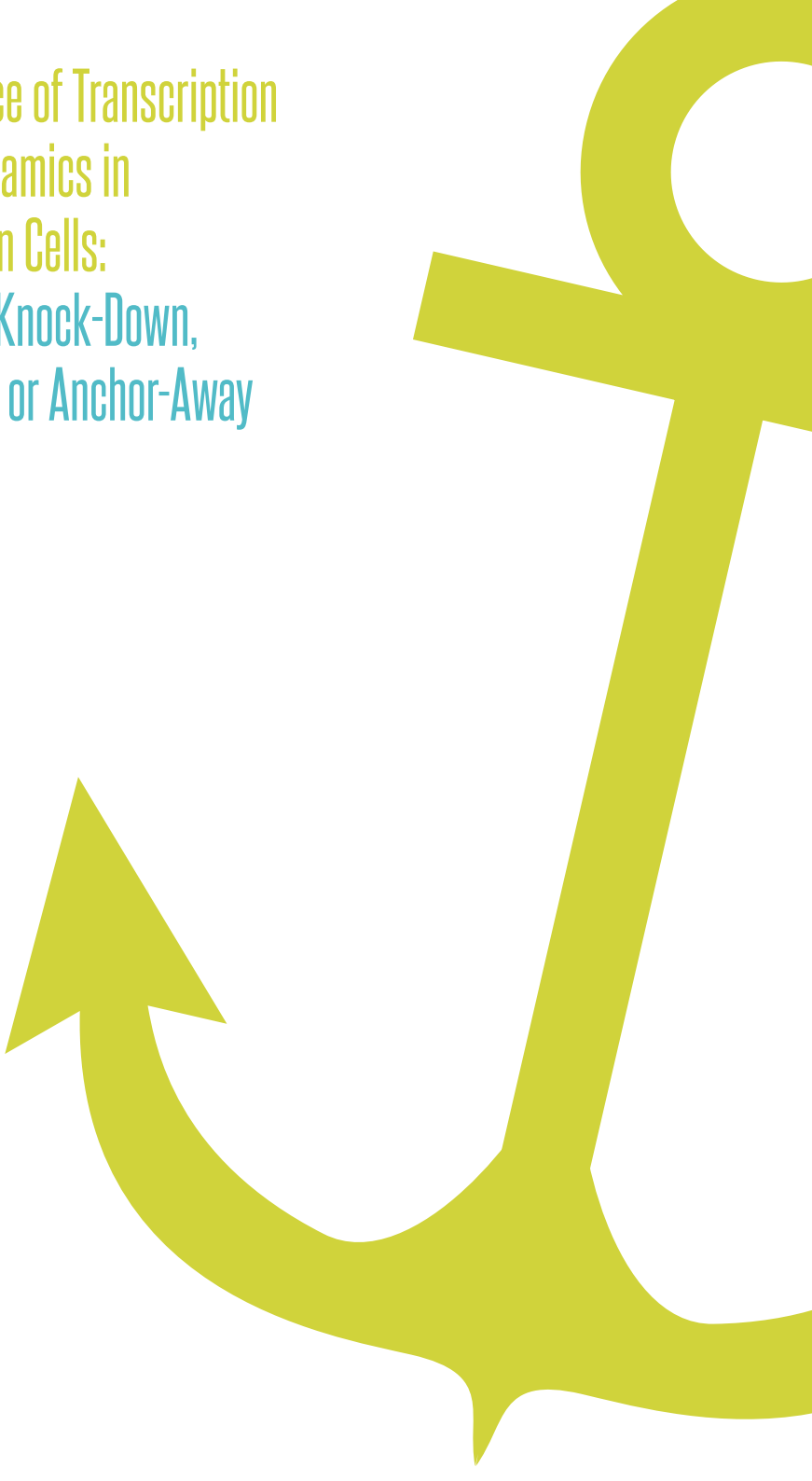
Disturbance of Transcription Factor Dynamics in Mammalian Cells: Knock-In, Knock-Down, Knock-Out or Anchor-Away

Ruud Jorna



Disturbance of Transcription Factor Dynamics in Mammalian Cells: Knock-In, Knock-Down, Knock-Out or Anchor-Away

Ruud Jorna

The studies presented in this thesis were performed in the Department of Cell Biology of the Erasmus MC in Rotterdam, The Netherlands. The department is a member of the Medisch Genetisch Centrum Zuid-West Nederland (MGC).

ISBN: 978-94-6228-243-8

Cover Design en Lay-Out: Appelsan
E-mail: info@appelsan.net

Printed by SMG Groep Hasselt

Disturbance of Transcription Factor Dynamics in Mammalian Cells: Knock-In, Knock-Down, Knock-Out or Anchor-Away

Het verstoren van de dynamiek van transcriptiefactoren in zoogdiercellen: knock-in, knock-down, knock-out of Anchor-Away

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam

op gezag van de rector magnificus
Prof. dr. H.G. Schmidt
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
dinsdag 1 oktober 2013 om 15:30 uur

door

Ruud Jacobus Johannes Jorna

geboren te Made en Drimmelen

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor:	Prof. dr. F.G. Grosveld
Overige leden:	Prof. dr. J.N.J. Philipsen Dr. R.A. Poot Dr. J. Essers
Copromotor:	Dr. E. Soler

Voor mijn ouders,
Voor Hilde en Marit

Table of Contents

List of abbreviations

3C	Chromatin conformation capture
3C-seq	3C combined with next generation sequencing
AA	Anchor-Away
ACH	Active chromatin hub
BAC	Bacterial artificial Chromosome
bp	Basepair
BL-CFC	Blast colony forming cell
CFC	Colony forming cell
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIP	Chromatin immuno-precipitation
ChIP-seq	ChIP combined with next generation sequencing
DNA	Deoxyribonucleic acid
ES cells	Embryonic stem cells
EB	Embryoid body
EHT	Endothelial to hematopoietic transition
GRN	Gene regulatory network
HEK293T cells	Human Embryonic Kidney 293 cells, stably expressing large T-antigen
Kbp	kilo base pair
LCR	Locus control region
LDB1	LIM domain binding protein 1
Mbp	Mega base pair
MYB	Myeloblastosis oncogene
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNA pol II	RNA polymerase II
TF	Transcription factor
TRN	Transcriptional regulatory network
WT	Wild type

Chapter 1
Introduction

Mouse embryonic stem cells

The mouse (***Mus Musculus***) has been extensively studied as a model organism. It is a placental mammal (like human), relatively cheap to house, small in size, has a short gestation time (19-20 days), produces large litters (6-8 animals) and becomes sexually mature quickly (6-8 weeks). Moreover, mouse inbred strains are available, considerably increasing the statistical power of genetic experiments.

A mouse embryo starts to develop after an oocyte is fertilized by a sperm cell. The fertilized oocyte or zygote will undergo several mitotic divisions until it has reached the 16 cell stage. This stage, where the conceptus consists of tightly coupled cells, is called morula. The first fate decisions are then made: part of the cells will contribute only to extra-embryonic tissues, while others will form the intra-embryonic tissues [1]. One stage later, at around embryonic day 3.5 (referred to as E3.5), the developing embryo is in the blastocyst stage. An overview of these first developmental steps is presented in figure 1A-F. The blastocyst is characterized by the presence of the inner cell mass and a fluid-filled cavity called the blastocoele [2]. Trophectoderm cells surround both (figure 1G). Trophectoderm cells are responsible for invading the uterine wall during implantation and they will also make up the fetal part of the placenta [3; 4]. The inner cell mass will be segregated into two distinct cell types: the epiblast cells that will form the embryo proper and a layer of primitive endoderm cells that separates the epiblast cells from the blastocoele (figure 1H, [5]). The primitive endoderm cells will divide upon implantation and develop into extra-embryonic visceral and parietal endoderm [6].

Mouse embryonic stem (ES) cells are isolated from E3.5 blastocysts, where they reside in the inner cell mass [7; 8]. ES cells are defined by two remarkable properties. The first is that ES cells are pluripotent, meaning they can differentiate toward any somatic lineage found in the mouse, both ***in vitro*** and ***in vivo***. In culture ES cells can be differentiated into all three primary germ layers (endoderm, ectoderm and mesoderm) [9; 10]. When ES cells are injected into E3.5 blastocysts and transferred to the uterus of pseudo-pregnant female mice, they can contribute to the generation of all tissues of the resulting chimeric animal [11; 12]. The second property is that ES cells can self-renew, preserving their pluripotent state. ES cells can be kept in culture indefinitely by keeping them on a monolayer of mouse embryonic fibroblasts (used as feeder cells), and in the presence of leukemic inhibitory factor (LIF) and fetal calf serum [13; 14]. Serum can be replaced by bone morphogenetic protein 4 (BMP4). BMP4 induces expression of ***Id*** genes via the Smad pathway, it was shown that forced ***Id*** expression combined with LIF addition is also sufficient for ES cell self-renewal [15]. However, ES cells do not grow well in that defined medium and are prone to differentiate. More recently it was shown that ES cells can also be kept in their non-differentiated ground state by addition of inhibitors of two kinases, MEK and GSK3 [16-18] to the culture medium. This chemically defined medium is called 2i medium. The inhibitors block the ERK and fibroblast growth factor (FGF) signalling cascades. It was shown that levels of NANOG, a transcription factor known to be essential to maintain pluripotency of ES cells, fluctuate much

less in 2i medium than in other media [291]. Fluctuations of Nanog levels do happen when ES cells are cultured in other media. Rat ES cells can be derived from rat blastocysts in 2i medium, a feat that could not be accomplished in other culture media. These experiments led the authors to conclude that 2i medium brings ES cells in the 'true' undifferentiated ground state [16]. Many labs across the world work on ES cells for a variety of reasons: ES cells provide a platform to study normal and abnormal (e.g. in absence of a gene) early embryonic development, ES cells are a great source for tissue engineering purposes because they can be differentiated and ES cells could be used to cure genetic diseases in the future. This last application has been illustrated in literature, as genetic mutations causing diseases have been corrected in human ES cells and shown to be cured in the specific adult cell type on several occasions [19-22].

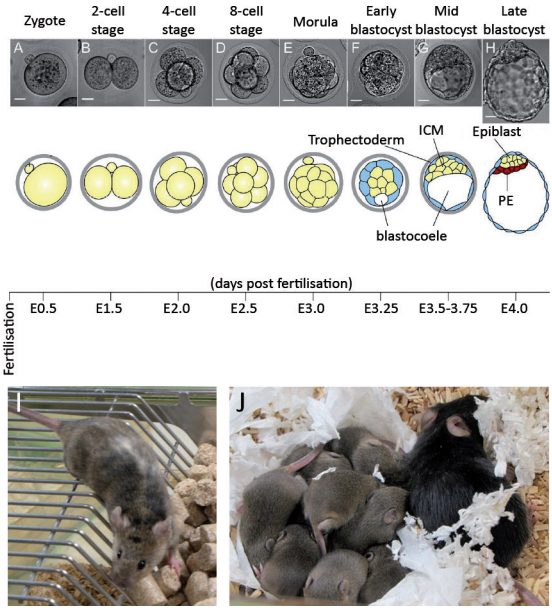


Figure 1. A-H) Stages in pre-implantation development of a mouse embryo. The inner cell mass (ICM) in the mid blastocyst at E3.5 contains pluripotent ES cells that can be extracted and cultured. Scale bars = 20μM. PE = primitive endoderm. Adapted from [1]. I) A chimaeric male was obtained by injecting ES cells obtained from a 129P2 mouse (light/agouti fur colour) into a C57/Bl6 blastocyst (black fur colour). Therefore the mouse has a striped black/brown appearance. J) A C57/Bl6 (black) female mouse was bred to the chimaera and gave birth to pups that were derived from the injected 129P2 ES cells (light/agouti). The pups are half C57/Bl6 (mother) and half 129P2 (light/agouti) and therefore have obtained a mixed fur colour (dark brown).

Genetic manipulation of ES cells can be achieved using (viral) random integration of deoxyribonucleic acid (DNA), insertion of DNA by transposases or homologous recombination (gene targeting, discussed below). The genetically modified ES cells can be used to obtain chimeric mice by injecting the ES cells into recipient

blastocysts. The resulting chimeric mouse embryos contain tissues that are derived partly from the wild type cells of the blastocyst and partly from the genetically manipulated ES cells. If the manipulated ES cells also contribute to the germ cells, the modification can be transmitted to the offspring, resulting in the establishment of a modified mouse line. An easy way to detect chimerism is to inject ES cells derived from a mouse strain with dark fur into blastocysts of white fur animals. From the fur colour of the offspring it is easy to judge the chimerism, as is shown in figure 1I and J. Usually a high level of chimerism translates into a high level of germline transmission of the ES cell derived modification.

Determination of gene function

DNA is said to bear the blueprint of life. The DNA within cells of organisms harbours all the information that makes up that specific organism. In order to use the information contained within the genomic DNA, cells first need to change the information to another, intermediate molecule: ribonucleic acid (RNA). This is achieved through a process called transcription. Finally, RNA can be used as a template for the ribosome to produce proteins. This process is called translation. The hierarchy of DNA-RNA-protein is called the central dogma of life. Proteins are the molecules that perform most of the chemical processes that take place within cells. All parts of the DNA that code for proteins are referred to as genes, the RNA that is transcribed from genes is referred to as messenger RNA (mRNA). The human genome contains approximately 20,000 protein coding genes [23; 24], while the mouse genome encodes for approximately 23,000 such genes [25; 26]. With the discovery of non-coding RNA the number of genes has been increasing. Non-coding genes are transcribed parts of the genome that are associated with genes as regulatory region or genes that code for functional RNA molecules that are not translated into protein by the ribosome. Several classes of non-coding RNAs have been discovered, they are conserved from bacteria to humans and have functions in all cellular processes [27].

Eight to ten percent of protein coding genes encode transcription factors (TFs). TFs are proteins found in all living organisms, and are essential in the control of gene expression. The different cells within an organism, despite having the same genome, are able to acquire different fates (e.g. liver cell versus lymphocyte) and perform different functions through the action of specific TFs. Expression of different sets of TFs in a given cell type will dictate whether certain genes are turned on or off, or whether their expression is positively or negatively modulated. TFs therefore play a central role in many biological processes such as cellular proliferation, differentiation, or responses to stimuli, and are involved in many diseases such as cancer or developmental disorders. TFs act at the level of transcription. They perform their functions by binding the DNA at specific sequences and gene loci, either via direct protein-DNA interactions, or indirectly as part of multiprotein complexes with DNA-binding activity. TFs and TF complexes recognize specific DNA motifs, partly explaining the selectivity of target genes between different TFs. Within TF complexes there may be chromatin modifying

enzymes, which are enzymes that modulate gene activity by chemically modifying the DNA or the histone proteins around which the DNA is wrapped. In this way TF complexes may have positive or negative activity on gene transcription, depending on the composition of the complexes. Specific interactions formed between cooperating or antagonizing TFs yield a dynamic network of interactions which fine-tune gene transcription levels and dictate whether a gene should be expressed or not expressed in a given cell type and at a given developmental stage. The different TFs present in a cell are able to regulate each other's expression and that of numerous other genes. Indeed, recent studies have shown that TFs are able to bind thousands of different loci simultaneously in a given cell type, forming transcriptional regulatory networks controlling cell behaviour. Elucidating the function of TFs and chromatin remodelling enzymes has been one of the key pursuits in cell biological research. One of the most straightforward approaches to assess protein function is by deleting (part of) its genomic locus, rendering the protein non-existent or non-functional (i.e. gene knock out). By observing the phenotype of a mouse after knocking out specific gene(s), one may draw conclusions about the function of the protein under investigation. Gene knock out cells can be obtained through gene targeting. The next part of this chapter will discuss the gene targeting process.

Gene targeting

The process of manipulating a specific gene by homologous recombination (i.e. non-random, targeted genetic modification) is referred to as gene targeting. This strategy takes advantage of the relatively high homologous recombination capability of mouse ES cells: when exogenous sequences with complete homology to a particular genomic region are introduced in ES cells, the endogenous DNA repair/recombination machinery will catalyse the exchange of homologous DNA strands thereby replacing the endogenous genomic sequence with the exogenous DNA. This process will keep the DNA sequence unaltered at sites of homologous recombination but can be used to introduce additional non homologous sequences in the genome (see below and figure 2A). Gene targeting will therefore modify genomic sequences at a specific locus, but leaves the rest of the genome unaltered. It can be used for several purposes, usually knocking-out of a specific gene or mimicking disease mutations or knocking-in exogenous sequences into a specific gene. In a gene knock out strategy, the function of a specific gene is disrupted by deleting (part of) the coding sequence on the DNA. A knock in refers to the targeted addition to a specific gene. The added sequence may be a traceable marker (e.g. a fluorescent protein) allowing spatio-temporal visualization of gene expression or a tag sequence. Tag sequences code for short stretches of amino acids specifically recognized by high-affinity antibodies or biochemical compounds. Tags are frequently used for biochemical purification of gene products or when antibodies raised against the endogenous factor are not available or have insufficient specificity.

Targeted mutation of a gene in ES cells was first described for the ***Hprt*** gene in 1987 [28]. Loss of the X-linked ***Hprt*** gene can be selected for in cell culture, making it an easy target. It took several more years before targeted mutation of genes without selectable phenotypes was reported [29-32]. These genes were mutated through gene targeting. A targeting experiment consists of several steps: first designing a targeting vector with selectable markers to select for properly targeted clones, performing homologous recombination in ES cells, selecting for the targeted clones and verifying the vector integration site and proper homologous recombination in the selected clones. Gene targeting in mouse ES cells has been optimized in many ways, dramatically increasing efficiency of the process. Despite these advances the whole process is still time consuming and necessitates custom work for each region to be targeted. Please note that the term gene targeting refers to manipulation of a gene. However any region of the genome, with the possible exception of centromeres and telomeres, can be targeted. In recent years, deletions of microRNA or enhancers are becoming more frequent (e.g. [33-35]) and also targeted mutation of a specific sequence/motif is possible [36; 37].

Steps in gene targeting

1. Obtain *E.coli* host strain containing BAC of interest
2. Electroporate plasmid harbouring recombineering genes into the same *E.coli*
3. Generate the adaptor cassette containing homology arms and desired mutation by PCR
4. Perform recombineering to insert adaptor cassette into the BAC to generate mutated BAC
5. Generate minimal plasmid backbone containing homology arms for targeting vector by PCR
6. Perform recombineering with minimal backbone and mutated BAC to obtain targeting vector
7. Verify targeting vector by digest pattern and sequencing
8. Purify targeting vector, linearize and electroporate into ES cells
9. Select ES cells that have integrated targeting vector by antibiotic selection
10. Pick ES cell colonies and expand
11. Verify correct integration in ES cell colonies

Box 1. An overview of the steps involved in a typical gene targeting experiment.

In order to design a successful targeting strategy it is important to have detailed knowledge of the region that needs to be manipulated. Genome browsers such as Ensembl, UCSC and VEGA and the ENCODE database are indispensible tools and made the generation of gene targeting vectors a lot easier. These online resources can be consulted to determine protein binding and expression of RNAs from the targeted region that may influence the outcome of the experiment. They also provide the genomic sequence, facilitating the design of the targeting vector itself and the downstream analyses (Southern Blot, PCR) to confirm targeting.

A gene targeting vector generally consists of two regions of homology to the genome that needs to be manipulated, the region of the DNA that is manipulated in the desired way, an antibiotic resistance marker that makes selection with antibiotics possible and a counter-selectable marker (see figure 2A). The counter-selectable marker kills cells that have integrated the vector randomly rather than by homologous recombination (figure 2B). Diphtheria toxin A has been shown to account for a three time higher targeting efficiency [38-40] and is widely used. An overview of the steps involved in a gene targeting experiment is given in box 1.

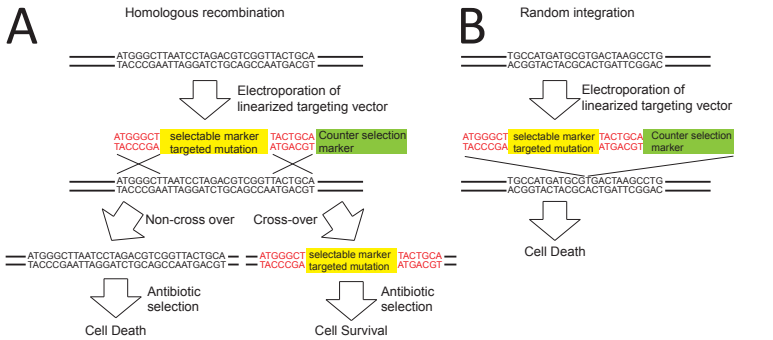


Figure 2. Homologous recombination. A) Principle of homologous recombination, this is the same for mammalian cells and bacteria. The example shown resembles targeting of a certain genomic sequence with a targeting vector. Exchange of homologous DNA takes place and results in a non-crossover or cross-over event. Typically, the cross-over event can be selected for with antibiotics. B) Random integration of the same targeting vector results in cell death because of integration of the counter-selection marker. This does not happen when homologous recombination takes place (compare to A).

Recombineering and BAC libraries

The construction of gene targeting vectors using traditional molecular cloning techniques in bacteria used to be tedious work. Development of BAC (Bacterial Artificial Chromosome) libraries and recombineering has dramatically improved that process. Recombineering (***recombin***ation-based genetic engin***eering***) is a technique that allows fast, seamless manipulation of DNA in bacteria without the need of restriction enzymes or size constraints. It makes use of homologous recombination in bacteria, which is performed using short homology arms (usually 50 base pair (bp) long). Recombineering was first developed by three independent labs [41]. Although the various methods are similar, there are small differences. The most widely used and most versatile is the method developed by the Stewart laboratory [42-45]. The method relies on recombination proteins found in the λ phage or the Rac prophage [43] that can be expressed in bacteria from a plasmid under control of an arabinose-inducible promoter. These recombinases are able to catalyse the insertion of exogenous sequences into a target DNA molecule (e.g. plasmid or BAC) at a given site using homology arms for homologous recombination. Standard vectors that can be used for gene targeting have been developed by the Stewart laboratory [46]. This makes construction of a targeting vector fast and straightforward.

Another advantage of recombineering is that the method can be used conveniently with large constructs such as BACs. BACs are DNA constructs mimicking bacterial chromosomes, containing a fertility origin of replication, a single copy origin that can be propagated in bacteria. Due to the capacity of BACs to host very large size constructs (up to ±400 kbp), they have been used to create libraries of mammalian genomic DNA [47; 48]. BAC libraries are a splendid resource for researchers to manipulate a certain part of the genome, because it provides practically any region of interest with homology arms suitable for homologous recombination [47; 48]. A graphical representation of the process performed to obtain a gene targeting vector from a BAC by recombineering is shown in figure 3.

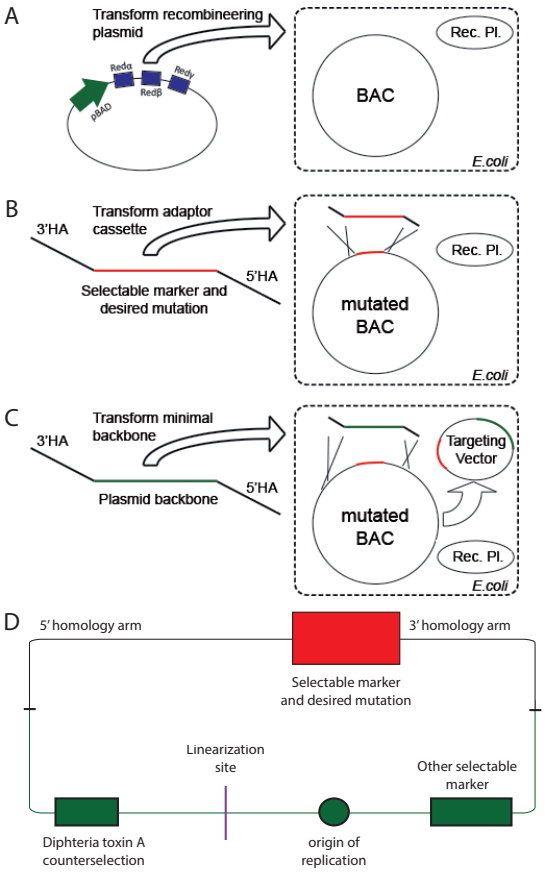


Figure 3. A-C) Recombineering involves homologous recombination in bacteria, for the principle see figure 2A. The pictures give a stepwise overview how a gene targeting vector can be made from a BAC containing wild-type genomic sequence in two rounds of recombineering. The red colour indicates the sequence containing the desired mutation and the selectable marker. In green is the plasmid backbone. 5'HA = 5' homology arm; 3'HA = 3' homology arm; Rec. Pl. = recombineering plasmid.
D) A typical gene targeting vector. The 5' and 3' homology arms (black line) are long (2-10 kb) stretches of DNA homologous to that of the target genome. The origin of replication and other selectable marker are for propagation and selection in bacteria and reside on the plasmid backbone (green line). The selectable marker (red box) is for antibiotic selection in mammalian cells and bacteria. Diptheria toxin A provides counter selection.

BACs can be modified by recombineering so that they can be used for transfection in ES cells. If the whole gene locus of a gene is present on the BAC, this should faithfully recapitulate expression of the target gene (also ***in vivo***) [46; 49]. It is also possible to fluorescently label a gene **and** make point mutations in e.g. enhancers, to follow what happens with the spatio-temporal expression of the gene when

specific regulatory sequences are mutated. Linear integration of a BAC into the genome has been shown by the use of the PiggyBac transposon system in fertilized oocytes [50] as well as ES cells [51]. Combining these techniques harbours a promise for elegant experiments that can be done in future. For now, the most frequent use of BACs in recombineering is to obtain homology arms for targeting constructs. After the production of a targeting construct, one proceeds to ES cells to perform homologous recombination (box 1).

Homologous recombination in ES cells

Homologous recombination is an essential part of machinery for the cell to repair DNA damage and sort sister chromatids during meiosis [52], but the process can be manipulated to incorporate specific mutations into a cell's genome [29]. Many cell lines show proficiency for targeting by homologous recombination, mouse and human ES cells included [29; 53]. Mouse ES cells seem to be particularly efficient in homologous recombination [40], unlike human ES cells. A homologous recombination reaction involves two stretches of identical or similar molecules of DNA. The target DNA, that needs to be replaced (e.g. because of a double stranded break) is cut around at the 5'ends in a process called resection. It is thought that the linear homologous DNA presented to the cell during gene targeting is recognized by the cell's DNA repair machinery as a double stranded break. Next, in the strand invasion step, the overhanging 3' end of the replacing DNA molecule 'invades' a target strand, forming a structure that is known as a Holiday junction. The Holiday junction is an unstable structure that needs to be removed by nucleases that cut at the 3' ends of the junction (the process of homologous recombination is reviewed in [54]). The resulting sequence of DNA can be one of two things: cross-over or non-crossover, see figure 2B. Proper gene targeting results from a cross-over event and can be selected for by a selectable marker. Of note, one cannot distinguish between genuinely targeted loci and random integration (which is not mediated by homologous recombination) of the target DNA by the selectable marker. The counter selection marker helps selecting homologous recombined clones over clones with random integrations (see fig. 2A vs 2B).

Although the process of targeting by homologous recombination is not completely understood, some basic rules to increase efficiency have been described. Longer homology arms give better targeting efficiency [55], but the effect is dampened after ±10kb of homology [56]. Second, isogenic DNA performs better than non-isogenic DNA [57]. This may account for lower efficiency of targeting in human ES cells: inbred humans do not exist, and BAC library are usually from a different human genome than that present in the different human ES cell lines. The third measure that makes targeting more efficient is the addition of the counter-selectable marker [39].

Several strategies were adopted to delete protein function by gene targeting in order to create a specific gene knock out mouse. The most important ones are discussed in the next section.

Gene targeting strategies

Gene traps and traditional knock out alleles

A gene trap strategy relies on integration of a 'trap' vector in the genome. Gene trap vectors typically consist of a splice acceptor, a selectable marker without promoter, a reporter gene (lacZ, antibiotic resistance gene or a fluorescent protein in general) and a stop signal (figure 4B). When the vector is integrated in introns of an expressed gene, it will be spliced to the gene exons during transcription. Hence integration of the 'trap' vector leads to a modified protein sequence and as a result will often generate truncated protein products. Although a gene trap can be targeted to a specific gene, their main application is to produce a library of knock out cells by random integration of the gene trap vector [58-61]. If the cells used are mouse ES cells, they can be turned into reporter mice. Reporter mice show staining only where the gene is active in the mouse embryo [59]. The mice can be crossed to become homozygous gene trapped and theoretical knock out mice.

The advantage of the technique is that is relatively fast, easy to screen by expression of the reporter gene and flexible. A drawback is that splicing around the gene trap has been reported on several occasions [59; 62; 63], as a consequence the gene trapped allele is not a knock out allele. Another consideration is that the gene trap could integrate at the 3' side of an expressed gene, leading to a protein product that is truncated close to the C-terminus. Such protein products may be (partially) functional, not recapitulating a full knock out phenotype in cells or mice. Finally, selection of gene trapped cells is limited to expressed genes only since transcription needs to be active on the locus in order to splice to the gene trap construct.

A way to confidently knock out a protein fully is to target a selectable marker to (part of) a gene (figure 4C). This leads to the complete deletion of that part of the gene, and loss of (functional) protein. This strategy has been successful in elucidating protein function for many proteins [31; 32; 64-67]. As understanding of the genome improved, scientists learned that knocking out a critical exon of a gene is a better strategy than knocking out the whole or big parts of the genomic locus of a gene. When big parts of the genome are taken away, one may affect microRNAs and/or regulatory sequences for other genes [68; 69].

The advantage of the traditional knock out over gene traps is that it is more controlled, i.e. the modification is user defined. A knock out can be confirmed at the protein level, the risk of splicing around the cassette can be avoided. If a reporter gene is added to the selectable marker, it is possible to obtain a reporter knock in, functional knock out mouse. The traditional knock out is less suitable for genome-wide studies or screens but offers the possibility to be spatially and temporally controlled allowing precise characterization of gene function. The generation of conditional knock out alleles is detailed in the next section.

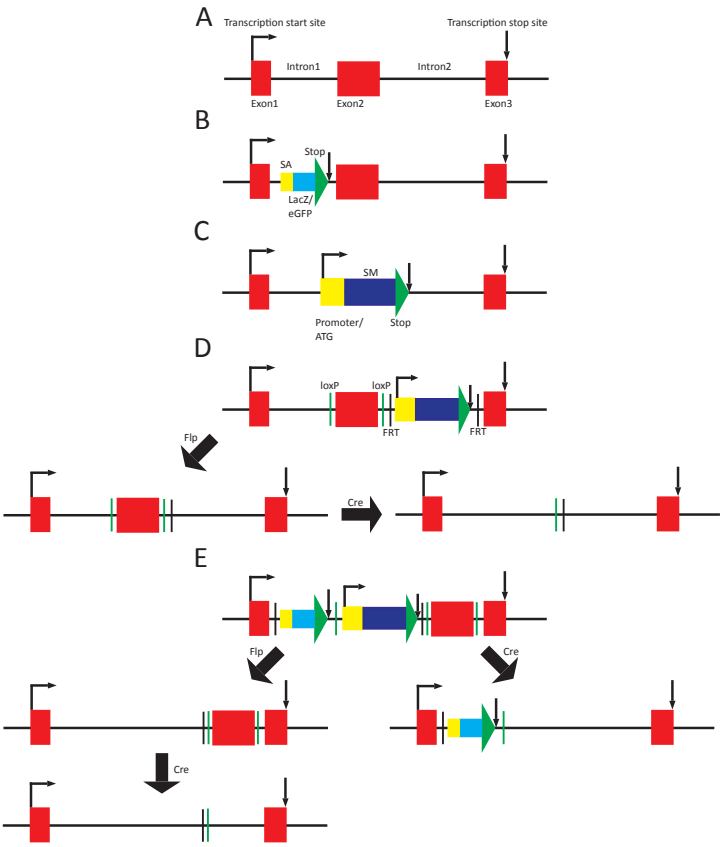


Figure 4. An overview of different targeting cassettes used to knock out a hypothetical gene. Sizes are not drawn to scale. A) normal, non-targeted situation. Exon 1 of this hypothetical gene contains the start codon (ATG) and exon 3 contains the stop codon (stop). B) Gene trap, the gene trapped gene is interrupted by alternative splicing via the splice acceptor. The colour marker (LacZ / eGFP) and/or selectable marker are expressed after the first exon of the gene and then expression is stopped. SA = Splice Acceptor C) Traditional knock out allele, a selectable marker (SM) is expressed instead of the second exon. This completely disrupts expression of the second and third exon, resulting in a knock out. D) The conditional knock out allele. The second (critical) exon can be knocked out by Cre recombinase expression, causing the gene to be expressed faulty. The selectable marker can be removed by Flp recombinase expression. E) Combination of gene trap and conditional knock out allele. After targeting, the gene is gene trapped and thus knock out. To ensure full knock out of the protein the selectable marker and second exon of the gene can be removed by Cre expression. If Flp recombinase is first expressed, the selectable marker and gene trap are removed, resulting in a conditional knock out allele. The conditional knock out allele is converted to knock out by Cre expression after Flp expression.

Conditional knock out alleles

The techniques described earlier lack conditionality. This can be problematic when a very early embryonic lethal phenotype is observed, or if one wants to study the effects of a knock out in a specific (adult) tissue whereas it may be essential in another tissue at an earlier stage. In these cases, the option to conditionally knock out a gene in a spatio-temporally controlled manner is needed.

Thus conditional knock out allele strategies were developed. These strategies rely on inducible recombination of a genomic area of choice. The area to be conditionally knocked out is placed in between sites recognized by a site-specific recombinase [40; 70-75]. The selection marker can be removed, as well as the area of interest by expressing the site-specific recombinases (figure 4D). More advanced variations, for instance combining a conditional gene trap with a conditional allele have also been devised and are routinely used (figure 4E).

Three site-specific recombinases can be used to generate conditional knock outs: Cre, Flp and Dre. All site-specific recombinases work in a similar way: they have a recognition sequence of 34 bp (33 for Dre) consisting in two palindromic repeats of 14 bp and a 6 bp (5 for Dre) binding sequence [70-74]. Variations in the palindromic repeats can be tolerated, this is well documented for Cre recombinase [76], but the binding sequence is rigid. One recombinase molecule binds its recognition sequence, after which it can make contact another recombinase molecule bound to another recognition sequence. When the two are brought together, the DNA in between the two recognition sequences is removed [77]. Many Cre knock in mouse lines exist, expressing (inducible or tissue/stage specific) Cre on certain developmental windows/tissues as specific proteins. These mouse lines are used to remove conditional regions in the conditional knock out allele.

The conditional knock out strategy has great advantages: gene function can be shut down in a tissue of choice, at the preferred time. Typically the knock out is relatively fast, after ±12h of Cre expression a locus is removed and a typical protein is ablated ±48 hours after Cre expression [78; 79], although these numbers vary on a gene-to-gene basis as they depend on protein half-life. All other advantages of a traditional knock out still apply. The need for Cre transgenic mice is a disadvantage, it complicates breeding schemes and the Cre knock in of choice may not always be available. In any case it lengthens the time to obtain the mouse model desired. In general, obtaining a (conditional) knock out mouse is a time-consuming process.

Novel ways of genome engineering

A major drawback of homologous recombination has been that only one allele can be targeted at once. The majority of genes are present in two copies in the genome. If both need to be mutated by homologous recombination, one has to perform serial targeting experiments, or breed the generated animals to get homozygous lines.

Several new ways of engineering genomes, that show promise to target two alleles at once, have been described in literature recently. The three most important candidates are zinc finger nucleases, Transcription Activator-Like Effector (TALE) nucleases and Cas9 nuclease (figure 5). Zinc finger and TALE nucleases both rely on fusion of a DNA binding protein domain to a restriction enzyme, FokI. Cas9 is an enzyme that cuts DNA, when guided by RNA molecules. All systems introduce double stranded breaks by digesting the DNA. These double stranded breaks attract molecules for homologous recombination. This targeted recruitment of the homologous recombination machinery makes the process of homologous recombination more efficient. It has been shown that gene targeting with these systems can be efficient when only short stretches of homologous DNA are used in the targeting constructs (0.5-1kb), and they offer the possibility to target two alleles at once [80-84].

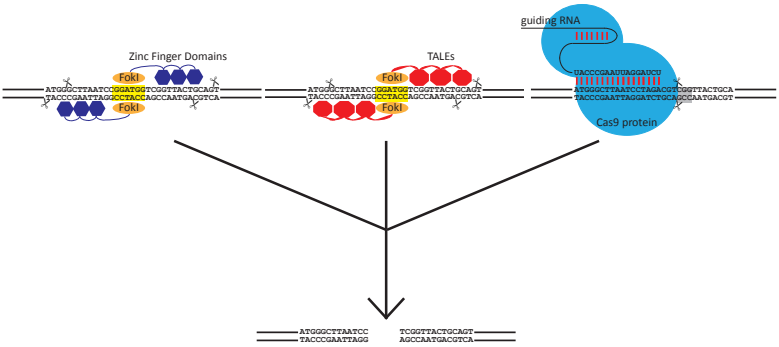


Figure 5. Overview of three novel genome editing methods. From left to right: zinc finger nucleases (ZFNs), TALENs and the Cas9 system. The three systems guide a nuclease (FokI for TALENs and ZFNs, Cas9 in the last system) to a specific site in the genome (recognition sequence). The guiding elements are protein zinc finger domains for ZFNs, protein TALE domains for TALENs and a guiding RNA for the Cas9 system. With all the systems a double stranded break is made at a specific site in the genome. This double stranded break recruits homologous recombination factors, improving gene targeting efficiency (discussed above and figure 2). A more detailed description is given in the text. The FokI recognition sequence is depicted with a yellow box, the Cas9 NGG motif is shaded gray.

Zinc fingers are highly conserved protein domains known to bind DNA with reasonable specificity. The first next generation genome engineering tool to be devised relies on zinc finger domains fused to FokI. FokI has a GGATG recognition sequence, and makes a double stranded break 9bp upstream and 13bp downstream of this sequence [85]. FokI has two domains, a DNA-binding domain and a DNA-cleavage domain that need to hetero-dimerize in order to digest DNA [86-88]. The zinc fingers can be manipulated to recognize a target DNA sequence [80], although producing zinc finger pairs with good specificity is a challenging task. Good results have been obtained in cells of different model organisms [21; 80; 89]. The major problem seems to be that zinc fingers have a short recognition sequence, making it challenging to create specific binding zinc fingers with no

off-target binding [90]. For some loci (target sequences) very efficient zinc fingers were generated, but target sequences have to be chosen carefully [91]. Obviously, the system is dependent on the presence of a FokI recognition site. Targeting of both alleles at once has been reported [80; 92], but efficiencies differ greatly between different loci.

The second class of genome engineering tools are the TALE nucleases. This system consists of a domain of a TALE protein fused to a FokI domain. TALE proteins are proteins secreted by ***Xanthomonas*** bacteria and have longer DNA recognition sequences than zinc fingers [84]. The method to engineer TALE proteins to bind specific sequences of DNA has been elucidated and TALE domains with affinity for any target sequence can be generated by traditional cloning experiments [82]. TALENs have shown promising results in cell lines, and homozygous targeting is possible [93]. The possibility of off-target effects is still present, and it was found that a target sequence that is present many times in the genome can overwhelm the DNA repair machinery, leading to chromosomal rearrangements and/or cell death [90]. The TALE nucleases, like the zinc finger nucleases, rely on the presence of a FokI recognition sequence.

The most recent development on genome editing is the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated (Cas) system. Originally the system was identified as a principle defence system of bacteria against foreign (plasmid) DNA entering [94-96]. The method works via RNA, rather than protein, guiding the nuclease to the recognition DNA sequence [97]. The nuclease is the Cas9 protein, which is capable of making double stranded breaks in linear double stranded DNA and can be purified from bacteria. In nature two RNA molecules are needed to guide Cas9 to a target sequence, but it was shown that one artificial guide RNA of ±50 nucleotides can be used alternatively [97]. Cas9 cuts three base pairs after the recognition sequence and is dependent on a NGG motif that has to be present directly after the cut site [98; 99]. Since the Cas system has only been developed recently, more thorough testing for off-target effects remains to be done [100]. Other features, such as efficiency of targeting in general and homozygous targeting in particular will also have to be tested more thoroughly, although some quantitative information is becoming available [97-99]. For the moment, the Cas system seems very promising to edit the mammalian genome, especially because of the ease of use. Transfection of the target DNA, a plasmid coding for the Cas9 protein and a plasmid coding for the guiding RNA simultaneously suffices to perform gene targeting using the Cas system [100].

Other techniques to disturb gene function

Knock down by RNA interference (RNAi)

An easy and relatively fast solution to inactivate gene function is using RNA interference (RNAi), an RNA-dependent gene silencing process. This approach consists of knocking down the expression of a gene of choice by the use of

small interfering RNAs (siRNAs), leading to mRNA degradation. On the basis of this system revolutionary work was done on the RNAi pathway in the nematode ***Caenorhabditis elegans*** and plants [101-106]. The cellular system seems to be conserved among most (but not all) eukaryotes [107-112]. In short, it works by introducing short double stranded pieces of RNA into the cell. This can be done directly, through transfection of long double stranded RNAs, or by incorporation of a plasmid that harbours a short hairpin RNA (shRNA, coding for siRNA precursors) under the control of an RNA polymerase III promoter [106; 113-115]. The shRNA and long double stranded RNA methods were developed to avoid direct introduction of siRNAs into the cells, which made using the RNAi pathway to knock down genes more robust [116]. Both long double stranded RNAs and shRNAs are recognized by a protein called Dicer, which processes it to produce 15-21 nucleotides long double stranded siRNAs [117-119]. The strands of the siRNAs are separated, and they integrate into an active RISC complex. After integration into the RISC complex, single stranded siRNAs can bind to their complement sequences on mRNA. The mRNA cannot be used in translation and is targeted for degradation [120]. Since the mRNA cannot be used for protein production, a drop in protein level is generally observed. An overview of the system is given in figure 6. The precise timing to reach efficient knock down of a gene depends on transcript abundance, protein abundance and half-life of the protein. However, despite its ease of use, some drawbacks are associated with this system such as low specificity caused by off-target knock down of genes. With improved understanding of the system off-target effects may be avoided [117]. An example of this is the development of esiRNA [121; 122], short siRNAs generated from a long dsRNA by an endoribonuclease digest (hence the name esiRNA). The long dsRNA is stable throughout the transfection and after it is cut by a simultaneously transfected endoribonuclease (the E.coli Dicer or RNase III are mostly used), a big heterogeneous pool of esiRNAs is produced. All esiRNAs target one specific mRNA, making esiRNA a specific and effective way to knock down proteins. Overall, RNAi mediated knock down has proven to be a very good method to elucidate protein function. The RNAi pathway is also amenable to high-throughput applications, therefore it can be used to perform drug target screening.

The advantages of the RNAi system are its ease of use and its flexibility. Repositories containing shRNA sources for virtually every gene in a number of genomes are available. It is possible to obtain shRNA for your favourite gene and knock it down in any cell amenable to transfection or viral transduction. Viral transduction of primary cells has been successfully performed for a number of tissues [123-127]. siRNA mediated knockdowns can be applied in large screens, targeting many genes, or targeting several members of the same protein complex [128]. The time needed to reach significant knock down efficiency, i.e. the time needed to significantly decrease mRNA and protein levels after introduction of siRNA/shRNA, is comparable to the time needed to inactivate a gene after induction of Cre recombinase expression in conditional gene knock outs (i.e. 24-72h). When compared to a knock out the drawback of the RNAi is that knock down is never complete, some target protein remains in the treated cells [117]. Most often a full knock out is not needed to see a phenotypic effect, however in some cases siRNAs

do not provide sufficient reduction of protein levels. A second drawback is that some cells are not amenable to transfection and are too delicate to survive viral transduction of RNAi constructs. Such cells cannot be manipulated by the RNAi system.

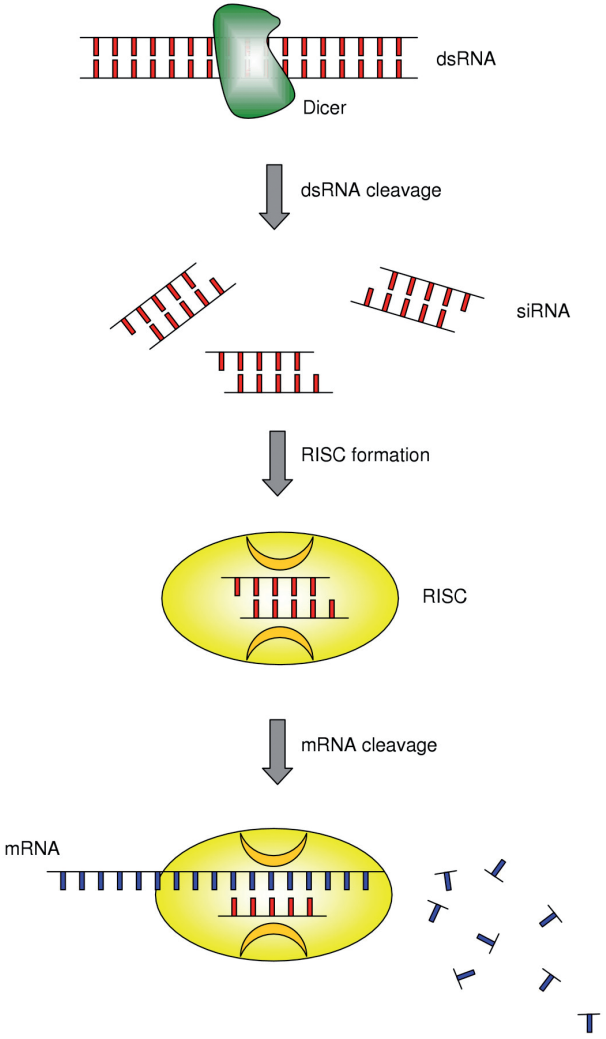


Figure 6. RNAi knockdown in mammalian cells. The protein DICER recognizes double stranded RNA (or shRNA) and cleaves them to short, double stranded siRNAs. The siRNAs bind the RISC complex and the strands are segregated. Single stranded siRNA in the RISC complex binds to mRNA, mediating cleavage of the target mRNA. Picture is taken from [288].

Finally, a common drawback between both conditional knock out and RNAi mediated knock down is the time needed to deplete the gene product (protein). Indeed, the time needed by Cre recombinase or RNAi to achieve gene knock out or down is usually within the range of several hours to several days (24-72h). This time frame is very long compared to the time needed for transcription to take place, which occurs within the frame of minutes to a few hours. Hence, when knocking out or down a gene involved in transcription regulation such as a gene coding for a TF, it is almost impossible to discriminate between primary (direct and specific, genes regulated by the TF) and secondary (indirect and non-specific, genes not directly regulated by the TF but e.g. by genes regulated by the TF) effects of gene removal on global transcription regulatory networks. This is illustrated by data showing that typically not more than 30% of the genes that are misregulated after knock out of a specific TF are also bound by the TF as determines by massive parallel sequencing after chromatin immune-precipitation (unpublished observations, [130; 131]). This implies that the amount of secondary effects is substantial. In order to separate primary from secondary targets, and better characterize TF functions in normal development and disease, faster ways to perform a knock down in mammalian cells are clearly needed.

Anchor-Away

An example of a particularly promising faster way to knock out protein function is the Anchor-Away technique published in 2008 [131]. The authors show that a target nuclear protein can be sequestered in the cytoplasm in baker's yeast (***Saccharomyces cerevisiae***). It is achieved by fusing both the target nuclear protein and a ribosomal protein (the anchor) to tags consisting of cellular protein domains that are able to heterodimerize in the presence of the small drug rapamycin. The anchor ribosomal protein is normally incorporated into the ribosome, and maturing ribosomal subunits are transported in a large flow from the nucleus to the cytoplasm [132; 133]. Upon addition of rapamycin to the culture medium, [134], the two rapamycin binding domains (tags) bind to rapamycin and to each other forming a high affinity ternary complex. This causes the target nuclear protein to become physically attached to the ribosomal subunit, and to be transported out of the nucleus into the cytoplasm. An overview figure of this procedure is provided in figure 7A. Once the nuclear protein resides in the cytoplasm, it cannot go back to the nucleus, so its nuclear activity is effectively inactivated. In the original study Haruki ***et al.*** observe growth phenotypes of yeast, comparable to those observed in knock outs of the same gene [131]. The advantage compared to conventional knock out or knock down is its incredible speed since due to the large flow of ribosomal proteins throughout the nucleus towards the cytoplasm, the nucleus can be depleted of TF in less than one hour in yeast.

Anchor-Away is a system which, if applicable to mammalian cells, may be fast enough to dissect primary and secondary effects of a knock down. It may also offer the possibility to differentiate between genes that are primary targets, but react with different kinetics to the ablation of a target protein. In addition, thanks to its speed Anchor-Away may allow the study of effects of specific nuclear proteins

and TFs during development and differentiation where transitions in cell fate occur rapidly during short time windows. In addition, Anchor-Away may offer the possibility to study the role played by essential factors for which knock out and/or knock down are toxic and cause cell death. Being able to analyze phenotypic effects quickly after inactivation and before lethality occurs may overcome this limitation.

Some characteristics of the Anchor-Away system have not been fully established at the moment. One major question is whether only the TF of choice is pulled out of the nucleus, or direct binding partners will also move along with the TF to the cytoplasm. Secondly, the system has not been tested in mammalian cells, and transferring it from yeast to mammals may be challenging (see chapter 3 & 4).

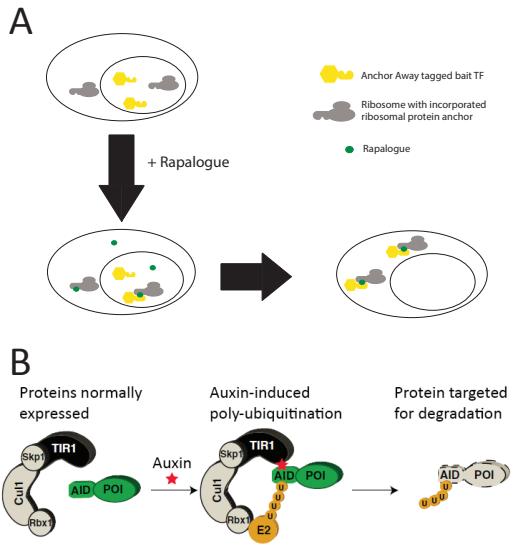


Figure 7. Potential systems that can achieve fast protein disruption.
A) Anchor-Away. A ribosomal protein anchor and a target transcription factor are expressed with rapalogue-binding tags. Upon addition of the rapalogue, ternary complex formation takes place. This causes the transcription factor to be dragged along with the ribosome to the cytoplasm.
B) auxin inducible degron (AID) system. The protein of interest (POI) is expressed with an AID domain tagged to it. TIR1 (black) has to be separately expressed in the same cell, the rest of the complex (grey) is present in mammalian cells. Upon addition of Auxin, binding between the AID domain and the TIR1 takes place. This is followed by E2 ubiquitin ligase recruitment and poly-ubiquitination of the AID domain. Poly-ubiquitination is a signal for degradation by the proteasome. Picture is adapted from [140].

As stated above, the main advantage of Anchor-Away in mammalian cells would be its speed. Since Anchor-Away directly acts at the protein level rather than the RNA or DNA level, one does not have to wait until the endogenous protein is

degraded. Anchor-Away is conditional and can in theory be applied to any nuclear protein(s), to inactivate the nuclear function of these protein(s). A disadvantage of the Anchor-Away technique is that it needs engineered cells expressing both the tagged ribosomal protein and the homozygously tagged protein of interest. In its present form it is not amenable to high-throughput analyses of nuclear factors, but rather represents an original tool for in depth characterization of selected factors.

The auxin inducible degron system

Another way to ablate a protein is to be able to conditionally influence its stability, by targeting it to degradation by the proteasome. Several systems have been described that explore this avenue [135-139]. One of the most recent and most promising is the auxin inducible degron (AID) system [135; 140]. It relies on the expression of three proteins that are conserved among all eukaryotes, the ubiquitin ligases SKP1, CULLIN and F-BOX protein, which form a complex (SCF-complex). In plants there is another protein called TIR1 that binds to the SCF-complex to poly-ubiquitinate proteins that contain an AID domain. Poly-ubiquitination is a well-known signal for degradation by the proteasome [141]. There are no known proteins with an AID domain in mammalian cells, nor is ***Tir1*** present in the mammalian genome. But if plant ***Tir1*** is expressed in mammalian cells, it does bind to the SCF-complex and is shown to degrade proteins fused to an AID domain (if expressed) [135; 142]. This makes it possible to engineer cells that express both TIR1 and target protein of choice fused with an AID domain, making inducible degradation of the target possible. The system is described in figure 7B. AID was recently tested in human cells [140], and shows promising results. The real test, endogenous degradation of tagged proteins in a target cell, still needs to be done to test the system to its full extent.

Hematopoiesis

Hematopoiesis refers to the process of blood cell production. It represents the largest quantitative cellular output of the adult body and has been a paradigm for many developmental processes and diseases. The function of the cell types differs and ranges from oxygen transport (red blood cells or erythrocytes) to defence of the body against disease (white blood cells) and blood clotting (platelets). The different blood cell types have limited life spans, and need to be replenished continuously. All the blood cells are continuously generated and replenished from a rare pool of hematopoietic stem cells (HSCs). Apart from differentiating, the HSCs also need to self-renew in order to sustain a pool of HSCs. In order to produce the enormous amount of new blood cells generated each day, immature hematopoietic progenitors which derive from differentiation of a small number of HSCs need to expand tremendously to meet the body's demands. An overview of the hematopoietic system is given in figure 8A. HSCs reside in the bone marrow during adult life. The bone marrow HSCs receive stimuli from their micro-environment [143-146] to maintain the balance between self-renewal and differentiation of HSCs [147].

(Mis)regulation of hematopoiesis

Hematopoiesis is tightly regulated by external signals to the developing blood cells. Externally, the micro-environment provides molecules that can bind to receptors on the cell-surface of the developing blood cells. This general process is called cell signalling and has been shown to stimulate many different cellular processes [52]. In this particular case, cell signalling can lead to the establishment of gene expression programs favouring the differentiation to a particular kind of mature blood cell. Essential proteins in these gene expression programs are transcription factors TFs. TFs achieve the dynamic transitions in gene expression profiles during hematopoiesis, acting during small developmental windows or within specific hematopoietic lineages. For instance, the GATA family of essential TFs act differently during hematopoiesis: ***Gata2*** is expressed early in HSCs and immature progenitors, whereas ***Gata1*** expression is switched on later and drives the differentiation towards the erythroid and megakaryocytic lineages. ***Gata3*** is expressed in the lymphoid lineage and participates in the production of T cells. TFs do not act alone, they need other TFs to make up a TF complex that exerts its function. This gives some flexibility, as specific TFs can form complexes with different factors to regulate expression of different sets of genes. It has been reported for TFs that they are part of different complexes, even within the same cell type [148; 149]. Within a complex, TFs can have different functions, from DNA binding to recruitment of other complexes or act as a docking site for other TFs that need to be in close proximity. A description of particular TFs playing key roles in hematopoiesis follows later in this chapter.

The balance between proliferation and differentiation of hematopoietic progenitors is tightly controlled in order to maintain tissue homeostasis. TFs play an essential role in controlling this delicate balance and alterations in TF function have been linked to the development of disease. Different kinds of leukemia are examples of this [150-155]. The problem of misregulated differentiation can occur at any point along each of the differentiation pathways (see figure 8A), resulting in different disease phenotypes. In acute lymphoblastic leukemia for instance, where too many lymphoid cells are produced. This results in accumulation of immature lymphocytic cells in the bone marrow, overcrowding other cells [151]. The cause of such misregulation can stem from within the (developing) blood cells, but faulty or absent stimuli of the micro-environment can also be causative. The diseased and healthy state of HSC differentiation is subject of intensive studies. A better understanding of these may help in developing new and better drugs to cure leukemia. Another subject that attracts a lot of interest of the scientific community is the characterization of the molecular events underlying the emergence of blood (stem) cells in the embryo.

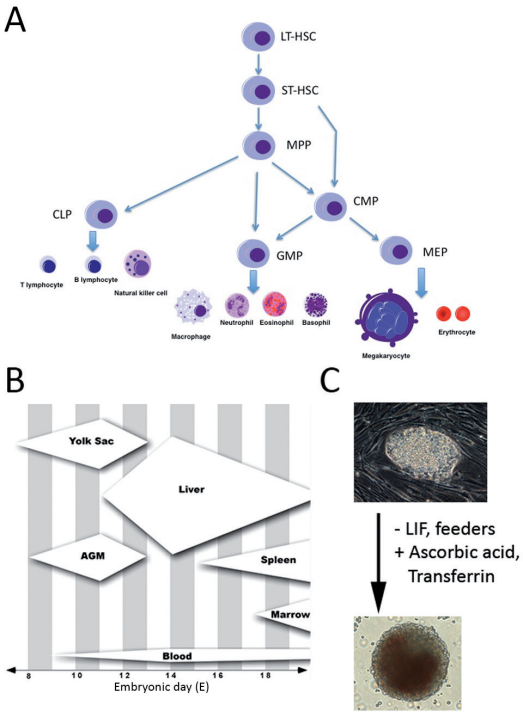


Figure 8. A) The hematopoietic system. LTHSC = Long-term repopulating HSC, STHSC = Short-term repopulating HSC, MPP = multipotent progenitor, CMP = common myeloid progenitor, CLP = common lymphoid progenitor, GMP = granulocyte/macrophage progenitor, MEP = Megakaryocyte/erythrocyte progenitor, RBC = red blood cell. Adapted from [289] B) Hematopoietic sites during mouse development. Day of development is indicated at the bottom of the picture. AGM = Aorta-gonads-mesonephros region. Adapted from [290] C) Picture of an ES cell colony and an embryoid body at day 7 of differentiation. A blood island has formed within the embryoid body (red colour). LIF = Leukemia inhibitory factor

Hematopoiesis in the mouse embryo

Hematopoiesis occurs in two waves in the mouse embryo. The first, or primitive, wave starts at embryonic day 7.5 (E7.5), corresponding to the presence of the first blood cells in the developing mouse embryo. They are erythroid progenitor cells and are located in the yolk sac [156]. Emergence of endothelial and hematopoietic cells coincides in the yolk sac, which led as early as the 1920's to the hypothesis that there is a common progenitor for endothelial and hematopoietic cells [157-159]. This common progenitor was named the hemangioblast. Cells with the expected hemangioblast-like properties (both endothelial and hematopoietic potential) have been isolated from developing mouse and human embryos [160-165], to demonstrate the existence of the hemangioblast. In 2000 it was shown that the

hemangioblast isolated from mouse embryos also harbours the potential to form smooth vascular muscle cells [166]. The hemangioblast is also referred to as the ***Flk1***+ cell, since this cell can be identified by the expression of cell-surface receptor fetal liver kinase 1 (***Flk1***) [167-169]. Other hematopoietic (progenitor) cells can be found in the yolk sac later in development (E7.5-E10), but no ***bona fide*** stem cell activity could be detected in these early yolk sac cells [170].

The second, or definitive, wave of hematopoiesis is characterized by the emergence of definitive progenitors in the mouse embryo. During this wave the ***bona fide*** long-term repopulating HSC is produced. This HSC is found from E10.5 onward in the developing embryo. It is first found in the developing aorta within the aorta-gonads-mesonephros region and in the viteline and umbilical artery [170-173]. The placenta harbours true HSC activity as well around E10.5, although it is not clear whether this stems from ***de novo*** generation or early colonization [174]. Much research has been performed to clarify that HSCs actually stem from hematopoietic clusters that bud out of the ventral wall of the dorsal aorta [170; 175]. The clusters emerge from an endothelial layer of cells, ***in vivo*** evidence that endothelial and hematopoietic cells share a common ancestry. Endothelium harbouring hematopoietic potential is referred to as hemogenic endothelium [160]. Under influence of specific factors hemogenic endothelium can give rise to hematopoietic (stem) cells in a process termed endothelial to hematopoietic transition [176; 177]. Recently cells that bud out of the clusters into the circulation of the mouse E10.5 AGM have been visualized [178], evidence is strong that this event is in fact the 'birth' of an HSC in the mouse embryo.

As soon as HSCs enter circulation they start to colonize other, secondary, hematopoietic tissues [179]. The fetal liver is an important secondary hematopoietic tissue that rapidly expands the amount of HSCs in between E12 and E15 [170]. During this time, the fetal liver is the most important hematopoietic organ in the mouse embryo. All kinds of hematopoietic progenitors can be found around this time in the fetal liver [180]. For instance, the fetal liver on E14 is often used as a potent source for erythroid progenitors. During the time, from E12 to E17, HSCs can be found in circulation. From E15.5 HSCs colonize the bone marrow [181], the place where they will reside from here on and the rest of adult life. An overview of hematopoietic organs during development can be seen in figure 8B. Animal models are indispensible for studying hematopoiesis, although many aspects of hematopoiesis can also be recapitulated ***in vitro*** in cell culture. In vitro models have the advantage of being cheaper, less time consuming, more accessible and more flexible than ***in vivo*** approaches, but have the disadvantage that they lack the normal microenvironment (hematopoietic niche).

Hematopoiesis in a dish

A valuable asset of ES cells is that they can be used to differentiate toward somatic cell lineages ***in vitro***. This is done in general by withdrawing LIF from the medium. Culture conditions vary largely though, depending of what kind of somatic cell type is wanted. In general, ES cells will form large spheric structures upon

withdrawal of LIF. For these spheric structures the term embryoid body (EB) was coined [9; 182]. Several labs confirmed that these EBs contain all three germ cell layers and that they do resemble to some extent the post-implantation embryo [9; 10]. There are limits to this resemblance, as there is no body-plan nor polarity in EBs [183]. It is possible to push EB differentiation toward the hematopoietic lineage ***in vitro***, by adding ascorbic acid and holo-transferrin to the differentiation medium [184]. The stepwise specification of blood progenitor formation from mouse ES cells through the epiblast and mesoderm stages has been studied in detail [185]. This specification leads to a peak of Flk1+ cell production after 3.25 days of differentiation [186; 187]. These Flk1+ cells have endothelial and hematopoietic potential and can form smooth vascular muscle as well [166; 188]. They can form blast colonies (BL-CFC), and as such are the ***in vitro*** equivalent of the hemangioblast. After six days of EB differentiation, blood islands can be seen in the EB which are similar to yolk sac blood islands in the developing embryo [156; 189]. Definitive hematopoietic cells (but not HSCs) can be observed from day 8 onwards. An overview of differentiation of ES cells to the hematopoietic lineage is presented in figure 8C.

More recent work by the laboratory of George Lacaud has shown that the hemangioblast, hemogenic endothelium and endothelial to hematopoietic transition (EHT) can also be studied ***in vitro*** [190-192]. In these elegant models it has been shown that the TF RUNX1 is an essential regulator of EHT. There is ***in vivo*** evidence in the frog ***Xenopus*** that the BMP and fibroblast growth factor (FGF) signaling pathways act upstream of RUNX1 [193]. The TFs GFI1/GFI1b were identified as downstream targets of RUNX1 in mouse cells, but are not able to fully rescue the EHT phenotype of hemogenic endothelium in the absence of ***Runx1***. Expression of ***Gfi1/Gfi1b*** did allow cells with the correct morphology to bud out of the endothelium; however these cells did not have hematopoietic competence. For ***Gfi1/Gfi1b*** knock out hemogenic endothelium the story was reversed: cells with hematopoietic competence were present, but they did not acquire the correct morphology. From this study it was concluded that ***Gfi1/Gfi1b*** is required for cells to acquire the correct identity, but not hematopoietic competence during EHT [191]. This study readily demonstrates the importance and flexibility of ***in vitro*** studies on TFs in hematopoiesis.

Transcription factors in hematopoiesis

TFs are important regulators of hematopoietic differentiation, ***in vivo*** and ***in vitro***. Many TFs that play a role in this process have been identified and their functions have been studied in detail. These studies often involve knock out or knock down techniques as described earlier in this chapter.

Runx1 / Cbf-β

RUNX1 (AML1, CBF-α2) is a DNA-binding TF that is a member of the family of α core binding factors. Members of the α core binding factor family can interact with another TF, the core binding factor β (CBF-β). The consensus binding motif

of RUNX1 is YGYGGTY, where Y can be either T or C [195; 196]. RUNX1 can bind to DNA alone, but the affinity of binding to its motif is greatly enhanced (around 10-fold) when it associates with CBF-β [197]. It was stated before that RUNX1 is an essential regulator for EHT, priming several gene loci for expression. Binding of RUNX1 to these loci was shown by DamId [193]. After this initial activating activity, RUNX1 is mainly thought of as a repressor of transcription in macrophages [198] and megakaryocytes [199-203], by interacting with the co-repressor mSIN3a [204-206]. It is known that ***Runx1*** is expressed in erythroid cells and that the protein interacts with essential regulators of erythropoiesis like LDB1, GATA1 and TAL1 [149; 207]. More recently, it was shown that RUNX1 acts as both a repressor and an activator of transcription in erythroid cells, and that it interacts with both LSD1 and MyEF2 [208].

The ***Runx1*** knock out phenotype generated a lot of attention, as mice without ***Runx1*** die at E12 due to a complete lack of definitive hematopoiesis [208], while primitive hematopoiesis is only mildly affected [209]. In fact, ***Runx1*** knock out embryos fail to produce hematopoietic clusters in the dorsal aorta [210], a situation that is mimicked in the ***in vitro*** phenotype discussed earlier. Hematopoiesis in ***Cbf-β*** knock out mice is affected in the same way as in the ***Runx1*** knock out mice, stressing the importance of the interaction between RUNX1 and CBF-β proteins [211]. In an effort to better understand the role of RUNX1 in hematopoiesis, ***Runx1*** was conditionally knocked-out at different times during development. This led to the conclusion that ***Runx1*** is essential early in development (until E10.5, concurrent with the emergence of HSCs and EHT), but is dispensable afterwards.

Another important feature of the ***RUNX1*** gene is its frequent involvement in human acute myeloid leukemia (AML). Many translocations found in AML material involving the RUNX1 gene have been described. Most frequent is the 8q22 t(8;21) translocation [212; 213], generating a RUNX1-ETO fusion protein that contains the DNA-binding runt domain of RUNX1 and all conserved domains of the transcriptional repressor ETO. As a consequence, the fusion protein is able to bind to ***RUNX1*** target genes and favours the aberrant recruitment of co-repressors via ETO [214]. It has been shown that the RUNX1-ETO fusion protein blocks differentiation of myeloid blast cells, whilst it promotes self-renewal, thereby contributing to the AML disease phenotype.

Ldb1

The ubiquitously expressed LIM domain binding protein 1 (LDB1, also known as CLIM2 or NL1) was first identified in yeast-2-hybrid screens for its capacity to bind to LIM homeodomains and LIM only (LMO) proteins [215-217]. It was found to be an interaction partner of GATA1 in erythroid cells [148; 218]. LDB1 has no DNA binding capacity or enzymatic activity. LDB1 rather seems to function as a docking protein, by interacting with specific protein partners [219]. Through its N-terminal homodimerization domain it is able to form homodimers, potentially bringing two distant sites on the DNA together (figure 9, [220; 221]). In this way it plays a role in long-range DNA interactions, as was highlighted in recent studies [222; 223]. For a

detailed introduction on long-range interactions the reader is referred to chapter 2. LDB1 has been shown to play a central role in a core complex controlling erythroid cell fate. Among the partners of this complex are GATA1/2, TAL1, ETO2 and LMO2 [148]. In recent work it has been shown that this complex is required to stimulate transcription of genes required in the late stages of erythropoies, although it suppresses the expression of some genes in this stage as well [206]. Throughout this thesis this complex will be referred to as the LDB1-complex.

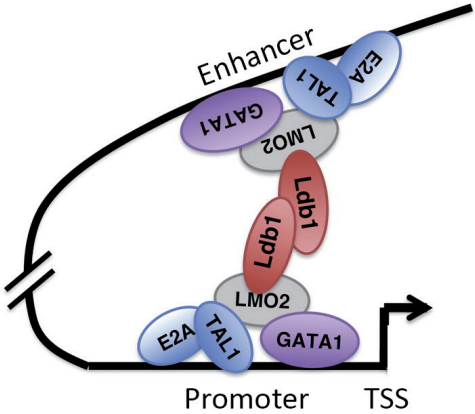


Figure 9. A theoretic model of long range chromatin interactions mediated by the LDB1 complex in late erythropoiesis. Two TF complexes bind distinct sites on the genome by factors able to directly interact with DNA (GATA1 and TAL1/E2A in this picture), LDB1 is part of this complex and indirectly interacts with the DNA binding factors. Through homodimerization of LDB1 the two complexes are linked and the two distinct genomic sites in close proximity. This process is called chromatin looping. Adapted from [222]. TSS = transcription start site

LDB1 is a highly conserved protein in eukaryotes and has a role many developmental processes in the mouse. This is exemplified by the ***Ldb1*** knock out mouse, which has a plethora of developmental defects [224]. Among them are patterning defects, defective neural development, aberrant development of the heart and a complete absence of hematopoiesis (primitive and definitive). More recently it was shown that ***Ldb1*** is required for hemangioblast function [225], HSC maintenance [226] and erythropoiesis [227]. The way LDB1 exerts its role is, as stated before, by binding to specific protein partners. The LMO proteins are important interaction partners for LDB1. This is exemplified by the knock out phenotypes of ***Lmo2*** and ***Lmo4***, which recapitulate the ***Ldb1*** knockout phenotype for blood [228; 229] and neural development [230] respectively. More recently, it was shown that LDB1 is an upstream regulator of hematopoietic/endothelial development [225].

Tal1

T-cell acute lymphocytic leukemia protein 1 (TAL1 or SCL) was first identified as a recurrent target of chromosomal translocations in T-cell acute lymphocytic

leukemia. Rearrangements causing leukemia often involve high expression of human TAL1 in differentiated T-cells, where TAL1 is normally suppressed [231]. TAL1 is a basic helix-loop-helix TF and binds so-called E-box motifs on the DNA. The E-box has the consensus motif CANNTG, where N can be any nucleotide. To efficiently bind DNA TAL1 needs to heterodimerize with another basic helix-loop-helix protein, such as TCF3, E2A, E2-2 or HEB [232]. In the LDB1-complex TAL1 cooperates with GATA1/2 to bind the DNA, resulting in a consensus motif for the LDB1-complex: (C)TGN$_{7\text{-}8}$WGATAR [206]. This motif contains a full GATA motif with half an E-box motif.

Different strategies for disrupting TAL1 protein function have been tried in order to elucidate the role of TAL1 in hematopoiesis. A complete ***Tal1*** knock out mouse generates neither primitive nor definitive blood and therefore dies at around E10.5 of severe anaemia [233]. ***Tal1*** knock out embryos show retarded growth and defective angiogenesis as well. ES cells that are homozygous knock out for ***Tal1*** are unable to contribute to blood in chimeric mice, but do contribute to other tissues [234]. EBs derived of ***Tal1*** knock out ES cells are devoid of hematopoietic potential, although Flk1+ cells were formed [188]. The differentiation potential to endothelial lineages was also affected. To further investigate TAL1 functions in hematopoiesis, conditional knock out studies were performed. They revealed that ***Tal1*** is not only important for the emergence of blood cells in the embryo, but also for maintenance of blood generation in the adult. Conditional knock out of ***Tal1*** in adult blood led to impaired differentiation of megakaryocytes and erythrocytes, as committed progenitors could not be generated [235], whereas other blood lineages seemed to be unaffected by ***Tal1*** knock out in adult blood. Further dissection of ***Tal1*** function in blood formation was performed by using cells with inducible TAL1 in cell culture. Using these ES cells it was shown that early ***Tal1*** expression, at the mesoderm stage (day 2-4 of EB differentiation), is essential for commitment to hematopoietic lineages later in development [236].

Myb

The myeloblastosis oncogene (***Myb***) encodes for the MYB TF. ***Myb*** is highly expressed in hematopoietic progenitors of all lineages (including HSCs), and is silenced upon differentiation toward the mature blood lineages [237; 238] (with the exception of megakaryocytes). This expression paradigm fits with a function for MYB in proliferation/differentiation decisions in progenitor blood cells. This is supported by the finding that human MYB is a protein misregulated in leukemia [239]. ***Myb*** knock out mice die at E15 because of an absence of definitive erythroid cells, although all blood lineages are affected [240]. Conditional knockout of the gene revealed roles of ***Myb*** in other blood cell types, mainly lymphoid [241; 242]. Expression of the ***Myb*** gene is regulated by the LDB1-complex through mediation of long-range promoter/enhancer interactions [223]. The expression of ***Myb*** and long-range chromatin interactions in its locus are topic of discussion in chapter 2 and 5.

Gata1 and *Gata2*

The GATA family of transcription factors consists in a group of six TFs containing a dual zinc-finger DNA-binding domain. They were named for the core nucleotide sequence that makes up their binding motif on the DNA (GATA). The complete motif is WGATAR or WGATAA [243]. The functional properties of GATA proteins are similar, as was shown in a study where GATA factors were swapped. The results of this study indicate that the dynamic regulation of Gata factors is more important than their identity [244]. GATA1, GATA2 and GATA3 have shown to play important roles in hematopoiesis, this discussion will focus on GATA1 and GATA2.

Knock out of the X-linked ***Gata1*** results in embryonic lethality between E10.5 and E11.5, as a result of severe anemia. Early yolk sac erythropoiesis is blocked in these animals [245; 246]. It was shown that these animals have fewer hematopoietic progenitor cells, and erythropoiesis is blocked at the pro-erythroblast stage. ***Gata1***-less ES cells can be differentiated toward the hematopoietic lineage and produce hemangioblasts at a normal rate, but erythropoiesis and megakaryopoiesis are blocked [247]. Further investigation of the ***Gata1*** knock out mouse showed that the megakaryocytic lineage is also affected in embryos [248]. Conditional knock out of ***Gata1*** in adult mice led to a decrease in the number of erythroid progenitors and a differentiation block at the pro-erythroblast stage [249]. This study indicates that ***Gata1*** may play a role in differentiation from early myeloid-erythroid progenitors to late erythropoiesis [250-252].

Gata2 knock out animals die between E10-11 due to severe anemia [253]. EBs derived from ***Gata2*** knock out ES cells produce hemangioblasts at a lower rate, as well as a reduced number of macrophages, very few mast cells and a significantly lower number of definitive and primitive erythrocytes [254; 255]. ***Gata2*** knock out ES cells cannot contribute to the hematopoietic lineage in chimaeric mice [255]. Moreover, it was shown that HSCs of ***Gata2*** haploinsufficient donors are significantly less able to reconstitute the hematopoietic system of sub-lethally irradiated mice [256]. Remarkably, high GATA2 levels were found to inhibit hematopoietic differentiation of human HSPCs (Hematopoietic Stem/Progenitor Cells) by influencing cell cycle [257]. These results combined point at the dose of GATA2 as being critical for correct functioning of HSPCs.

Gata1 and ***Gata2*** expression profiles are in concordance with their knockout phenotypes: ***Gata2*** is mainly expressed in HSPCs [258], but also in immature erythroid [259], megakaryocytes, mast and endothelial cells [259-261]. ***Gata1*** expression is turned on along erythrocytic specification [262], but is also found in committed hematopoietic progenitors, megakaryocytes, mast cells, eosinophils and dendritic cells [263-265]. Moreover, it was found that both factors can be part of the LDB1 complex [148; 266]. Combined, these findings suggest that a 'GATA-switch' may take place. This means that genomic sites that are occupied by GATA2 in early hematopoietic cells, may be occupied by GATA1 later during differentiation when ***Gata2*** expression is decreased [267]. It was shown that GATA2 is a far less stable protein than GATA1 [268] and that proteasome inhibition stabilizes GATA2,

blocking GATA-switching at least at several loci. GATA2 instability may account for rapid GATA2 loss during differentiation of HSCs, after which GATA1 can take GATA2's place in erythrocyte progenitors in TF complexes such as the LDB1-complex.

Other than the DNA motif itself, the chromatin state may be an important determinant of binding of TFs to the DNA. It is known that TFs are much less able to bind condensed chromatin [269]. However, not all GATA motifs in open accessible chromatin are bound. This suggests that additional motifs and/or co-factors are needed to target TFs to their cognate target sites in the genome. Indeed it was recently shown that GATA factor complexes binding to specific chromatin sites depend on the cofactors present in the complex [270].

Cofactors in transcription factor complexes

TF complexes need cofactors in order to be able to exert their function. For example, mass spectrometry of the individual members of GATA1 and LDB1-complexes in erythroid cells revealed the presence of histone modifying enzymes, chromatin remodelers and cell cycle regulators ([148; 271], and unpublished observations). It is hypothesized that these proteins are recruited to the genome along with the TF complex itself to i) influence the chromatin state leading to modified chromatin accessibility for TF binding and to activation or repression of transcription of genes, ii) maintain this chromatin state and iii) cross-talk between cellular functions (cell cycle, differentiation, apoptosis, etc.) to ensure proper regulation of these functions [272; 273].

Among the changes the cofactors promote are post-translational modifications (PTMs). PTMs are regulated changes of chemical properties of a specific amino acid within a protein by adding or removing a specific chemical compound to the amino acid. Examples of known PTMs that occur in mammalian cells are phosphorylation, acetylation, methylation or ubiquitination. There are many more known PTMs and new ones are still uncovered ([275], reviewed in [276]). Many specialized proteins that catalyze PTMs are identified in the human and mouse genome and many proteins that are subject to PTM have been identified. The function of PTMs is often providing a switch: regulating the specific affinity of a certain domain with its ligand or providing a platform for other domains to bind are examples of this [277]. It is appreciated that PTMs play a major role in the composition of TF complexes and determining their function [278], although the full extent of how PTMs influence cellular processes is far from being elucidated.

When the proteins present in a TF complex (TFs and cofactors with their associated cellular functions) as well as the target genes are known, this data can be combined to construct gene regulatory networks (GRNs). GRNs are blueprints of what is necessary to perform certain cellular functions. Many GRNs are presented in literature for developmental stages in different model systems, ranging from the reasonably simple to the immensely complex (e.g. [278-280]). All aim to provide a better molecular understanding of cellular functions.

Histone modifications

A special class of PTMs are the histone modifications. Histones are the core proteins around which the DNA is wrapped in the nucleus in order to achieve the first level of compaction needed to fit the $\approx 6.10^9$ base pairs of genomic DNA in the nucleus. Histones form octamers, with the N-termini protruding out of the complex. These N-termini are called the histone tails and they are heavily subjected to PTMs [282]. When first discovered, it was hypothesized that the state of PTMs on the histone tails would reflect the state of the chromatin [283]. The term histone code was coined for this hypothesis, in analogy with the genetic code. The histone code has been proven difficult to solve and to be ambiguous, but some rules have emerged from large-scale experimental data. Histone marks important for this work are discussed here, a more complete overview of histone marks that have been identified and chromatin states that have been (partly) ascribed to them is given in [282].

The histone octamer consists of two copies of each histone 1, histone 2A, histone 3 and histone 4 (H1, H2A, H3 and H4). The amino acids in the tails are referred to by their single letter code and the position within the tail, counted from the N-terminus. If a PTM is present it is referred to with the first two letters of the PTMs name and a possible number if more groups are present, e.g. H3K4me2 stands for di-methylation of the lysine at position 4 of the histone 3 tail. Condensed chromatin or heterochromatin is associated with H3K27me3 and H3K9me3 and an overall loss of histone acetylation. Moreover, DNA itself can be methylated in regions where the chromatin is in closed confirmation [269; 281; 283]. Open chromatin (euchromatin) is associated with high levels of histone acetylation, and regions of low density of nucleosomes and regions of active enhancers characterized by the presence of H3K4me1 and H3K27ac. Actively transcribed genes are marked by high levels of promoter H3K4me3, and the transcribed region is covered by the H3K36me3 and H3K79me2 marks highlighting productive transcriptional elongation [281; 283]. A special class are bivalent promoters, promoters of genes that are primed for active transcription, but not yet expressing. These are marked both H3K27me3 (repressive mark) and H3K4me3 (active mark) [284]. However the finding of bivalent promoters has been challenged as well, it has been suggested that the low H3K4me3 levels on H3K27me3-marked promoters is an artefact of suboptimal culture conditions [285; 286]. A comprehensive review on chromatin accessibility, histone modifications and transcriptional silencing/activation is given in [287].

References

1. Saiz, N., and Plusa, B. (2013). Early cell fate decisions in the mouse embryo. Reproduction ***145***, R65-80.
2. Kaufman, M.H., and Bard, J.B.L. (1999). The Anatomical Basis of Mouse Development (Academic Press).
3. Bazer, F.W., Spencer, T.E., ***et al.*** (2009). Comparative aspects of implantation. Reproduction ***138***, 195-209.
4. Chavatte-Palmer, P., and Guillomot, M. (2007). Comparative implantation and placentation. Gynecol Obstet Invest ***64***, 166-174.
5. Rossant, J., and Tam, P.P. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development ***136***, 701-713.
6. Zernicka-Goetz, M., Morris, S.A., ***et al.*** (2009). Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. Nat Rev Genet ***10***, 467-477.
7. Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature ***292***, 154-156.
8. Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A ***78***, 7634-7638.
9. Doetschman, T.C., Eistetter, H., ***et al.*** (1985). The ***in vitro*** development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. J Embryol Exp Morphol ***87***, 27-45.
10. Keller, G.M. (1995). In vitro differentiation of embryonic stem cells. Curr Opin Cell Biol 7, 862-869.
11. Bradley, A., Evans, M., ***et al.*** (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature ***309***, 255-256.
12. Nagy, A., Gocza, E., ***et al.*** (1990). Embryonic stem cells alone are able to support fetal development in the mouse. Development ***110***, 815-821.
13. Smith, A.G., Heath, J.K., ***et al.*** (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature ***336***, 688-690.
14. Williams, R.L., Hilton, D.J., ***et al.*** (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature ***336***, 684-687.
15. Ying, Q.L., Nichols, J., ***et al.*** (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell ***115***, 281-292.
16. Silva, J., Barrandon, O., ***et al.*** (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol ***6***, e253.
17. Leitch, H.G., Blair, K., ***et al.*** (2010). Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state. Development ***137***, 2279-2287.
18. Theunissen, T.W., van Oosten, A.L., ***et al.*** (2011). Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. Curr Biol ***21***, 65-71.
19. An, M.C., Zhang, N., ***et al.*** (2012). Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. Cell Stem Cell ***11***, 253-263.
20. Corti, S., Nizzardo, M., ***et al.*** (2012). Genetic correction of human induced pluripotent stem cells from patients with spinal muscular atrophy. Sci Transl Med ***4***, 165ra162.
21. Yusa, K., Rashid, S.T., ***et al.*** (2011). Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. Nature ***478***, 391-394.
22. Panopoulos, A.D., and Belmonte, J.C. (2012). Induced pluripotent stem cells in clinical hematology: potentials, progress, and remaining obstacles. Curr Opin Hematol ***19***, 256-260.
23. Ensembl (2013). Ensembl gene build 70.37, based on NCBI assembly Ch37.
24. International Human Genome Sequencing, C. (2004). Finishing the euchromatic sequence of the

human genome. Nature ***431***, 931-945.
25. Ensembl (2013). Ensembl gene build 70.38, based on NCBI assembly m38.
26. Mouse Genome Sequencing, C., Waterston, R.H., ***et al.*** (2002). Initial sequencing and comparative analysis of the mouse genome. Nature ***420***, 520-562.
27. Goodrich, J.A., and Kugel, J.F. (2009). From bacteria to humans, chromatin to elongation, and activation to repression: The expanding roles of noncoding RNAs in regulating transcription. Critical reviews in biochemistry and molecular biology ***44***, 3-15.
28. Kuehn, M.R., Bradley, A., ***et al.*** (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature ***326***, 295-298.
29. Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell ***51***, 503-512.
30. Mansour, S.L., Thomas, K.R., ***et al.*** (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature ***336***, 348-352.
31. Schwartzberg, P.L., Goff, S.P., ***et al.*** (1989). Germ-line transmission of a c-abl mutation produced by targeted gene disruption in ES cells. Science ***246***, 799-803.
32. McMahon, A.P., and Bradley, A. (1990). The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. Cell ***62***, 1073-1085.
33. Bouvier, G., Watrin, F., ***et al.*** (1996). Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. Proc Natl Acad Sci U S A ***93***, 7877-7881.
34. Gottgens, B., Broccardo, C., ***et al.*** (2004). The scl +18/19 stem cell enhancer is not required for hematopoiesis: identification of a 5' bifunctional hematopoietic-endothelial enhancer bound by Fli-1 and Elf-1. Mol Cell Biol ***24***, 1870-1883.
35. Wang, S., Aurora, A.B., ***et al.*** (2008). The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis. Dev Cell ***15***, 261-271.
36. Splinter, E., Heath, H., ***et al.*** (2006). ***CTCF*** mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20, 2349-2354.
37. Seitzer, N., Mayr, T., ***et al.*** (2010). A single nucleotide change in the mouse genome accelerates breast cancer progression. Cancer Res ***70***, 802-812.
38. Yagi, T., Nada, S., ***et al.*** (1993). A novel negative selection for homologous recombinants using diphtheria toxin A fragment gene. Anal Biochem ***214***, 77-86.
39. Yanagawa, Y., Kobayashi, T., ***et al.*** (1999). Enrichment and efficient screening of ES cells containing a targeted mutation: the use of DT-A gene with the polyadenylation signal as a negative selection maker. Transgenic Res ***8***, 215-221.
40. Skarnes, W.C., Rosen, B., ***et al.*** (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature ***474***, 337-342.
41. Copeland, N.G., Jenkins, N.A., ***et al.*** (2001). Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet ***2***, 769-779.
42. Zhang, Y., Buchholz, F., ***et al.*** (1998). A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet ***20***, 123-128.
43. Muyrers, J.P., Zhang, Y., ***et al.*** (2000). Point mutation of bacterial artificial chromosomes by ET recombination. EMBO Rep ***1***, 239-243.
44. Zhang, Y., Muyrers, J.P., ***et al.*** (2000). DNA cloning by homologous recombination in Escherichia coli. Nat Biotechnol ***18***, 1314-1317.
45. Muyrers, J.P., Zhang, Y., ***et al.*** (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res ***27***, 1555-1557.
46. Hofemeister, H., Ciotta, G., ***et al.*** (2011). Recombineering, transfection, Western, IP and ChIP methods

for protein tagging via gene targeting or BAC transgenesis. Methods ***53***, 437-452.
47. Osoegawa, K., Tateno, M., ***et al.*** (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome Res ***10***, 116-128.
48. Adams, D.J., Quail, M.A., ***et al.*** (2005). A genome-wide, end-sequenced 129Sv BAC library resource for targeting vector construction. Genomics ***86***, 753-758.
49. Coutinho, J.M., Singaraja, R.R., ***et al.*** (2005). Complete functional rescue of the ABCA1-/- mouse by human BAC transgenesis. J Lipid Res ***46***, 1113-1123.
50. Rostovskaya, M., Naumann, R., ***et al.*** (2013). Transposon mediated BAC transgenesis via pronuclear injection of mouse zygotes. Genesis ***51***, 135-141.
51. Rostovskaya, M., Fu, J., ***et al.*** (2012). Transposon-mediated BAC transgenesis in human ES cells. Nucleic Acids Res ***40***, e150.
52. Alberts, B., Johnson, A., ***et al.*** (2002). Molecular biology of the cell, 4th edn (New York: Garland science).
53. Ruby, K.M., and Zheng, B. (2009). Gene targeting in a HUES line of human embryonic stem cells via electroporation. Stem Cells ***27***, 1496-1506.
54. Paigen, K., and Petkov, P. (2010). Mammalian recombination hot spots: properties, control and evolution. Nat Rev Genet ***11***, 221-233.
55. Hasty, P., Rivera-Perez, J., ***et al.*** (1991). The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol ***11***, 5586-5591.
56. Lu , Z.H., Books, J.T., ***et al.*** (2003). Long targeting arms do not increase the efficiency of homologous recombination in the beta-globin locus of murine embryonic stem cells. Blood 102, 1531-1533.
57. te Riele, H., Maandag, E.R., ***et al.*** (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. Proc Natl Acad Sci U S A ***89***, 5128-5132.
58. Gossler, A., Joyner, A.L., ***et al.*** (1989). Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science ***244***, 463-465.
59. Skarnes, W.C., Auerbach, B.A., ***et al.*** (1992). A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes Dev ***6***, 903-918.
60. Joyner, A.L., Auerbach, A., ***et al.*** (1992). The gene trap approach in embryonic stem cells: the potential for genetic screens in mice. Ciba Found Symp ***165***, 277-288; discussion 288-297.
61. von Melchner, H., DeGregori, J.V., ***et al.*** (1992). Selective disruption of genes expressed in totipotent embryonal stem cells. Genes Dev ***6***, 919-927.
62. Voss, A.K., Thomas, T., ***et al.*** (1998). Compensation for a gene trap mutation in the murine microtubule-associated protein 4 locus by alternative polyadenylation and alternative splicing. Dev Dyn ***212***, 258-266.
63. McClive, P., Pall, G., ***et al.*** (1998). Gene trap integrations expressed in the developing heart: insertion site affects splicing of the PT1-ATG vector. Dev Dyn ***212***, 267-276.
64. Koller, B.H., Marrack, P., ***et al.*** (1990). Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8+ T cells. Science ***248***, 1227-1230.
65. Thomas, K.R., and Capecchi, M.R. (1990). Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. Nature ***346***, 847-850.
66. Donehower, L.A., Harvey, M., ***et al.*** (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature ***356***, 215-221.
67. Weeda, G., Donker, I., ***et al.*** (1997). Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. Curr Biol ***7***, 427-439.
68. Muller, U. (1999). Ten years of gene targeting: targeted mouse mutants, from vector design to

phenotype analysis. Mech Dev ***82***, 3-21.
69. Osokine, I., Hsu, R., ***et al.*** (2008). Unintentional miRNA ablation is a risk factor in gene knockout studies: a short report. PLoS Genet ***4***, e34.
70. Anastassiadis, K., Fu, J., ***et al.*** (2009). Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. Dis Model Mech ***2***, 508-515.
71. Dymecki, S.M. (1996). Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. Proc Natl Acad Sci U S A ***93***, 6191-6196.
72. Gu, H., Marth, J.D., ***et al.*** (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. Science ***265***, 103-106.
73. Kuhn, R., Schwenk, F., ***et al.*** (1995). Inducible gene targeting in mice. Science ***269***, 1427-1429.
74. Logie, C., and Stewart, A.F. (1995). Ligand-regulated site-specific recombination. Proc Natl Acad Sci U S A ***92***, 5940-5944.
75. Gu, H., Zou, Y.R., ***et al.*** (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Cell ***73***, 1155-1164.
76. Lee, G., and Saito, I. (1998). Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. Gene ***216***, 55-65.
77. Grindley, N.D., Whiteson, K.L., ***et al.*** (2006). Mechanisms of site-specific recombination. Annu Rev Biochem 75, 567-605.
78. Zaprazna, K., and Atchison, M.L. (2012). YY1 controls immunoglobulin class switch recombination and nuclear activation-induced deaminase levels. Mol Cell Biol ***32***, 1542-1554.
79. Morris, C.R., Stanton, M.J., ***et al.*** (2012). A knockout of the Tsg101 gene leads to decreased expression of ErbB receptor tyrosine kinases and induction of autophagy prior to cell death. PLoS One ***7***, e34308.
80. Urnov, F.D., Miller, J.C., ***et al.*** (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature ***435***, 646-651.
81. Bedell, V.M., Wang, Y., ***et al.*** (2012). In vivo genome editing using a high-efficiency TALEN system. Nature ***491***, 114-118.
82. Li, T., Huang, S., ***et al.*** (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res ***39***, 6315-6325.
83. Li, T., Huang, S., ***et al.*** (2011). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res ***39***, 359-372.
84. Miller, J.C., Tan, S., ***et al.*** (2011). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol ***29***, 143-148.
85. Sugisaki, H., and Kanazawa, S. (1981). New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). Gene ***16***, 73-78.
86. Wah, D.A., Hirsch, J.A., ***et al.*** (1997). Structure of the multimodular endonuclease FokI bound to DNA. Nature ***388***, 97-100.
87. Li, L., Wu, L.P., ***et al.*** (1992). Functional domains in Fok I restriction endonuclease. Proc Natl Acad Sci U S A ***89***, 4275-4279.
88. Skowron, P., Kaczorowski, T., ***et al.*** (1993). Atypical DNA-binding properties of class-IIS restriction endonucleases: evidence for recognition of the cognate sequence by a FokI monomer. Gene ***125***, 1-10.
89. Morton, J., Davis, M.W., ***et al.*** (2006). Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. Proc Natl Acad Sci U S A ***103***, 16370-16375.
90. Mussolino, C., Morbitzer, R., ***et al.*** (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res ***39***, 9283-9293.
91. Cathomen, T., and Joung, J.K. (2008). Zinc-finger nucleases: the next generation emerges. Mol Ther ***16***, 1200-1207.

92. Hockemeyer, D., Soldner, F., ***et al.*** (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol ***27***, 851-857.
93. Hockemeyer, D., Wang, H., ***et al.*** (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol ***29***, 731-734.
94. Bhaya, D., Davison, M., ***et al.*** (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet ***45***, 273-297.
95. Wiedenheft, B., Sternberg, S.H., ***et al.*** (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature ***482***, 331-338.
96. Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. Curr Opin Microbiol ***14***, 321-327.
97. Jinek, M., Chylinski, K., ***et al.*** (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science ***337***, 816-821.
98. Hwang, W.Y., Fu, Y., ***et al.*** (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol.
99. Cong, L., Ran, F.A., ***et al.*** (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823.
100. Jinek, M., East, A., ***et al.*** (2013). RNA-programmed genome editing in human cells. Elife ***2***, e00471.
101. Fire, A., Xu, S., ***et al.*** (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature ***391***, 806-811.
102. Tabara, H., Sarkissian, M., ***et al.*** (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell ***99***, 123-132.
103. Chuang, C.F., and Meyerowitz, E.M. (2000). Specific and heritable genetic interference by double-stranded RNA in Arabidopsis thaliana. Proc Natl Acad Sci U S A ***97***, 4985-4990.
104. Mourrain, P., Beclin, C., ***et al.*** (2000). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell ***101***, 533-542.
105. Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science ***286***, 950-952.
106. Elbashir, S.M., Harborth, J., ***et al.*** (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature ***411***, 494-498.
107. Kennerdell, J.R., and Carthew, R.W. (2000). Heritable gene silencing in Drosophila using double-stranded RNA. Nat Biotechnol ***18***, 896-898.
108. Zhou, Y., Ching, Y.P., ***et al.*** (2002). Post-transcriptional suppression of gene expression in ***Xenopus*** embryos by small interfering RNA. Nucleic Acids Res ***30***, 1664-1669.
109. Li, Y.X., Farrell, M.J., ***et al.*** (2000). Double-stranded RNA injection produces null phenotypes in zebrafish. Dev Biol ***217***, 394-405.
110. McCaffrey, A.P., Meuse, L., ***et al.*** (2002). RNA interference in adult mice. Nature ***418***, 38-39.
111. Drinnenberg, I.A., Weinberg, D.E., ***et al.*** (2009). RNAi in budding yeast. Science ***326***, 544-550.
112. Scannell, D.R., Frank, A.C., ***et al.*** (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Natl Acad Sci U S A ***104***, 8397-8402.
113. Wang, Z., Rao, D.D., ***et al.*** (2011). RNA interference and cancer therapy. Pharm Res ***28***, 2983-2995.
114. Lebbink, R.J., Lowe, M., ***et al.*** (2011). Polymerase II promoter strength determines efficacy of microRNA adapted shRNAs. PLoS One ***6***, e26213.
115. Giering, J.C., Grimm, D., ***et al.*** (2008). Expression of shRNA from a tissue-specific pol II promoter is an effective and safe RNAi therapeutic. Mol Ther ***16***, 1630-1636.
116. Brummelkamp, T.R., Bernards, R., ***et al.*** (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science ***296***, 550-553.

117. Siomi, H., and Siomi, M.C. (2009). On the road to reading the RNA-interference code. Nature ***457***, 396-404.
118. Zamore, P.D., Tuschl, T., ***et al.*** (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell ***101***, 25-33.
119. Vermeulen, A., Behlen, L., ***et al.*** (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. RNA ***11***, 674-682.
120. Ahlquist, P. (2002). RNA-dependent RNA polymerases, viruses, and RNA silencing. Science ***296***, 1270-1273.
121. Yang, D., Buchholz, F., ***et al.*** (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proc Natl Acad Sci U S A ***99***, 9942-9947.
122. ***Kit***tler, R., Putz, G., ***et al.*** (2004). An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. Nature ***432***, 1036-1040.
123. Karra, D., and Dahm, R. (2010). Transfection techniques for neuronal cells. J Neurosci ***30***, 6171-6177.
124. Finer, M.H., Dull, T.J., ***et al.*** (1994). kat: a high-efficiency retroviral transduction system for primary human T lymphocytes. Blood ***83***, 43-50.
125. Millington, M., Arndt, A., ***et al.*** (2009). Towards a clinically relevant lentiviral transduction protocol for primary human CD34 hematopoietic stem/progenitor cells. PLoS One ***4***, e6461.
126. Cornetta, K., Pollok, K.E., ***et al.*** (2008). Transduction of primary hematopoietic cells by retroviral vectors. CSH Protoc ***2008***, pdb prot4884.
127. Teramoto, S., Matsuse, T., ***et al.*** (1997). Kinetics of adenovirus-mediated gene transfer to human lung fibroblasts. Life Sci ***61***, 891-897.
128. Bernards, R., Brummelkamp, T.R., ***et al.*** (2006). shRNA libraries and their use in cancer genetics. Nat Methods ***3***, 701-706.
129. Yuan, P., Han, J., ***et al.*** (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev ***23***, 2507-2520.
130. Zhao, L., Glazov, E.A., ***et al.*** (2011). Integrated genome-wide chromatin occupancy and expression analyses identify key myeloid pro-differentiation transcription factors repressed by ***Myb***. Nucleic Acids Res ***39***, 4664-4679.
131. Haruki, H., Nishikawa, J., ***et al.*** (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell ***31***, 925-932.
132. Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol ***8***, 761-773.
133. Zemp, I., and Kutay, U. (2007). Nuclear export and cytoplasmic maturation of ribosomal subunits. FEBS Lett ***581***, 2783-2793.
134. Chen, J., Zheng, X.F., ***et al.*** (1995). Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue. Proc Natl Acad Sci U S A ***92***, 4947-4951.
135. Nishimura, K., Fukagawa, T., ***et al.*** (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods ***6***, 917-922.
136. Iwamoto, M., Bjorklund, T., ***et al.*** (2010). A general chemical method to regulate protein stability in the mammalian central nervous system. Chem Biol ***17***, 981-988.
137. Bonger, K.M., Chen, L.C., ***et al.*** (2011). Small-molecule displacement of a cryptic degron causes conditional protein degradation. Nat Chem Biol ***7***, 531-537.
138. Banaszynski, L.A., Chen, L.C., ***et al.*** (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell ***126***, 995-1004.
139. Neklesa, T.K., Tae, H.S., ***et al.*** (2011). Small-molecule hydrophobic tagging-induced degradation of

HaloTag fusion proteins. Nat Chem Biol ***7***, 538-543.
140. Holland, A.J., Fachinetti, D., ***et al.*** (2012). Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc Natl Acad Sci U S A ***109***, E3350-3357.
141. Weissman, A.M., Shabek, N., ***et al.*** (2011). The predator becomes the prey: regulating the ubiquitin system by ubiquitylation and degradation. Nat Rev Mol Cell Biol ***12***, 605-620.
142. Kanke, M., Nishimura, K., ***et al.*** (2011). Auxin-inducible protein depletion system in fission yeast. BMC Cell Biol ***12***, 8.
143. Stier, S., Cheng, T., ***et al.*** (2002). Notch1 activation increases hematopoietic stem cell self-renewal ***in vivo*** and favors lymphoid over myeloid lineage outcome. Blood ***99***, 2369-2378.
144. Reya, T., Duncan, A.W., ***et al.*** (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature ***423***, 409-414.
145. Blank, U., Karlsson, G., ***et al.*** (2008). Signaling pathways governing stem-cell fate. Blood ***111***, 492-503.
146. Zhang, J., Niu, C., ***et al.*** (2003). Identification of the haematopoietic stem cell niche and control of the niche size. Nature ***425***, 836-841.
147. Li, Z., and Li, L. (2006). Understanding hematopoietic stem-cell microenvironments. Trends Biochem Sci ***31***, 589-595.
148. Meier, N., Krpic, S., ***et al.*** (2006). Novel binding partners of ***Ldb1*** are required for haematopoietic development. Development ***133***, 4913-4923.
149. Grosveld, F., Rodriguez, P., ***et al.*** (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.
150. Goldman, J.M., and Melo, J.V. (2003). Chronic myeloid leukemia--advances in biology and new approaches to treatment. N Engl J Med ***349***, 1451-1464.
151. Horton, S.J., and Huntly, B.J. (2012). Recent advances in acute myeloid leukemia stem cell biology. Haematologica ***97***, 966-974.
152. Mullighan, C.G. (2012). Molecular genetics of B-precursor acute lymphoblastic leukemia. J Clin Invest ***122***, 3407-3415.
153. Van Vlierberghe, P., and Ferrando, A. (2012). The molecular basis of T cell acute lymphoblastic leukemia. J Clin Invest ***122***, 3398-3406.
154. Curtis, D.J., and McCormack, M.P. (2010). The molecular basis of ***Lmo2***-induced T-cell acute lymphoblastic leukemia. Clin Cancer Res ***16***, 5618-5623.
155. Gaidano, G., Foa, R., ***et al.*** (2012). Molecular pathogenesis of chronic lymphocytic leukemia. J Clin Invest ***122***, 3432-3438.
156. Palis, J., Robertson, S., ***et al.*** (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Development ***126***, 5073-5084.
157. Shalaby, F., Rossant, J., ***et al.*** (1995). Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. Nature ***376***, 62-66.
158. Sabin, F.R. (1920). Studies on the Origin of Blood-vessels and of Red Blood-corpuscles as Seen in the Living Blastoderm of Chicks During the Second Day of Incubation. Contrib Embryol 9, 215-262.
159. His, W. (1900). Lecithoblast und angioblast der wirbelthiere. Histogenetische studien. Abhandl KS Ges Wis Math-Phys ***22***, 171-328.
160. Jaffredo, T., Nottingham, W., ***et al.*** (2005). From hemangioblast to hematopoietic stem cell: an endothelial connection? Exp Hematol ***33***, 1029-1040.
161. Jaffredo, T., Bollerot, K., ***et al.*** (2005). Tracing the hemangioblast during embryogenesis: developmental relationships between endothelial and hematopoietic cells. Int J Dev Biol ***49***, 269-277.
162. Zambidis, E.T., Oberlin, E., ***et al.*** (2006). Blood-forming endothelium in human ontogeny: lessons from in utero development and embryonic stem cell culture. Trends Cardiovasc Med ***16***, 95-101.

163. North, T.E., de Bruijn, M.F., ***et al.*** (2002). ***Runx1*** expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. Immunity ***16***, 661-672.
164. Oberlin, E., Tavian, M., ***et al.*** (2002). Blood-forming potential of vascular endothelium in the human embryo. Development ***129***, 4147-4157.
165. Tavian, M., and Peault, B. (2005). The changing cellular environments of hematopoiesis in human development in utero. Exp Hematol ***33***, 1062-1069.
166. Yamashita, J., Itoh, H., ***et al.*** (2000). Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. Nature ***408***, 92-96.
167. Eichmann, A., Marcelle, C., ***et al.*** (1993). Two molecules related to the VEGF receptor are expressed in early endothelial cells during avian embryonic development. Mech Dev ***42***, 33-48.
168. Eichmann, A., Corbel, C., ***et al.*** (1997). Ligand-dependent development of the endothelial and hemopoietic lineages from embryonic mesodermal cells expressing vascular endothelial growth factor receptor 2. Proc Natl Acad Sci U S A ***94***, 5141-5146.
169. Choi, K., Kennedy, M., ***et al.*** (1998). A common precursor for hematopoietic and endothelial cells. Development ***125***, 725-732.
170. Dzierzak, E., and Speck, N.A. (2008). Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nat Immunol ***9***, 129-136.
171. Muller, A.M., Medvinsky, A., ***et al.*** (1994). Development of hematopoietic stem cell activity in the mouse embryo. Immunity ***1***, 291-301.
172. Medvinsky, A., and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. Cell ***86***, 897-906.
173. de Bruijn, M.F., Speck, N.A., ***et al.*** (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. EMBO J ***19***, 2465-2474.
174. Ottersbach, K., and Dzierzak, E. (2005). The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. Dev Cell ***8***, 377-387.
175. Taoudi, S., and Medvinsky, A. (2007). Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta. Proc Natl Acad Sci U S A ***104***, 9399-9403.
176. Kissa, K., and Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. Nature ***464***, 112-115.
177. Bertrand, J.Y., Chi, N.C., ***et al.*** (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. Nature ***464***, 108-111.
178. Boisset, J.C., van Cappellen, W., ***et al.*** (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. Nature ***464***, 116-120.
179. Medvinsky, A.L., Samoylina, N.L., ***et al.*** (1993). An early pre-liver intraembryonic source of CFU-S in the developing mouse. Nature ***364***, 64-67.
180. Cumano, A., and Godin, I. (2007). Ontogeny of the hematopoietic system. Annu Rev Immunol ***25***, 745-785.
181. Dzierzak, E., Medvinsky, A., ***et al.*** (1998). Qualitative and quantitative aspects of haematopoietic cell development in the mammalian embryo. Immunol Today ***19***, 228-236.
182. Martin, G.R., and Evans, M.J. (1975). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies ***in vitro***. Proc Natl Acad Sci U S A ***72***, 1441-1445.
183. Rippon, H.J., and Bishop, A.E. (2004). Embryonic stem cells. Cell Prolif ***37***, 23-34.
184. Carotta, S., Pilat, S., ***et al.*** (2004). Directed differentiation and mass cultivation of pure erythroid progenitors from mouse embryonic stem cells. Blood ***104***, 1873-1880.
185. Pearson, S., Sroczynska, P., ***et al.*** (2008). The stepwise specification of embryonic stem cells to hematopoietic fate is driven by sequential exposure to Bmp4, activin A, bFGF and VEGF. Development ***135***, 1525-1535.

186. Lacaud, G., Gore, L., ***et al.*** (2002). ***Runx1*** is essential for hematopoietic commitment at the hemangioblast stage of development ***in vitro***. Blood ***100***, 458-466.
187. Lacaud, G., Kouskoff, V., ***et al.*** (2004). Haploinsufficiency of ***Runx1*** results in the acceleration of mesodermal development and hemangioblast specification upon ***in vitro*** differentiation of ES cells. Blood ***103***, 886-889.
188. D'Souza, S.L., Elefanty, A.G., ***et al.*** (2005). SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development. Blood ***105***, 3862-3870.
189. Keller, G., Kennedy, M., ***et al.*** (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol Cell Biol ***13***, 473-486.
190. Lichtinger, M., Ingram, R., ***et al.*** (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. EMBO J ***31***, 4318-4333.
191. Lancrin, C., Mazan, M., ***et al.*** (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. Blood ***120***, 314-322.
192. Hoogenkamp, M., Lichtinger, M., ***et al.*** (2009). Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. Blood ***114***, 299-309.
193. Walmsley, M., Cleaver, D., ***et al.*** (2008). Fibroblast growth factor controls the timing of Scl, ***Lmo2***, and ***Runx1*** expression during embryonic blood development. Blood ***111***, 1157-1166.
194. Bravo, J., Li, Z., ***et al.*** (2001). The leukemia-associated AML1 (***Runx1***)--CBF beta complex functions as a DNA-induced molecular clamp. Nat Struct Biol ***8***, 371-378.
195. de Bruijn, M.F., and Speck, N.A. (2004). Core-binding factors in hematopoiesis and immune function. Oncogene ***23***, 4238-4248.
196. Ogawa, E., Inuzuka, M., ***et al.*** (1993). Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2 alpha. Virology ***194***, 314-331.
197. Hu, Z., Gu, X., ***et al.*** (2011). RUNX1 regulates corepressor interactions of PU.1. Blood 117, 6498-6508.
198. Elagib, K.E., Racke, F.K., ***et al.*** (2003). RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. Blood ***101***, 4333-4341.
199. Goldfarb, A.N. (2009). Megakaryocytic programming by a transcriptional regulatory loop: A circle connecting RUNX1, GATA-1, and P-TEFb. J Cell Biochem ***107***, 377-382.
200. Pencovich, N., Jaschek, R., ***et al.*** (2011). Dynamic combinatorial interactions of RUNX1 and cooperating partners regulates megakaryocytic differentiation in cell line models. Blood ***117***, e1-14.
201. Tijssen, M.R., Cvejic, A., ***et al.*** (2011). Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. Dev Cell ***20***, 597-609.
202. Xu, G., Kanezaki, R., ***et al.*** (2006). Physical association of the patient-specific GATA1 mutants with RUNX1 in acute megakaryoblastic leukemia accompanying Down syndrome. Leukemia ***20***, 1002-1008.
203. Zhao, X., Jankovic, V., ***et al.*** (2008). Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity. Genes Dev ***22***, 640-653.
204. Imai, Y., Kurokawa, M., ***et al.*** (2004). The corepressor mSin3A regulates phosphorylation-induced activation, intranuclear location, and stability of AML1. Mol Cell Biol ***24***, 1033-1043.
205. Bakker, W.J., Blazquez-Domingo, M., ***et al.*** (2004). FoxO3a regulates erythroid differentiation and induces BTG1, an activator of protein arginine methyl transferase 1. J Cell Biol ***164***, 175-184.
206. Soler, E., Andrieu-Soler, C., ***et al.*** (2010). The genome-wide dynamics of the binding of ***Ldb1*** complexes during erythroid differentiation. Genes Dev ***24***, 277-289.
207. van Riel, B., Pakozdi, T., ***et al.*** (2012). A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells. Mol Cell Biol ***32***, 3814-3822.

208. Wang, Q., Stacy, T., ***et al.*** (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc Natl Acad Sci U S A ***93***, 3444-3449.
209. Yokomizo, T., Hasegawa, K., ***et al.*** (2008). ***Runx1*** is involved in primitive erythropoiesis in the mouse. Blood ***111***, 4075-4080.
210. North, T., Gu, T.L., ***et al.*** (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. Development ***126***, 2563-2575.
211. Wang, Q., Stacy, T., ***et al.*** (1996). The CBFbeta subunit is essential for CBFalpha2 (AML1) function ***in vivo***. Cell ***87***, 697-708.
212. Miyoshi, H., Shimizu, K., ***et al.*** (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. Proc Natl Acad Sci U S A ***88***, 10431-10434.
213. Gao, J., Erickson, P., ***et al.*** (1991). Isolation of a yeast artificial chromosome spanning the 8;21 translocation breakpoint t(8;21)(q22;q22.3) in acute myelogenous leukemia. Proc Natl Acad Sci U S A ***88***, 4882-4886.
214. Elagib, K.E., and Goldfarb, A.N. (2007). Oncogenic pathways of AML1-ETO in acute myeloid leukemia: multifaceted manipulation of marrow maturation. Cancer Lett 251, 179-186.
215. Agulnick, A.D., Taira, M., ***et al.*** (1996). Interactions of the LIM-domain-binding factor ***Ldb1*** with LIM homeodomain proteins. Nature ***384***, 270-272.
216. Bach, I., Carriere, C., ***et al.*** (1997). A family of LIM domain-associated cofactors confer transcriptional synergism between LIM and Otx homeodomain proteins. Genes Dev ***11***, 1370-1380.
217. Jurata, L.W., Pfaff, S.L., ***et al.*** (1998). The nuclear LIM domain interactor NLI mediates homo- and heterodimerization of LIM domain transcription factors. J Biol Chem ***273***, 3152-3157.
218. Wadman, I.A., Osada, H., ***et al.*** (1997). The LIM-only protein ***Lmo2*** is a bridging molecule assembling an erythroid, DNA-binding complex which includes the ***TAL1***, E47, GATA-1 and ***Ldb1***/NLI proteins. EMBO J ***16***, 3145-3157.
219. Matthews, J.M., and Visvader, J.E. (2003). LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. EMBO Rep ***4***, 1132-1137.
220. Palstra, R.J., Tolhuis, B., ***et al.*** (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet ***35***, 190-194.
221. Song, S.H., Hou, C., ***et al.*** (2007). A positive role for NLI/***Ldb1*** in long-range beta-globin locus control region function. Mol Cell ***28***, 810-822.
222. Deng, W., Lee, J., ***et al.*** (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell ***149***, 1233-1244.
223. Stadhouders, R., Thongjuea, S., ***et al.*** (2012). Dynamic long-range chromatin interactions control ***Myb*** proto-oncogene transcription during erythroid development. EMBO J ***31***, 986-999.
224. Mukhopadhyay, M., Teufel, A., ***et al.*** (2003). Functional ablation of the mouse ***Ldb1*** gene results in severe patterning defects during gastrulation. Development ***130***, 495-505.
225. Mylona, A., Andrieu-Soler, C., ***et al.*** (2013). Genome-wide analysis shows that ***Ldb1*** controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis. Blood.
226. Li, L., Jothi, R., ***et al.*** (2011). Nuclear adaptor ***Ldb1*** regulates a transcriptional program essential for the maintenance of hematopoietic stem cells. Nat Immunol ***12***, 129-136.
227. Li, L., Lee, J.Y., ***et al.*** (2010). A requirement for Lim domain binding protein 1 in erythropoiesis. J Exp Med ***207***, 2543-2550.
228. Yamada, Y., Warren, A.J., ***et al.*** (1998). The T cell leukemia LIM protein ***Lmo2*** is necessary for adult mouse hematopoiesis. Proc Natl Acad Sci U S A ***95***, 3890-3895.
229. Warren, A.J., Colledge, W.H., ***et al.*** (1994). The oncogenic cysteine-rich LIM domain protein rbtn2 is

essential for erythroid development. Cell ***78***, 45-57.
230. Lee, S.K., Jurata, L.W., ***et al.*** (2005). The LIM domain-only protein LMO4 is required for neural tube closure. Mol Cell Neurosci ***28***, 205-214.
231. Begley, C.G., Aplan, P.D., ***et al.*** (1989). The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc Natl Acad Sci U S A ***86***, 10128-10132.
232. Murre, C., McCaw, P.S., ***et al.*** (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell ***56***, 777-783.
233. Robb, L., Lyons, I., ***et al.*** (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. Proc Natl Acad Sci U S A ***92***, 7075-7079.
234. Porcher, C., Swat, W., ***et al.*** (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell ***86***, 47-57.
235. Hall, M.A., Curtis, D.J., ***et al.*** (2003). The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. Proc Natl Acad Sci U S A ***100***, 992-997.
236. Endoh, M., Ogawa, M., ***et al.*** (2002). SCL/tal-1-dependent process determines a competence to select the definitive hematopoietic lineage prior to endothelial differentiation. EMBO J ***21***, 6700-6708.
237. Gonda, T.J., and Metcalf, D. (1984). Expression of ***myb***, myc and fos proto-oncogenes during the differentiation of a murine myeloid leukaemia. Nature ***310***, 249-251.
238. Emambokus, N., Vegiopoulos, A., ***et al.*** (2003). Progression through key stages of haemopoiesis is dependent on distinct threshold levels of c-***Myb***. EMBO J ***22***, 4478-4488.
239. Ramsay, R.G., and Gonda, T.J. (2008). ***MYB*** function in normal and cancer cells. Nat Rev Cancer ***8***, 523-534.
240. Mucenski, M.L., McLain, K., ***et al.*** (1991). A functional c-***myb*** gene is required for normal murine fetal hepatic hematopoiesis. Cell ***65***, 677-689.
241. Bender, T.P., Thompson, C.B., ***et al.*** (1987). Differential expression of c-***myb*** mRNA in murine B lymphomas by a block to transcription elongation. Science ***237***, 1473-1476.
242. Thomas, M.D., Kremer, C.S., ***et al.*** (2005). c-***Myb*** is critical for B cell development and maintenance of follicular B cells. Immunity ***23***, 275-286.
243. Ko, L.J., and Engel, J.D. (1993). DNA-binding specificities of the GATA transcription factor family. Mol Cell Biol ***13***, 4011-4022.
244. Ferreira, R., Wai, A., ***et al.*** (2007). Dynamic regulation of Gata factor levels is more important than their identity. Blood ***109***, 5481-5490.
245. Fujiwara, Y., Browne, C.P., ***et al.*** (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proc Natl Acad Sci U S A ***93***, 12355-12358.
246. Pevny, L., Simon, M.C., ***et al.*** (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature ***349***, 257-260.
247. Weiss, M.J., Keller, G., ***et al.*** (1994). Novel insights into erythroid development revealed through ***in vitro*** differentiation of GATA-1 embryonic stem cells. Genes Dev ***8***, 1184-1197.
248. Stachura, D.L., Chou, S.T., ***et al.*** (2006). Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. Blood ***107***, 87-97.
249. Gutierrez, L., Tsukamoto, S., ***et al.*** (2008). Ablation of ***Gata1*** in adult mice results in aplastic crisis, revealing its essential role in steady-state and stress erythropoiesis. Blood ***111***, 4375-4385.
250. Gutierrez, L., Lindeboom, F., ***et al.*** (2004). Homotypic signalling regulates ***Gata1*** activity in the erythroblastic island. Development ***131***, 3183-3193.
251. Whyatt, D., Lindeboom, F., ***et al.*** (2000). An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. Nature ***406***, 519-524.

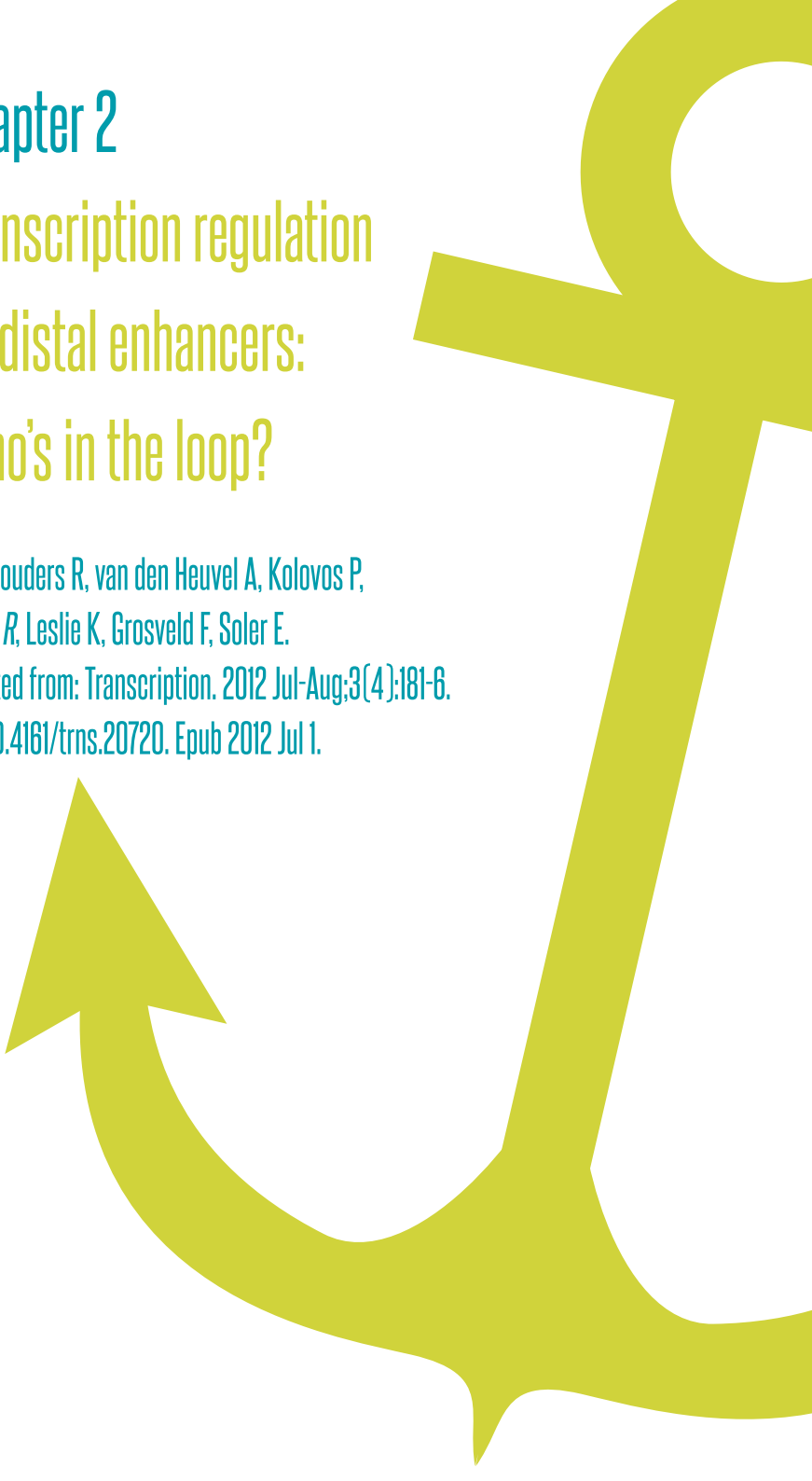
252. Whyatt, D.J., Karis, A., ***et al.*** (1997). The level of the tissue-specific factor GATA-1 affects the cell-cycle machinery. Genes and function ***1***, 11-24.
253. Tsai, F.Y., Keller, G., ***et al.*** (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature ***371***, 221-226.
254. Lugus, J.J., Chung, Y.S., ***et al.*** (2007). GATA2 functions at multiple steps in hemangioblast development and differentiation. Development ***134***, 393-405.
255. Tsai, F.Y., and Orkin, S.H. (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood ***89***, 3636-3643.
256. Rodrigues, N.P., Janzen, V., ***et al.*** (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. Blood ***106***, 477-484.
257. Tipping, A.J., Pina, C., ***et al.*** (2009). High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. Blood ***113***, 2661-2672.
258. Orlic, D., Anderson, S., ***et al.*** (1995). Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45 NF-E2, and c-***myb*** and low levels or no mRNA for c-fms and the receptors for granulocyte colony-stimulating factor and interleukins 5 and 7. Proc Natl Acad Sci U S A ***92***, 4601-4605.
259. Orkin, S.H. (1992). GATA-binding transcription factors in hematopoietic cells. Blood ***80***, 575-581.
260. Jippo, T., Mizuno, H., ***et al.*** (1996). Abundant expression of transcription factor GATA-2 in proliferating but not in differentiated mast cells in tissues of mice: demonstration by in situ hybridization. Blood ***87***, 993-998.
261. Lee, M.E., Temizer, D.H., ***et al.*** (1991). Cloning of the GATA-binding protein that regulates endothelin-1 gene expression in endothelial cells. J Biol Chem ***266***, 16188-16192.
262. Leonard, M., Brice, M., ***et al.*** (1993). Dynamics of GATA transcription factor expression during erythroid differentiation. Blood ***82***, 1071-1079.
263. Shivdasani, R.A., and Orkin, S.H. (1996). The transcriptional control of hematopoiesis. Blood ***87***, 4025-4039.
264. Martin, D.I., Zon, L.I., ***et al.*** (1990). Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. Nature ***344***, 444-447.
265. Zon, L.I., Yamaguchi, Y., ***et al.*** (1993). Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription. Blood ***81***, 3234-3241.
266. Lecuyer, E., Herblot, S., ***et al.*** (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood ***100***, 2430-2440.
267. Bresnick, E.H., Lee, H.Y., ***et al.*** (2010). GATA switches as developmental drivers. J Biol Chem ***285***, 31087-31093.
268. Lurie, L.J., Boyer, M.E., ***et al.*** (2008). Differential GATA factor stabilities: implications for chromatin occupancy by structurally similar transcription factors. Biochemistry ***47***, 859-869.
269. Schotta, G., Lachner, M., ***et al.*** (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev ***18***, 1251-1262.
270. Chlon, T.M., Dore, L.C., ***et al.*** (2012). Cofactor-mediated restriction of GATA-1 chromatin occupancy coordinates lineage-specific gene expression. Mol Cell ***47***, 608-621.
271. Rodriguez, P., Bonte, E., ***et al.*** (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. EMBO J ***24***, 2354-2366.
272. Narlikar, G.J., Fan, H.Y., ***et al.*** (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell ***108***, 475-487.
273. Krebs, J.E., Kuo, M.H., ***et al.*** (1999). Cell cycle-regulated histone acetylation required for expression of the yeast HO gene. Genes Dev ***13***, 1412-1421.

274. Tan, M., Luo, H., ***et al.*** (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell ***146***, 1016-1028.
275. Tootle, T.L., and Rebay, I. (2005). Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. Bioessays ***27***, 285-298.
276. Ahearn, I.M., Haigis, K., ***et al.*** (2012). Regulating the regulator: post-translational modification of RAS. Nat Rev Mol Cell Biol ***13***, 39-51.
277. Everett, L., Hansen, M., ***et al.*** (2010). Regulating the regulators: modulators of transcription factor activity. Methods Mol Biol ***674***, 297-312.
278. Loose, M., and Patient, R. (2004). A genetic regulatory network for ***Xenopus*** mesendoderm formation. Dev Biol ***271***, 467-478.
279. Rual, J.F., Venkatesan, K., ***et al.*** (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature ***437***, 1173-1178.
280. Li, S., Armstrong, C.M., ***et al.*** (2004). A map of the interactome network of the metazoan C. elegans. Science ***303***, 540-543.
281. Kouzarides, T. (2007). Chromatin modifications and their function. Cell ***128***, 693-705.
282. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature ***403***, 41-45.
283. Zhou, V.W., Goren, A., ***et al.*** (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet ***12***, 7-18.
284. Bernstein, B.E., Kamal, M., ***et al.*** (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell ***120***, 169-181.
285. Hong, S.H., Rampalli, S., ***et al.*** (2011). Cell fate potential of human pluripotent stem cells is encoded by histone modifications. Cell Stem Cell ***9***, 24-36.
286. Marks, H., Kalkan, T., ***et al.*** (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell ***149***, 590-604.
287. Bell, O., Tiwari, V.K., ***et al.*** (2011). Determinants and dynamics of genome accessibility. Nat Rev Genet ***12***, 554-564.
288. Mocellin, S., and Provenzano, M. (2004). RNA interference: learning gene knock-down from cell physiology. J Transl Med ***2***, 39.
289. Chute, J.P., Ross, J.R., ***et al.*** (2010). Minireview: Nuclear receptors, hematopoiesis, and stem cells. Mol Endocrinol ***24***, 1-10.
290. Christensen, J.L., Wright, D.E., ***et al.*** (2004). Circulation and chemotaxis of fetal hematopoietic stem cells. PLoS Biol ***2***, E75.
291. Abranches, E., Bekman, E., ***et al.*** (2013). Generation and characterization of a novel mouse embryonic stem cell line with a dynamic reporter of Nanog expression. PLoS One ***8***, e59928.

Chapter 2

Transcription regulation by distal enhancers: Who's in the loop?

Stadhouders R, van den Heuvel A, Kolovos P,
Jorna R, Leslie K, Grosveld F, Soler E.
Adapted from: Transcription. 2012 Jul-Aug;3(4):181-6.
doi: 10.4161/trns.20720. Epub 2012 Jul 1.



Abstract

Genome-wide chromatin profiling efforts have shown that enhancers are often located at large distances from gene promoters within the non-coding genome. Whereas enhancers can stimulate transcription initiation by communicating with promoters via chromatin looping mechanisms, we propose that enhancers may also stimulate transcription elongation by physical interactions with intronic elements. We review here recent findings derived from the study of the hematopoietic system.

Introduction

The development of multicellular organisms relies upon the capacity of stem and progenitor cells to respond to their microenvironment and differentiate upon exposure to specific stimuli. This multi-step process involves complex epigenetic changes within regulatory transcriptional networks, which contribute to the timely activation and repression of key developmental genes. Transcription of mammalian genes relies on the presence of a variety of ***cis***-DNA regulatory sequences such as promoters and enhancers. Whereas it is relatively easy to identify gene promoters (i.e., at the 5' end of transcriptional units), locating and characterizing enhancers is more complicated. Transgenic experiments carried out over the last few decades have taught us important lessons about transcriptional enhancers and genomic organization. Attempts to express transgenes in animals, under the control of endogenous promoter sequences, often resulted in weak expression, altered tissue specificity, and frequent transcriptional silencing after stable integration of the transgene into the genome. Efficient transgene expression (maintaining developmental transcription dynamics and expression levels) required the use of large genomic DNA sequences. Besides promoter elements, these large DNA fragments contained introns and sequences surrounding the genes. It was deduced that natural sequences surrounding the genes containing tissue-specific transcriptional enhancers were essential for its proper expression, because promoters work in combination with additional regulatory sequences that may be remote from transcription start sites (TSS). The ***de novo*** identification of transcriptional enhancers is difficult because they show not only a great variety in localization with respect to their target genes but also in sequence composition. The recent advances in high throughput sequencing technologies, such as ChIP Sequencing (ChIP-Seq), have allowed chromatin structure and transcription factor occupancy to be analyzed on a genome-wide scale. These techniques have resulted in a recent redefinition of enhancers as discrete genomic sites harboring a local combination of open chromatin structure (hypersensitivity to DNAse I), specific covalent histone modifications like mono- and di-methylation of histone 3 lysine 4 (H3K4me1, H3K4me2), acetylation of H3K27, low levels of H3K4me3, and occupancy of RNA polymerase II (RNA Pol II) and transcription factors (TFs) [1-3]. Based on this (epigenetic) definition, thousands of potential enhancers were localized genome-wide, some of which have been functionally validated ***in vivo*** to

show that they fulfill the old criteria of the definition of an enhancer as a sequence that enhances the transcription of a gene [4]. Nevertheless, this redefinition of enhancers may not be sufficient to predict enhancer function in a spatio-temporal fashion during development. Furthermore, with the recent identification of 67 novel types of histone modifications [5], the typical enhancer signature is likely to evolve and reveal a high degree of complexity and diversity. Presently, the best way to identify these critical regulatory elements is by combining histone modification profiling with the binding of general and tissue-specific TFs.

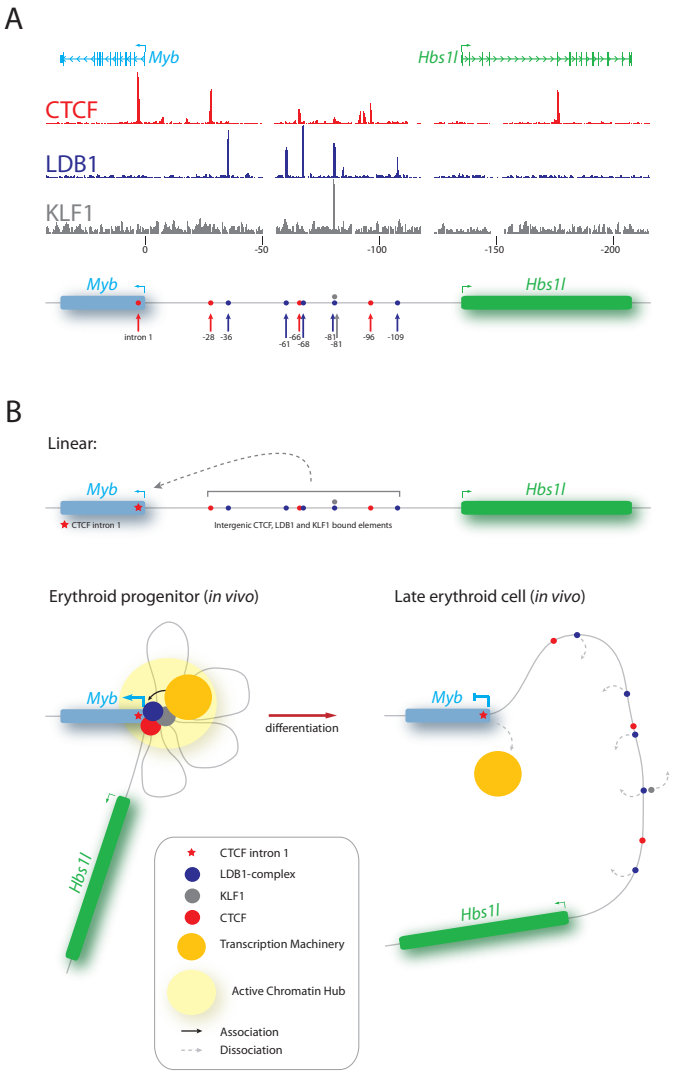
The Genome is Big: Why Not Just Use All This Space?

ChIP-Seq-based studies have determined the genome-wide binding sites of TFs with unprecedented ease and speed. A highly complex picture of transcriptional regulatory networks (TRNs) emerges from the data obtained. Textbook examples of TF binding at promoters are in fact exceptional cases, with several studies showing that critical tissue-specific TFs bind at large distances from genes TSS either in intergenic regions or within introns. For example, the binding profiles of the essential hematopoietic factors GATA1, TAL1, LDB1 and RUNX1 show that at least 90% of binding events are not at the promoter but intronic or intragenic [207; 296]. Binding sites are often localized several dozen to hundreds of kbp from the nearest TSS, indicative of long-range transcriptional regulation. In addition, it appears that developmentally regulated genes may harbor multiple binding sites for the same TF complexes, raising the possibility that TF-bound regulatory elements may act in a cooperative and/ or specialized fashion during development [6-9], underscoring the complexity of TRNs. Similar findings were reported for non-hematopoietic tissues [10], indicating that transcriptional regulatory elements are generally located in the non-coding fraction of mammalian genomes. This is not surprising in light of the necessity to retain the coding and function of a gene during evolution. Importantly, enhancer location is clearly not restricted to the immediate vicinity of their cognate target genes as they may be found upstream, downstream or within genes. Long-range transcriptional regulation by distal enhancers hence emerges as an important mechanism driving proper spatio-temporal regulation of gene expression during development. In agreement with this observation, examination of genome-wide association studies suggests that mutations in non protein-coding genomic regions contribute to disease traits in a significant number of cases [11]. For example, single nucleotide polymorphisms (SNPs) affecting the severity of the erythroid disorders beta-thalassemia and sickle cell anemia were found within the ***HBS1L-MYB*** and ***BCL11A*** loci [12]. The causative SNPs fall into intergenic and intronic regions, respectively, up to 80 kb away from the gene promoters. An intronic SNP within the ***HERC2*** gene has recently been linked to the regulation of the downstream ***OCA2*** gene which is involved in human pigmentation [13]. One of the most extreme examples is the location of an enhancer 1 Mbp upstream from the Sonic Hedgehog (***SHH***) gene, within an intron of the unrelated ***LMBR1*** gene [14]. SNPs were found in this region in humans and shown to affect spatio-temporal ***SHH*** expression resulting in the congenital abnormality preaxial polydactyly, one of the most frequently observed

hand malformations [15]. These intriguing findings reveal the incredible functional complexity of the non-coding genome where intergenic, intronic and even gene desert areas [16; 17] have the potential to play critical roles in gene regulatory networks both in development and disease. This raises the question of how mammalian genome organization relates to transcriptional regulation, and how this organization dynamically changes during cellular differentiation to allow distal enhancers to regulate transcription over large distances ***in vivo***.

Long-range Transcription Regulation by Chromatin Looping

New insights derived from ChIP-Seq analyses have provided a very detailed view of the regulatory potential of the genome although it is restrained to a linear perspective. Functional genomics studies are now facing the challenge of linking distal enhancers to their cognate genes, functionally dissecting enhancer-gene relationships, and understanding the impact of non-coding sequence variations in disease. The current dominant model for long-range transcriptional regulation proposes that distal enhancers are brought into physical proximity to their target genes in the three-dimensional nuclear space by chromatin looping mechanisms [18; 19]. The analysis of such spatial organization has been made possible thanks to the development of Chromosome Conformation Capture (3C) technology [20] and its high throughput derivatives (4C, 5C, 3C-Seq, HiC) [21]. 3C allows measuring the interaction frequency between two distal DNA elements and thereby provides information about local genomic topology and chromatin looping. 3C was originally used to study chromosome conformation in yeast [20] and the regulation of the ***ß-globin*** gene cluster by distal regulatory elements during erythroid development in mice [22]. We recently developed 3C-Seq technology which couples chromosome conformation capture to high throughput sequencing [6; 9]. 3C-Seq measures interaction frequencies between a viewpoint (a DNA fragment of choice, e.g., gene promoter) and (distal) regulatory elements on a genome-wide scale. We used 3C-Seq for the unbiased analysis of the spatial organization of the ***Myb*** proto-oncogene locus in murine erythroid cells [9]. ***Myb*** is a critical hematopoietic regulator required for the proliferation and expansion of all blood progenitors, and is dramatically down regulated during terminal differentiation. Failure to silence ***Myb*** expression is linked to impaired differentiation and may play a key role in leukemogenesis [23]. We showed that ***Myb*** transcription is regulated by an array of distal intergenic enhancers localizing up to 109 kb upstream of the gene. The enhancers are occupied by the essential hematopoietic TFs GATA1, TAL1, and LDB1. 3C-Seq profiling revealed that the enhancers loop to the ***Myb*** gene when it is transcriptionally active, forming an active chromatin hub resembling the one detected on the ***ß-globin*** locus. Importantly, the spatial organization of the locus is highly dynamic. During terminal differentiation the active chromatin hub is destabilized and the enhancers no longer loop to the ***Myb*** gene, a feature correlating with a loss of TF occupancy at the distal sites, and a loss of transcriptional activity of the locus (figure 1) [9]. This and earlier studies suggest that dynamic chromatin looping and changes in spatial organization represent important features within gene regulatory networks [9; 15; 22; 24-26].



Stadhouders et al. - Figure 1

Figure 1. Transcription factor occupancy and three-dimensional structure of the ***Myb*** locus. (A) ChIP-Seq profiles of CTCF (red), LDB1 (blue) and KLF1 (grey) at the ***Myb-Hbs1l*** intergenic region. A schematic of the area including all TF-binding sites and their distance relative to ***Myb*** TSS is shown. (B) Spatial organization of the ***Myb*** locus in erythroid cells. On top, a linear schematic of the locus is shown (as in A), with the looping events towards the promoter summarized by the gray arrow. Below, the actual model of the three-dimensional conformation of the locus ***in vivo*** is shown, for both erythroid progenitors (expressing ***Myb***) and differentiated erythroid cells (silencing ***Myb*** expression). TSS = transcription start site.

Chromatin Loop Formation and Maintenance

The mechanisms by which specific chromatin loops are established, maintained or lost remain unclear. Furthermore, whether chromatin looping is a cause or a consequence of gene activity remains a matter of debate, although some data has been generated suggesting that chromatin looping can occur before a gene is expressed [27]. However, it is clear that chromatin loops depend on the local binding of structural and regulatory transcription factors. Structural proteins such as CTCF and Cohesin have been shown to participate in three-dimensional genomic interactions [28-30]. For instance, both CTCF and Cohesin were shown to be crucial for imprinting at the H19/IGF2 locus [31], a locus subjected to long-range regulation by differential looping. It is worth noting that CTCF is also well known for its enhancer-blocking function and, as such, can limit the range of activity of nearby enhancers [28; 32]. Within the immunoglobulin κ light chain locus (Igκ), conditional inactivation of the ***CTCF*** gene in pre-B cells results in increased usage of the proximal Vκ-3 gene family, which is rarely used in normal B cells. Increased Vκ-3 genes usage correlates with increased interaction between the Igκ locus enhancers and the Vκ-3 genes in the absence of ***CTCF***, suggesting that CTCF drives the specificity of enhancer-genes contacts at the Igκ locus [29]. The absence of ***CTCF*** has been linked to disruption of loop formation at several other developmentally regulated loci. For instance, targeted disruption of a CTCF binding motif in the ***ß-globin*** locus 3'HS1 element, abolishing CTCF binding, disrupts local loop formation [30]. The Cohesin complex has also been linked to higher order chromatin structure formation and/or maintenance and it was shown that depletion of the Cohesin complex subunit SMC1 resulted in reduced enhancer-promoter loop formation at the Nanog locus in ES cells [33]. A further elegant study using HiC reveals that depletion of CTCF or Cohesion both lead to a loss of overall 3D configuration of the genome in mouse ES cells [personal communication K.Wendt]. In addition, TFs were also shown to play a role in long-range gene regulation, e.g., the hematopoietic TFs LDB1, GATA1, FOG1, KLF1 and BCL11a are required to maintain chromatin looping within the ***ß-globin***, ***Myb*** and other loci [9; 24; 34-37]. Differential enhancer-gene looping correlating with gene expression was also observed at the ***Kit*** oncogene locus. ***Kit*** expression in hematopoietic progenitors is controlled by a distal enhancer -114 kb upstream of the gene, which is occupied by GATA2 TF complexes and loops to the ***Kit*** gene when transcriptionally active. At the onset of terminal differentiation, the GATA2 complexes are replaced by GATA1-nucleated complexes, correlating with a spatial reorganization of the locus, a modification of enhancer-gene interactions, and a loss of ***Kit*** expression [24]. These findings emphasize that complex interplay between regulatory factors binding to distal enhancers takes place during development, and suggest that the dynamic and timely establishment of higher order chromatin structures is involved in establishing and maintaining transcriptional regulatory networks.

Regulation of Transcriptional Elongation by Distal Enhancers

Despite detailed information from a number of model loci [9; 22; 26; 28], higher order chromatin structure and local genomic reorganization upon signaling remain poorly understood or even completely uncharacterized for the vast majority of genes. Importantly, the functional relationship between distal enhancer-gene interactions and transcriptional activity is still a matter of debate. The prevalent model is that distal enhancers loop to target gene promoters where they stimulate transcription by providing an increased local concentration of positive acting factors [28; 33; 38]. However, this model does not apply to all cases. Sometimes distal enhancers show a preferential interaction with the transcribed part of their target genes (e.g., at intronic sites) rather than at promoter regions [9; 24; 39]. These observations raise questions regarding the functionality of such enhancer-gene contacts. We recently showed that the -81 kb ***Myb*** enhancer preferentially associates with the first intron of the gene. This region is strongly occupied by CTCF and was previously shown to harbor an 'attenuator' site regulating transcription elongation [23]. Accordingly, we demonstrated that this region represents the site where RNA Pol II switches from the initiating to the elongating form, as characterized by phosphorylation of serine (Ser) residues 5 and 2. The appearance of transcription elongation-associated chromatin marks (e.g., H3K36me3) also occurs just downstream of the intronic CTCF site [9]. However, both this site and the ***Myb*** promoter harbor only minor quantities of the positive elongation factor CDK9, a kinase involved in the phosphorylation of RNA Pol II Ser 2, which regulates transcription elongation. Instead, strong enrichments of CDK9 and an additional positive elongation factor, TIF1γ, were found at the upstream regulatory sites, including the -81 kb enhancer. We proposed a model where RNA Pol II stalls at ***Myb*** intron 1, close to the CTCF site, and requires stimulatory activity from the distal enhancers to bypass the attenuator element. Interestingly, when erythroid cells were treated with the CDK9 kinase inhibitor DRB to inhibit transcriptional elongation, distal enhancers still looped to ***Myb*** intron 1. This suggests that the loops became non-functional due to their inability to provide kinase activity. Intriguingly, our unpublished observations suggest that this mechanism also operates at other developmentally regulated genes in erythroid cells (van den Heuvel, Kolovos ***et al.*** unpublished). Furthermore, previous experiments have shown that the ***ß-globin*** LCR controls high level globin transcription primarily through a stimulatory effect on transcription elongation [40]. Similar to the ***Myb*** upstream regulatory elements, the LCR was highly enriched for positive elongation factors, while proximal promoter sequences showed less binding of these factors [41]. Together, these data suggest that the function of at least a subclass of distal enhancers may be to provide direct local stimulation of transcription elongation (figure 2). In support of this view, a recent genome-wide histone modification profiling study, performed in differentiating erythroid cells, suggested that the regulation of transcription elongation plays a key role in gene induction and repression processes during cellular differentiation [42]. Future investigations will reveal whether direct transcription elongation stimulation by enhancers is a general mechanism.

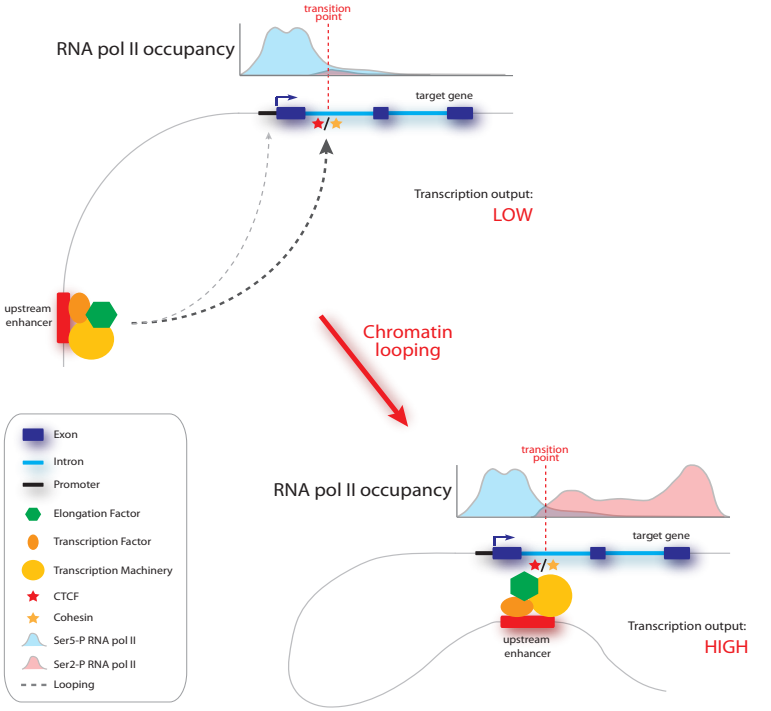


Figure 2. Speculative model of enhancer-mediated long-range stimulation of transcription elongation. The upper half shows a model gene with an upstream enhancer occupied by transcription factors, elongations factors and the transcription machinery. In the absence of chromatin looping, expression of the gene is kept low due to inefficient transcriptional elongation. Enhancer looping towards the gene results in the stimulation of elongation by increased RNA Pol II Ser 2 phosphorylation and high level gene expression. Structural factors involved in chromatin looping (i.e., CTCF and/or Cohesin, depicted by star symbols) possibly contribute to establishing local enhancer-gene communication.

Important Technical Challenges and Remaining Questions

Since 3C-based technologies only provide topological information, their functional relevance should be interpreted with caution and needs to be supported by additional experiments. These experiments typically aim at correlating gene expression and TF occupancy with chromatin looping dynamics but assessing the functionality of a looping event remains a difficult task. One way to address this question is to generate mutant alleles and conditional enhancer deletions to address their roles ***in vivo***, and to selectively disrupt specific loop formation [8; 30]. In the case of genes controlled by multiple regulatory elements (e.g., ***Myb***), this will

show whether transcriptional activity directly depends on all the active regulatory elements or whether there are specific elements and/or subsets driving stage-specific high level expression. Studies performed to address these questions for the human β-globin locus include the insertion of a human β-globin mini-locus in the mouse [38] and targeted disruption of the LCR in a cell line [43]. Despite the availability of high throughput recombineering technologies and novel ways to engineer genomes, such approaches remain laborious and time consuming. Recombineering and genome engineering are discussed in chapter one. It remains a challenge to obtain a broader picture of genomic architecture with sufficient resolution to visualize individual enhancer-gene contacts, and explore the correlation with gene transcription. Several new technological developments have provided new possibilities with approaches such as HiC allowing the capture of the genomic "interactome" or ChIA-PET, which allows the detection of genome-wide loop formation nucleated by specific transcription factors (for review see ref. [20]). Defining the nuclear architecture, its dependence on regulatory factors and its impact on gene expression remains an important challenge in the field of functional genomics. However, it is likely to highlight key features that will provide a superior understanding of the regulatory role of the non-coding genome. One of the major challenges in this field may be to decipher the mechanism by which long-range interactions can switch a stalled to an elongating form of polymerase.

Concluding Remarks

The current genome-wide characterizations of enhancers provide a picture of increasing complexity and diversity in both enhancer structure and function [28; 44]. We expect that different classes of enhancers will fulfill specific functions, such as facilitating transcriptional pause release or enhancing transcription elongation. The presence of multiple enhancers at single gene loci suggests that subsets of functionally specialized enhancers may provide a means to precisely drive transcription during specific developmental windows within specific lineages. Analyzing transcriptional regulation in both time and space emphasized the highly dynamic nature of the genome, which has recently been compared to a "regulatory jungle" [17], bearing "regulatory archipelagos" [16]. The rules governing the genomic regulatory landscape in its incredible complexity are only now just being discovered. Understanding the interplay between distal enhancers, their target genes, and their individual roles within complex genetic loci will remain a major task both in basic and disease-driven research.

References

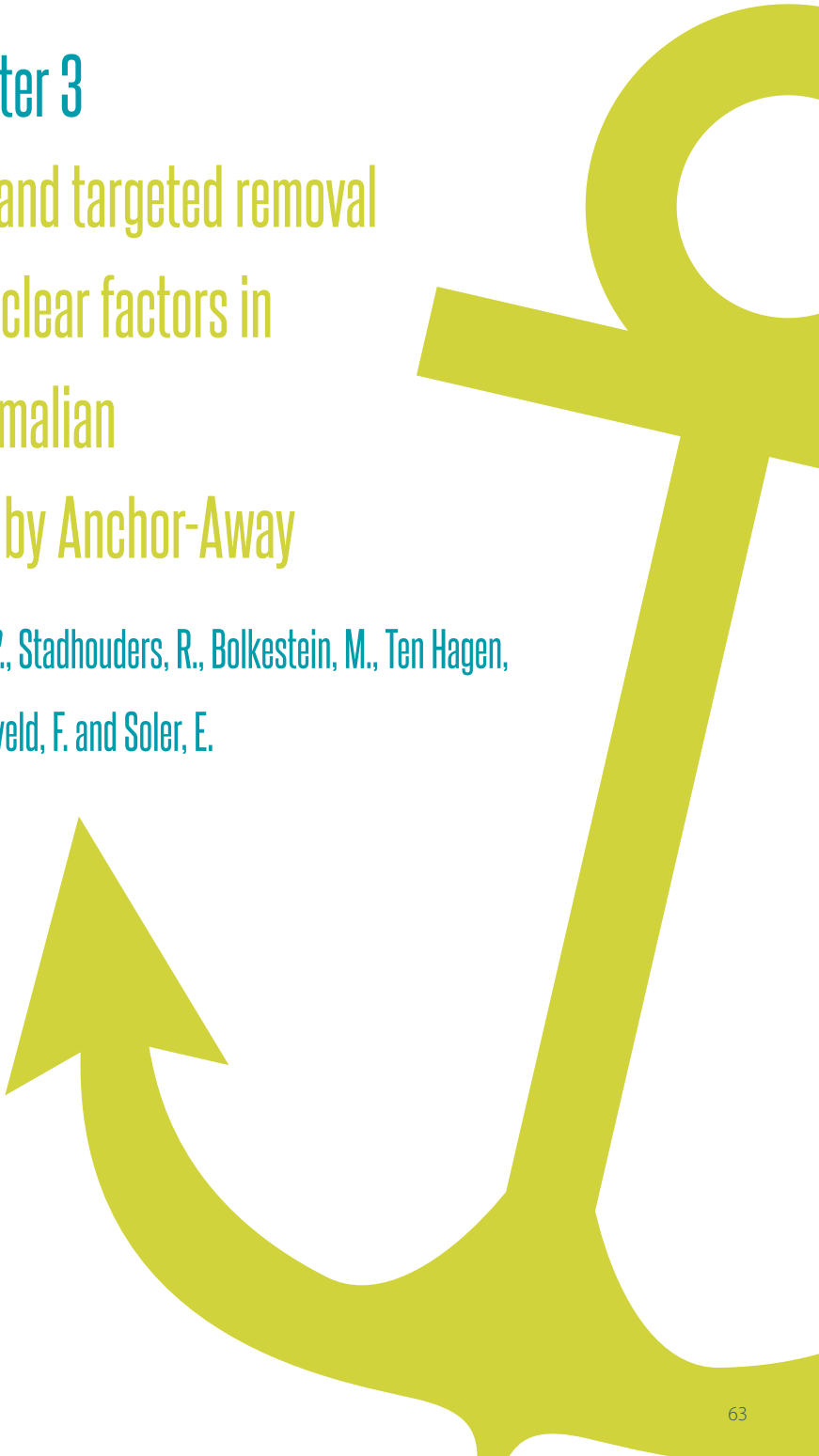
1. Pekowska, A., Benoukraf, T., ***et al.*** (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. EMBO J ***30***, 4198-4210.
2. Heintzman, N.D., Hon, G.C., ***et al.*** (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature ***459***, 108-112.
3. Koch, F., and Andrau, J.C. (2011). Initiating RNA polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity. Transcription ***2***, 263-268.
4. Visel, A., Blow, M.J., ***et al.*** (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature ***457***, 854-858.
5. Tan, M., Luo, H., ***et al.*** (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell ***146***, 1016-1028.
6. Soler, E., Andrieu-Soler, C., ***et al.*** (2010). The genome-wide dynamics of the binding of ***Ldb1*** complexes during erythroid differentiation. Genes Dev ***24***, 277-289.
7. Wilson, N.K., Foster, S.D., ***et al.*** (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell ***7***, 532-544.
8. Snow, J.W., Trowbridge, J.J., ***et al.*** (2010). A single cis element maintains repression of the key developmental regulator ***Gata2***. PLoS Genet ***6***.
9. Stadhouders, R., Thongjuea, S., ***et al.*** (2012). Dynamic long-range chromatin interactions control ***Myb*** proto-oncogene transcription during erythroid development. EMBO J ***31***, 986-999.
10. Schmidt, D., Wilson, M.D., ***et al.*** (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science ***328***, 1036-1040.
11. Visel, A., Rubin, E.M., ***et al.*** (2009). Genomic views of distant-acting enhancers. Nature ***461***, 199-205.
12. Thein, S.L., Menzel, S., ***et al.*** (2009). Control of fetal hemoglobin: new insights emerging from genomics and clinical implications. Hum Mol Genet ***18***, R216-223.
13. Visser, M., Kayser, M., ***et al.*** (2012). ***HERC2*** rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the ***OCA2*** promoter. Genome Res ***22***, 446-455.
14. Sagai, T., Hosoya, M., ***et al.*** (2005). Elimination of a long-range ***cis***-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development ***132***, 797-803.
15. Lettice, L.A., Heaney, S.J., ***et al.*** (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet ***12***, 1725-1735.
16. Montavon, T., Soshnikova, N., ***et al.*** (2011). A regulatory archipelago controls Hox genes transcription in digits. Cell ***147***, 1132-1145.
17. Ruf, S., Symmons, O., ***et al.*** (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat Genet ***43***, 379-386.
18. Hanscombe, O., Whyatt, D., ***et al.*** (1991). Importance of globin gene order for correct developmental expression. Genes Dev ***5***, 1387-1394.
19. Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature ***322***, 697-701.
20. Dekker, J., Rippe, K., ***et al.*** (2002). Capturing chromosome conformation. Science ***295***, 1306-1311.
21. de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev ***26***, 11-24.
22. Palstra, R.J., Tolhuis, B., ***et al.*** (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet ***35***, 190-194.
23. Ramsay, R.G., and Gonda, T.J. (2008). ***MYB*** function in normal and cancer cells. Nat Rev Cancer ***8***, 523-534.
24. Jing, H., Vakoc, C.R., ***et al.*** (2008). Exchange of GATA factors mediates transitions in looped chromatin

organization at a developmentally regulated gene locus. Mol Cell ***29***, 232-242.
25. Kolovos, P., Knoch, T.A., ***et al.*** (2012). Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin ***5***, 1.
26. Noordermeer, D., Leleu, M., ***et al.*** (2011). The dynamic architecture of Hox gene clusters. Science ***334***, 222-225.
27. Deng, W., Lee, J., ***et al.*** (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell ***149***, 1233-1244.
28. Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet ***12***, 283-293.
29. Ribeiro de Almeida, C., Stadhouders, R., ***et al.*** (2011). The DNA-binding protein ***CTCF*** limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus. Immunity ***35***, 501-513.
30. Splinter, E., Heath, H., ***et al.*** (2006). ***CTCF*** mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev ***20***, 2349-2354.
31. Wendt, K.S., Yoshida, K., ***et al.*** (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature ***451***, 796-801.
32. Ribeiro de Almeida, C., Stadhouders, R., ***et al.*** (2012). DNA-binding factor ***CTCF*** and long-range gene interactions in V(D)J recombination and oncogene activation. Blood ***119***, 6209-6218.
33. Kagey, M.H., Newman, J.J., ***et al.*** (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature ***467***, 430-435.
34. Drissen, R., Palstra, R.J., ***et al.*** (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev ***18***, 2485-2490.
35. Song, S.H., Hou, C., ***et al.*** (2007). A positive role for NLI/***Ldb1*** in long-range beta-globin locus control region function. Mol Cell ***28***, 810-822.
36. Vakoc, C.R., Letting, D.L., ***et al.*** (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell ***17***, 453-462.
37. Xu, J., Sankaran, V.G., ***et al.*** (2010). Transcriptional silencing of {gamma}-globin by ***BCL11A*** involves long-range interactions and cooperation with SOX6. Genes Dev ***24***, 783-798.
38. Grosveld, F., van Assendelft, G.B., ***et al.*** (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell ***51***, 975-985.
39. Krumm, A., Hickey, L.B., ***et al.*** (1995). Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. Genes Dev ***9***, 559-572.
40. Sawado, T., Halow, J., ***et al.*** (2003). The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. Genes Dev ***17***, 1009-1018.
41. Song, S.H., Kim, A., ***et al.*** (2010). Multiple functions of ***Ldb1*** required for beta-globin activation during erythroid differentiation. Blood ***116***, 2356-2364.
42. Wong, P., Hattangadi, S.M., ***et al.*** (2011). Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. Blood ***118***, e128-138.
43. Kim, C.G., Epner, E.M., ***et al.*** (1992). Inactivation of the human beta-globin gene by targeted insertion into the beta-globin locus control region. Genes Dev ***6***, 928-938.
44. Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell ***144***, 327-339.

Chapter 3

Fast and targeted removal of nuclear factors in mammalian cells by Anchor-Away

Jorna, R., Stadhouders, R., Bolkestein, M., Ten Hagen, T., Grosveld, F. and Soler, E.



Abstract

The slow kinetics of RNAi knock down or site-specific recombinase mediated conditional knock out in mammalian cells obscure the primary effects of the depletion of factors by these methods. Therefore, it would be advantageous to quickly inactivate or remove the target protein rather than the specific mRNA or genes. Here we apply the Anchor-Away technique [132] to mammalian cells. We show that Anchor-Away is capable of depleting factors from the nuclei of mammalian cells within four hours. Depletion is mediated by an anchor ribosomal protein (rpL13) fused to the FKBP12 domain and the target nuclear factor (LDB1 in this work) fused to the FRB* domain. We show that addition of AP21967 (rapalog) induces ternary complex formation of the two domains and causes the target nuclear factor to be pulled from the nucleus to the cytoplasm by the ribosomal protein. We conclude that Anchor-Away is an important new technique to deplete nuclear proteins in mammalian cells and perform functional studies.

Introduction

In order to investigate target genes regulated by a transcription factor (TF) it is common practice that the TF of interest is either knocked down using RNAi or (conditionally) knocked out by gene targeting [207; 247; 328; 329]. Although these strategies have been very successful in generating functional data for many TFs, they nevertheless have their drawbacks. This is illustrated by data showing that typically fewer than 30% of the genes that are misregulated after knock down/out of a specific TF are also bound by the TF as determined by massive parallel sequencing after chromatin immune-precipitation (e.g. [130; 131]). A potential contributor to the low overlap are indirect, secondary effects. The problem is caused by timing: enough of the target protein has to be degraded by the proteasome in order to obtain a good knock down/out. The time this takes is dependent protein half-life and this can vary by orders of magnitude. In general one has to wait for at least 48 hours, and as a result the cells spend a considerable time in an environment where the targeted protein is "slowly" disappearing providing time for secondary effects to take place. Thus during incubation time, genes primarily regulated by the target genes will be differentially regulated, but these primary regulated genes will also exert their influence on their transcriptional targets. These are the secondary effects of the knockdown.

A RNAi knock down specific drawback is that the interfering RNA should be specific and not cause non-specific effects, hence usually more than one shRNA needs to be used to obtain a reliable result. Moreover not all shRNAs result in the attended knock down and a number of different ones (3-5) need to be tested. In addition cells that are transfected with shRNA are not synchronous and in many cell types it is very difficult to deliver the shRNA to all the cells, causing the knock down to be incomplete. Although this last problem may be overcome by using

lentivirusses to transduce cells with shRNA constructs. Similarly a conditional knock out takes time to take effect, is usually not synchronous in all cells of a population and often incomplete. Secondary effects would be lowered if gene expression profiling could be performed much faster after the initiation of a knock down/out, preferably by removing the target protein directly instead of (part of) its gene or mRNA.

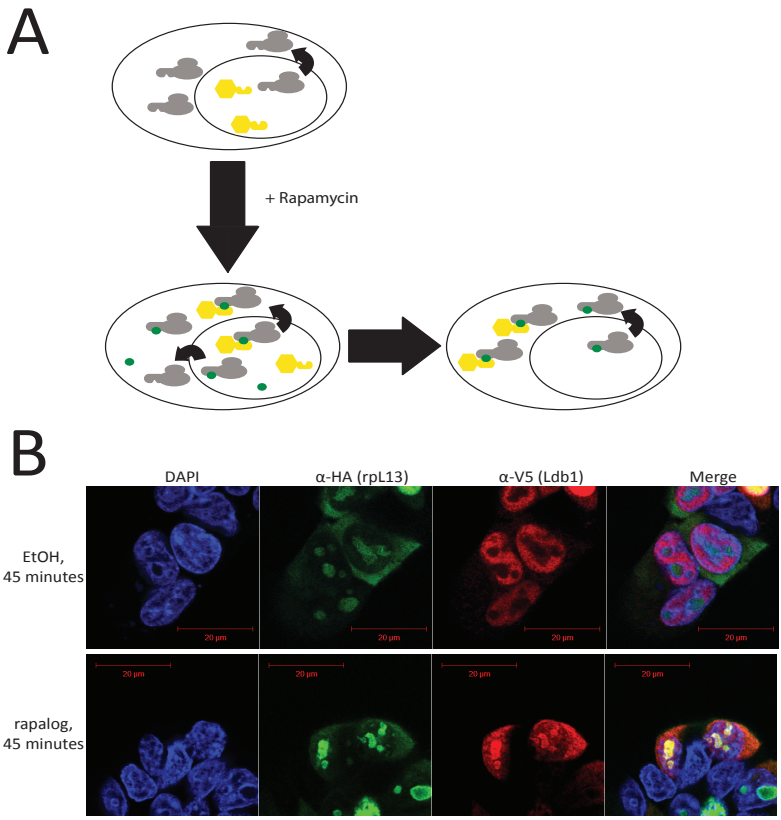


Figure 1. A) Schematic view of AA in yeast as presented by Haruki et al [1]. Both target nuclear protein (yellow) and an anchor ribosomal protein (incorporated in ribosome, in grey) are tagged with an AA tag. Upon addition of rapamycin (green) ternary complex formation takes place between the tagged proteins and rapamycin. In time, the cell's nucleus is cleared from the target protein. In mammalian cells the same mechanism is applied, although rapalog is used instead of rapamycin and a mammalian homolog of the ribosomal protein (rpL13 / rpL7a) is used.

B) Immunofluorescence images of AA in HEK293T cells. The transcription factor LDB1 is in red, the nuclear DAPI stain in blue and in green the ribosomal protein rpL13. The upper panel depicts a control situation, 45 minutes incubation with EtOH added to the medium. The lower panel shows pictures of cells 45 minutes after rapalog addition to the medium. The pictures shown are representative pictures of at least five different experiments, for those cells in the population that are double transfectants (rpL13-AA, LDB1-AA).

A new approach for disturbing target nuclear proteins in yeast was presented by Haruki et al. in 2008 [1]. Their work shows that a target nuclear protein could be sequestered to the cytoplasm within the hour. Furthermore, it was shown that sequestration of nuclear factors to the cytoplasm using Anchor-Away (AA) led to growth phenotypes similar to those observed after RNAi knock down, suggesting AA can be used to obtain functional data. AA works through fusion of each of two different partner rapamycin binding domains to a specific ribosomal protein (anchor) and the target nuclear protein that is to be removed. Upon addition of rapamycin strong ternary complex formation takes place between the ribosomal protein and the nuclear protein. The ribosomal protein is incorporated in a maturing ribosomal subunit that is transported in a large flow from the nucleoli to the cytoplasm, through the nucleus. The nuclear protein is dragged along (figure 1A) to the cytoplasm and remains there.

Here we present an adaptation of the original AA system for use in mammalian cells. In the original AA method, rapamycin was used as the inducing agent in rapamycin-resistant yeast strains [1]. Rapamycin toxicity in mammalian cells is well documented and rapamycin has a clinical application as an anti-cancer agent (e.g. recently reviewed in [8]). Toxicity of rapamycin is mediated through binding of the protein FK-binding protein 12 (FKBP12) and subsequent ternary complex formation with mammalian target of rapamycin (mTOR) protein. The ternary binding is mediated by the FRB domain of mTOR [9; 10]. An alternative to rapamycin is a homolog of rapamycin, AP21967, which will be referred to as rapalog from here on. Rapalog has no toxic effects on mammalian cells at low concentrations [11], because it has very low affinity for the original FRB domain. Rapalog does have a high affinity for the FKBP12 and a stable, mutated form of the FRB domain (K2095P, W2101F), the FRB* domain [12]. Upon binding of rapalog to FKBP12 an interaction surface is created for the FRB* domain to interact with both, establishing a strong ternary complex with nanomolar dissociation constant [10]. This inducible heterodimerization system has been used for other applications, such as identification of nuclear export sequences of proteins [334] and forced gene expression [332].

Here we report that AA functions in human HEK293T cells, mouse ES cells and cells differentiated toward the hematopoietic lineage from mouse ES cells. Target transcription factors can be sequestered to the cytoplasm within four hours in all cell types that were investigated. This time frame is one order of magnitude shorter than RNA interference and is therefore better suited to identify the direct effect of the functional removal of a target protein. AA works through addition of rapalog to the cell culture medium for HEK293T cells and mouse ES cells. Rapalog was unable to enter differentiated cells. This blockage was overcome by incorporating the rapalog in cationic liposomes that were subsequently fused with the differentiated cells.

Materials and Methods

Constructs

Original constructs of the AA paper [1] were a kind gift of Ulrich Laemmli's laboratory. Constructs with FRB* and FKBP12 domains, as well as rapalog were obtained from ARIAD (now ClonTech, Cat. 635067). cDNAs of transcription factors and ribosomal proteins were cloned into pBUD and pPyCAG vectors, respectively. The DNA coding for the FRB* domain and V5 tag were ligated onto the transcription factor cDNAs and the DNA coding for the FKBP12 domain and HA tag onto the ribosomal protein cDNA. To obtain a BAC tagged version of ***Ldb1***, recombineering on the 3' end of ***Ldb1*** was performed in the BAC RP23-106G13. The DNA for recombineering was obtained by a PCR on the constructs made for tagging TFs in HEK293T cells with the following primers: Forward 5'CATCCAGCCAAGAGAGCAAATCGGAGAATCCCACGTCACAGGCTTCCCAGGCC AGCACCAAGGGCCC 3' and Reverse 5'GAGCTGTGAGGGGTAGGCAGGCAGAGC GCTGGGTGGCCGTGGTAGGGCCTGGATCCCCTCGAGGGACC 3'. Recombineering plasmids were a kind gift of Francis Stewart's laboratory in Dresden.

Cell culture

HEK293T cells were cultured in standard culture medium (89% DMEM, 10% FCS, 1X Pen / Strep) and were transfected using lipofectamine. HEK293T cells were passaged every other day.

Mouse ES cells were cultured in 83% KO-DMEM (Gibco, Cat. 10829-018), 15% FCS, 1X Pen/Strep and 1X Non-essential amino acids (Invitrogen, Cat. 11140-050). Leukemia inhibitory factor (LIF, 10^3 U/ml) and 0.1mM 2-mercaptoethanol were added to the medium. For selection 200 µg/mL of G418 (Sigma, Cat. A1720) or 1.5 µg /mL puromycin (Sigma, Cat. P8833) was added. Medium was refreshed every day and cells were passaged every other day. Cells were transfected using lipofectamine according to manufacturer's protocol. For establishing stable ES cell clones, selection was started 24 hours after transfection for about 7 days and medium-sized, round colonies were picked. Rapalog was kept at -20°C in EtOH at a 1mM concentration and added to a final concentration of 5µM in the culture medium.

Differentiation of ES cells and rapalog incubation

ES cells were grown in suspension at 10.000 cells/ml on nonadherent dishes in IMDM medium (without LIF) with 15% FCS, 5% protein free hybridoma medium II (Gibco, Cat. 12040-077), 1% P/S, 1% Glutamax (Gibco, Cat. 25030-08), 0.05 µg/ml transferrin (Roche, Cat. 652-202), 0.05 µg/ml ascorbic acid (Sigma, Cat. A4544) and 1.8µl/ml monothioglycerol (Sigma, Cat. M6145). Embryoid bodies (EBs) were disrupted using trypsin-EDTA at day 3.25 of differentiation. EB cells were incubated for <4h with 5µM rapalog or EtOH and allowed to sediment on poly-prep slides (Sigma, Cat. P0425-72EA) after which the immunofluorescence protocol could be started.

Liposome Preparation

Cationic liposomes were prepared by a lipid film hydration and extrusion method [14]. Liposomes containing a mixture of egg phosphatidylcholine (EPC; Lipoid

GmbH, Ludwigshaven, Germany), cholesterol (Sigma-Aldrich, St. Louis, MO, USA), 1,2-dioleoyl-3-trimethylammonium-propane (chloride salt) (DOTAP; Avanti Polar lipids, Alabaster, AL, USA) in a molar ratio of 7:1:2 was dissolved in chloroform/methanol 9:1 (v/v). After addition of 0.1 mol% 1,2-dipalmitoyl-***sn***-glycero-3-phosphoethanolamine-N-(lissamine rhodamine B sulfonyl) (ammonium salt) (Rhodamine-PE; Avanti Polar lipids, Alabaster, AL, USA) the solvent was evaporated in vacuo using a rotary evaporator (Büchi Rotavapor R-210, Büchi Labortechnik, Flawil, Switzerland) until a homogeneous lipid film was formed. The film was hydrated in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffered saline (10 mM HEPES, 135 mM NaCl, pH adjusted to 7.4). The resulting vesicles were extruded at 60°C through a high-pressure Lipex thermoline extruder (Northern Lipids Inc, Vancouver, Canada) by passing through a Nucleopore polycarbonate membrane filter (Whatman, Newton, MA, USA) with pore diameters of 200, 100, 80, and 50 nm (5 extrusions per filter).

Film loading method [15]

AP21967 (ARIAD Pharmaceuticals, Inc, Cambridge, MA, USA) was dissolved in ethanol, which was evaporated with a stream of nitrogen gas. Empty liposomes were added to the rapalog film in a ratio 60:1 (total lipid: AP21967) and the mixture was sonicated for 10 min at 60°C in a sonication bath [16]. The suspension was left for 1 hour at room temperature, after which the free drug was separated from the cationic liposomes by size exclusion chromatography (Sephadex G-50 medium). Finally, the liposomes were passed through a 0.22 µm filter to remove any possible aggregates.

Characterization of Liposomes

The average diameter and size distribution of the liposomes were determined by dynamic light scattering (DLS) using a Zetasizer Nano (Malvern Instruments, Worcestershire, UK). The mean size of cationic liposomes was 91.7 ± 1.57 nm with a polydispersity index (PDI) of 0.091 ± 0.008, whereas the liposomes with rapalog were larger with sizes of 132 nm and PDI of 0.093. The zeta potential was on average 58.2 ± 10.5 mV. Total phospholipid content was determined by phosphate assay according to Rouser et al [17].

Immunofluorescence

Cells on glass cover-slips or poly-prep slides were fixed in 4% weight:volume PFA/PBS for 15' at RT, permeabilized with 0.1% volume:volume Triton®X-100/PBS, blocked with PBS/0.5% BSA/0.15% Glycine weight:volume, incubated overnight at 4°C with anti-V5 (Invitrogen, Cat. R960-25) and/or anti-HA (Sigma, Cat. 029K4788) and for 2h at RT with appropriate secondary antibodies. Coverslips were mounted on glass slides with DAPI/Vectashield for DNA staining.

Whole cell lysates and Western Blotting

Cells were lysed in lysis buffer (20mM Hepes pH7.5, 150mM KCl, 2.5mM EDTA, 5mM DTT, 10% Glycerol, 0.1% Triton, with 1X cOmplete, EDTA free protease inhibitors (Roche, Cat. 11873580001) added) and samples were subjected to immunoblotting using the LDB1 N-18 (Santa Cruz®, Cat. SC-11198), V5 or HA

antibodies. Fluorescently labeled secondary antibodies were used (Licor, IRDye 680LT series) and blots were imaged using the Odyssey imager.

Colony Forming Cell (CFC) assay

EBs were disrupted at day 6 with trypsin-EDTA. Cells were transferred to methycellulose-based media with 10% FCS, 1% L-glutamine, 0.25µg/ml transferrin, 0.25µg/ml ascorbic acid, 2µl/ml monothioglycerol, 5% protein free hybridoma medium II, 0.01µg/ml mIL6, 0.001µg/ml IL3 (R&D Systems, Cat.403-ML), 0.005µg/ml hIL11 (R&D Systems, Cat.418-ML), 0.003µg/ml GM-CSF (R&D Systems, Cat.415-ML), 4U/ml EPO (R&D Systems, Cat.959-ME), 0.005µg/ml TPO (R&D Systems, Cat.488-TO), 0.1 µg/ml SCF (R&D Systems, Cat.455-MC). Red primitive erythroid colonies and white macrophage colonies (composed of round cells growing in clumps) were identified microscopically after 6 days according to morphology and color.

Results

Anchor-Away tagged proteins are stable and properly localized, and are AA proficient

HEK293T human cells were co-transfected with constructs for TFs tagged with the FRB* domain and a V5 tag, as well as constructs for ribosomal proteins tagged with FKBP12 and an HA tag. HEK293T cells were chosen because they divide fast (every 24 hours), are easy to transfect (80% efficiency with lipofectamine is often achieved) and can be cultured in standard culture medium. Cells were fixed to a slide 48 hours after transfection and protein localization was visualized by immunofluorescence. The expected localization of the ribosomal protein is in nucleoli and cytoplasm while transcription factors should be localized in the nucleus, but not nucleoli. Pictures of stained cells confirmed these localizations (figure 1B, right panel).

Next the AA constructs were tested for their AA proficiency by adding rapalog overnight to the culture medium and fixing the cells the next day. Several ribosomal proteins were tested this way to determine which would be most effective to drag a TF to the cytoplasm. They were chosen according to the following criteria: i) homologs of ribosomal proteins that were reported to be efficient in AA [1] ii) homologs of ribosomal proteins that were determined to have their C-terminus exposed to the surface of the ribosome in yeast [18; 19]. Only when AA-tagged TFs were co-transfected with the ribosomal proteins rpL7a or rpL13 was a shift from nucleus to cytoplasm observed (supplementary table 1). Other ribosomal proteins showed no capacity as anchors. Further experiments were only conducted with these two AA tagged ribosomal proteins. Since the main advantage of AA is meant to be its speed, a time lapse experiment was conducted. These experiments with rpL7a or rpL13 showed a clear shift in localization of LDB1, RUNX1 and GATA1 from the nucleus to nucleoli and cytoplasm after 45 minutes incubation with rapalog, (figure 1B, data not shown).

Some concerns have been raised about the stability of FRB* tagged proteins [20] and although the double FRB* mutant is reported to be stable [12], we investigated the effects of FRB* fusion on protein stability. It has been reported that tagging a protein with an FRB domain destabilizes the protein by targeting it for degradation. Incubation with a heterodimerizing agent, such as rapamycin or rapalog, and binding to FKBP12 stabilizes the fusion protein. To investigate the stability of the fusion product, HEK293T cells were co-transfected with TF constructs as well as the ribosomal protein construct. After 48 hours, protein samples were taken. Other cells were incubated for three hours with ethanol or rapalog and subsequently protein samples were made. Stability of the protein was judged on Western blot, which showed no instability due to the tagging (Supplementary figure 1).

N- and C-terminal AA tagged LDB1 are both AA-proficient in combination with rpL13-AA in mouse ES cells, but only C-terminal AA tagged LDB1 (LDB1-AA) fully rescues *Ldb1 -/-* phenotype.

The next step after the proof-of-principle with overexpressed fusion proteins in HEK293T cells was to perform AA in relevant cells with (close to) endogenous levels. An ***Ldb1*** -/- ES cell was derived earlier in our laboratory [21] and was used to obtain an ES cell that expresses both the AA tagged ribosomal protein and ***Ldb1*** in a background that does not contain endogenous ***Ldb1***. ***Ldb1-/-*** ES cells do not have a phenotype, although ***Ldb1*** is normally expressed in ES cells. LDB1 function is important for the maintenance of long-term repopulating hematopoietic stem cells, the development of megakaryocytes and primitive erythrocytes [22; 23]. It was shown that LDB1 plays an important role in chromatin looping, by bringing together remote genomic sites, probably mediated through homodimerization of its N-terminus [24]. More recently it was shown that hemangioblast formation from ***Ldb1*** -/- ES cell derived embryoid bodies is severely hampered. Blood cell formation from the hemangioblast is blocked completely [21]. We first investigated the phenotype of ES cell differentiation after ***Ldb1*** expression was restored in these cells. To restore ***Ldb1*** expression, we used two different approaches: the cDNA of ***Ldb1*** was placed behind a constitutive EF1α promoter and expressed, and a BAC-tagged version of ***Ldb1*** was also prepared for expression. Either end was tried for the cDNA tagging, while the BAC-tagged version of ***Ldb1*** was only tagged at its C-terminus. Stable clones were obtained after selection, and clones that showed close to endogenous levels of LDB1 on Western blot were selected (data not shown). In these clones, AA-tagged rpL7a or rpL13 were introduced and again stable clones were obtained (figure 2A). These clones were tested for their hematopoietic differentiation phenotype in a colony forming cell (CFC) assay. Clones transfected with the N-terminal tagged version of ***Ldb1*** showed partial rescue of the ***Ldb1*** -/- phenotype, but clones transfected with C-terminal tagged versions of ***Ldb1*** were able to fully rescue the differentiation phenotype (figure 2B). After confirmation of the phenotype of the rescued ***Ldb1-AA*** ES cells, the AA system was tested in these cells, by addition of rapalog to the culture medium and immunofluorescence staining of the tagged LDB1 and ribosomal proteins (rpL13 and rpL7a). Both combinations showed proficiency for AA, but the localization shift for LDB1-AA/rpL7a-AA seemed unable to reach completion. A time course

experiment for LDB1-AA/rpL13-AA in ES cells shows that LDB1-AA is fully cleared from the nucleus after four hours, although not yet completely after two hours (figure 2C). This time window is a very significant improvement over that achieved with RNAi mediated knock-down, which takes at least 48 hours. Usually TFs exert their function during development in specific cell types rather than in ES cells as is the case for LDB1. In order to investigate whether AA is a widely applicable method, we therefore set out to show it is also working in differentiated cells.

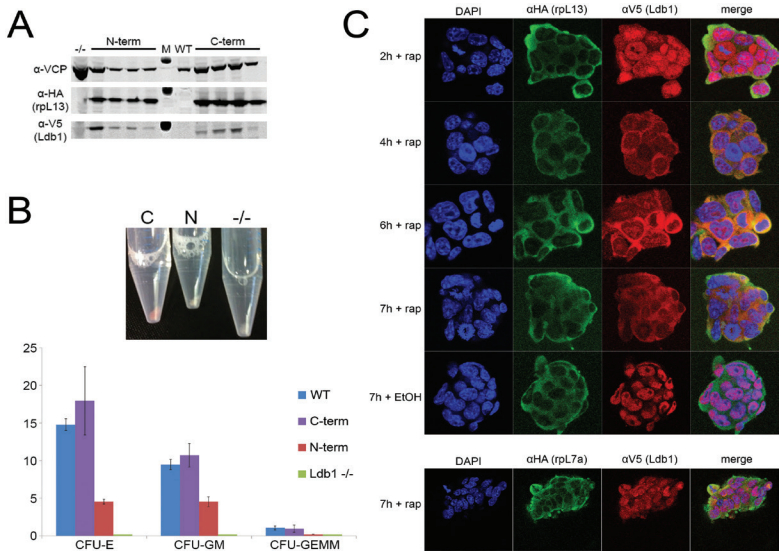


Figure 2. Analysis of clones transfected with AA tagged ***Ldb1*** in ***Ldb1-/-*** background.

A) Western blot on the V5 and HA tag and VCP (loading control). Shown are four clones with N-terminally tagged LDB1 (N-term) and four clones with C-terminally tagged clone (C-term). ***Ldb1-/-*** (-/-) and wildtype (WT) protein extracts are also shown as negative controls for HA and V5 staining. M = protein marker.

B) Rescue of ***Ldb1-/-*** hematopoietic differentiation phenotype by expression of AA tagged ***Ldb1***. A picture of sedimented EBs on day 6 of differentiation in PBS shows primitive erythrocytes (red color in the pellet) are produced in the C-terminally tagged clone (C), but not in the N-terminally tagged (N) or ***Ldb1-/-*** (-/-) EBs. Furthermore, the results for a CFC assay on WT, C-term, N-term and -/- cells is shown. The N-terminally tagged EBs do rescue the phenotype of ***Ldb1-/-*** in part, but C-terminally tagged clones rescue the phenotype fully. The result of three independent experiments is shown. CFU-E = Colony Forming Unit – Erythrocyte; CFU-GM = Colony Forming Unit – Granulocyte, Monocyte; CFU-GEMM = Colony Forming Unit – Granulocyte, Erythrocyte, Monocyte, Megakaryocyte

C) AA in ES cells with C-terminally tagged LDB1. Cells were incubated for indicated times with rapalog (rap) or EtOH. For rpL13 the LDB1-AA has already shifted substantially to the cytoplasm after 2h of incubation, and is completely gone from the nucleus after 4h. After 7h LDB1-AA is still in the cytoplasm. Incubation with EtOH does not trigger relocalization of LDB1-AA.

AA works in differentiated ES cells, when rapalog delivery is mediated by cationic liposomes

Ldb1-/- ES cells do not have a phenotype, and although ***Ldb1*** is expressed in ES cells, it is not associated with chromatin at this stage (A. Martella et al, unpublished). In order to show AA in cells where ***Ldb1*** is associated with other factors and functional, we differentiated ES cells for 3.25 days toward the hematopoietic lineage. At this stage, a peak of ***Flk1***+ cells is observed in the embryoid bodies. The ***Flk1***+ population is enriched for the hemangioblast, a hematopoietic/endothelial/vascular smooth muscle precursor [25-27]. It has been shown that LDB1 is associated with chromatin, and a complex containing LDB1 is regulating gene expression at this stage [21]. Embryoid bodies were disrupted and cells incubated in medium with rapalog or ethanol alone. Even after eight hours of incubation with rapalog, AA did not seem to work at all in these differentiated cells. Since the shift was observed in ES cells this suggested that either the rapalog was unable to enter the differentiated cells or that a conformational shift in the LDB1 protein caused the tag to be inaccessible in these cells. Haruki ***et al.*** had reported that rapalog could not be used for AA in baker's yeast, although the reason for this remained unclear. We reasoned that the problem could have been similar, suggesting that rapalog unable to enter the cell was the more likely possibility. We therefore tried approaches that could allow rapalog to penetrate the cell membrane.

The first approaches were to permeabilize the cell membrane with low concentrations of Triton (as described in [28]) or by adding lipofectamine to the medium with rapalog. Both approaches seemed to improve the situation to some extent and cells were present that showed a shift of localization indicating that penetration of the rapalog was indeed the problem (data not shown). Hydrophobic rapalog molecules can probably not be incorporated into the lipofectamine liposomes, as the inner part consists of a watery solution. Based on that it is to be expected that the chance of rapalog entering the cell is based on random proximity at the moment liposomes enter the cells. If true, the system could be improved by incorporating the hydrophobic rapalog into the lipid bilayer of liposomes using the film-loading method [15], as has been done for rapamycin [29]. We used cationic liposomes for this method, as the positive charge of these liposomes helps binding to the negatively charged cellular membrane [30]. Figure 3 shows that AA is working in differentiated cells after four hours of incubation with the rapalog-loaded cationic liposomes, but not for incubation with empty cationic liposomes.

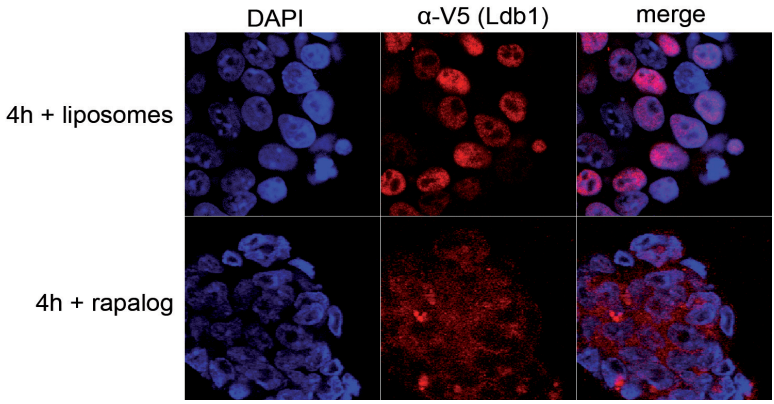


Figure 3. AA in differentiated ES cells. Images of a clone transfected with the C-terminally tagged ***Ldb1*** construct in ***Ldb1-/-*** ES cells. Differentiated cells were incubated with empty or rapalog-loaded cationic liposomes for four hours. LDB1 was imaged by immunofluorescence on the V5 tag. The picture shows that the rapalog-loaded, but not the empty liposomes induce a shift of LDB1-AA to the cytoplasm.

Discussion

The results presented here show a fast (<4 hours) removal of a target nuclear TF out of the nucleus in functionally relevant cells. LDB1 is a member of TF complexes in early and late erythropoiesis [2; 31; 32] and cannot act on its own due to an inability to bind DNA directly (reviewed in [33]). As such, the results presented here proof that AA can be used to knock out individual members of a complex. Moreover, the system can be used to knock out a nuclear target gene fast after cells are allowed to develop up to a specific point. This is impossible to accomplish with conventional siRNA-mediated knockdown or conditional knock out of genes, as these techniques are constrained by protein half-life. It has been noted in yeast that a gradual depletion of a target protein from the nucleus can lead to a different phenotype when compared to a sudden depletion [34].

A fast removal as presented here provides an opportunity to distinguish primary from secondary effects of the knock down/out. Improving the rate of primary over secondary effects of a knock down is an important priority, as it would allow us to better understand which genes are regulated by a specific TF. This in turn gives a better understanding of cellular functions and genes that are regulated in development. A last speculative feature of AA is that it is in theory more rapidly reversible than siRNA-mediated knock down: after removal / wash-out of the rapalogue, the system would be independent on transcription for reconstitution of nuclear protein levels. Reversibility of the system might allow temporary gene depletion at a specific developmental stage, after which protein levels could be restored to normal. This is a possible application that could be tested in future.

AA does not control the stability of the endogenous, untagged proteins. In

order to investigate fast nuclear factor depletion using AA, one has to (i) express an AA-tagged target nuclear protein and ribosomal protein in a cell of knock-out background or (ii) modify the endogenous protein by gene targeting. Improvements in homologous recombination and the development of recombineering techniques make this last approach more attractive for mouse ES cells [35]. I therefore introduced the recombineering technique in the laboratory and targeted mouse ES cells to develop five different cell lines that can be used to obtain mouse models of AA. The TFs targeted are ***Cbf-β***, ***Myb***, ***Gata1***, ***Gata2*** and ***Tal1*** (see Chapter 4). Furthermore, techniques such as zinc finger nucleases, transcription activator-like effector endonucleases (TALENs) and the most recently developed Cas9 system (zinc fingers and TALENs recently reviewed in [36], Cas9 is described in [37-40]) make the gene targeting approach feasible for many cell types including human ES cells and/or human somatic cells. An rpL13-targeted mouse ES cell line has been established and is confirmed to yield high chimaeras after blastocyst injection (supplementary figure 2), making it easier to obtain AA mice after one targeting event at the target gene. The C-terminal tagging by recombineering described here can easily be applied to any other gene, provided the correct BAC is available. Taken together these resources provide a short route to obtain AA proficient ES cells/mice of any gene.

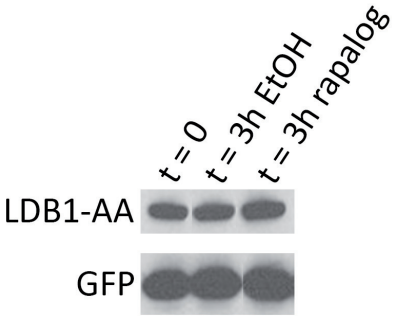
Other systems have been described that regulate posttranslational regulation of protein stability, most recently the auxin-induced degron (AID) system in human cells [41]. The AID system has yielded promising results in two human cell lines (DLD-1 and RPE-1) after cDNA overexpression. Although the results for the AID system are promising, it was not tested in a knock out background or with endogenous levels of proteins. We would therefore like to emphasize that a shift in localization for overexpression experiments in HEK293T cells was visible within 45 minutes (figure 1), but in ES cells no shift could be observed before approximately two hours (figure 2, data not shown). The exact timing in differentiated cells remains to be established, but appears at least as fast as observed in ES cells (figure 2 and 3).

A possible application of the system would be to target specific cells in mice. Results in this work (figure 3) indicate a potential inability of rapalog to enter somatic cells. This will have to be tested on a cell-to-cell basis and may be a hurdle to overcome if animal experiments are to be undertaken. The use of liposomes as presented here may help overcoming this hurdle, but potentially other avenues to target rapalog ***in vivo*** could be explored.

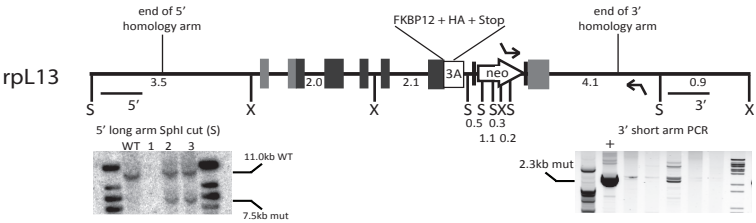
Supplementary table and figures for chapter 3

Supplementary table 1. Ribosomal proteins tested in AA in HEK293T cells. Only rpL13 showed positive results in mouse ES cells.

Ribosomal protein	AA proficient in HEK293T	Ribosomal protein	AA proficient in HEK293T
rpL7a	Yes	rpL23a	No
rpL13	Yes	rpL24	No
rpL4	No	rpSa	No
rpL10a	No	rpS2	No



Supplementary Figure 1. Stability of the FRB* fusion protein. The same amount of protein was loaded of protein samples taken on: 48 hours after transfection (0), 51 hours after transfection, including 3h of incubation with EtOH in the medium (3h EtOH) and 51 hours after transfection, including 3h of incubation with rapalog in the medium (3h rapalog). A plasmid expressing GFP in mammalian cells was co-transfected to serve as a loading control. No difference in amount of LDB1-AA is observed, hence the fusion protein's stability is similar in all cases. The blot shown is representative of three independent experiments.



Supplementary Figure 2. Overview of rpL13 targeting with AA tag. The targeted genomic locus of rpL13 is shown, exons are boxed, light gray indicates untranslated parts of the exons. Numbers underneath the locus correlate to the distance between restriction sites in kbps, where S = SphI and X = XbaI. Arrows indicate the location of primers used to screen for homologous recombination. Positive clones show a 2.3kb band, an amplicon that spans the 3' short homology arm (shown at the bottom right, positive is indicated with a +). The integration was confirmed at the 5' side with a Southern blot. Beneath the locus, the position of the probes are indicated with a horizontal line. The result of the Southern blot is indicated on the lower left. WT = wild type, mut = mutation after homologous recombination.

References

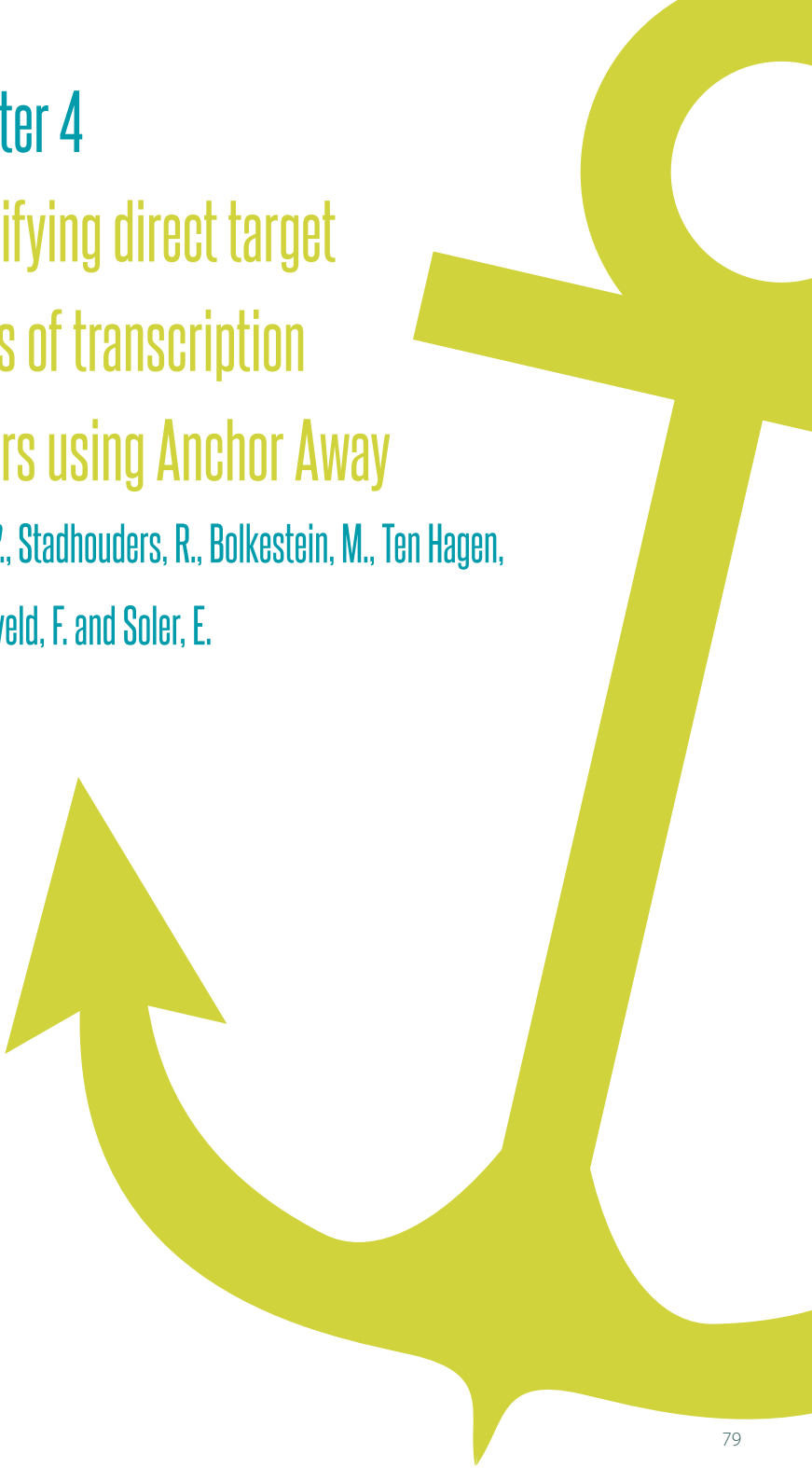
1. Haruki, H., Nishikawa, J., ***et al.*** (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell ***31***, 925-932.
2. Soler, E., Andrieu-Soler, C., ***et al.*** (2010). The genome-wide dynamics of the binding of ***Ldb1*** complexes during erythroid differentiation. Genes Dev ***24***, 277-289.
3. Pevny, L., Simon, M.C., ***et al.*** (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature ***349***, 257-260.
4. Gonczy, P., Echeverri, C., ***et al.*** (2000). Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. Nature ***408***, 331-336.
5. Fraser, A.G., Kamath, R.S., ***et al.*** (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature ***408***, 325-330.
6. Yuan, P., Han, J., ***et al.*** (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev ***23***, 2507-2520.
7. Zhao, L., Glazov, E.A., ***et al.*** (2011). Integrated genome-wide chromatin occupancy and expression analyses identify key myeloid pro-differentiation transcription factors repressed by ***Myb***. Nucleic Acids Res ***39***, 4664-4679.
8. Zagouri, F., Sergentanis, T.N., ***et al.*** (2012). mTOR inhibitors in breast cancer: a systematic review. Gynecol Oncol ***127***, 662-672.
9. Choi, J., Chen, J., ***et al.*** (1996). Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. Science ***273***, 239-242.
10. Chen, J., Zheng, X.F., ***et al.*** (1995). Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue. Proc Natl Acad Sci U S A ***92***, 4947-4951.
11. Belshaw, P.J., Ho, S.N., ***et al.*** (1996). Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins. Proc Natl Acad Sci U S A ***93***, 4604-4607.
12. Stankunas, K., Bayle, J.H., ***et al.*** (2007). Rescue of degradation-prone mutants of the FK506-rapamycin binding (FRB) protein with chemical ligands. Chembiochem ***8***, 1162-1169.
13. Klemm, J.D., Beals, C.R., ***et al.*** (1997). Rapid targeting of nuclear proteins to the cytoplasm. Curr Biol ***7***, 638-644.
14. Lasic, D.D. (1993). Liposomes from physics to applications (Amsterdam, The Netherlands: Elsevier).
15. Sadzuka, Y., Takabe, H., ***et al.*** (2005). Liposomalization of SN-38 as active metabolite of CPT-11. J Control Release ***108***, 453-459.
16. Haeri, A., Sadeghian, S., ***et al.*** (2011). Use of remote film loading methodology to entrap sirolimus into liposomes: preparation, characterization and ***in vivo*** efficacy for treatment of restenosis. Int J Pharm ***414***, 16-27.
17. Rouser, G., Fkeischer, S., ***et al.*** (1970). Two dimensional then layer chromatographic separation of polar lipids and determination of phospholipids by phosphorus analysis of spots. Lipids ***5***, 494-496.
18. Spahn, C.M., Beckmann, R., ***et al.*** (2001). Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-ribosome and subunit-subunit interactions. Cell ***107***, 373-386.
19. Ben-Shem, A., Garreau de Loubresse, N., ***et al.*** (2011). The structure of the eukaryotic ribosome at 3.0 A resolution. Science ***334***, 1524-1529.
20. Stankunas, K., Bayle, J.H., ***et al.*** (2003). Conditional protein alleles using knockin mice and a chemical inducer of dimerization. Mol Cell ***12***, 1615-1624.
21. Mylona, A., Andrieu-Soler, C., ***et al.*** (2013). Genome-wide analysis shows that ***Ldb1*** controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in

hematopoiesis. Blood.
22. Li, L., Jothi, R., ***et al.*** (2011). Nuclear adaptor ***Ldb1*** regulates a transcriptional program essential for the maintenance of hematopoietic stem cells. Nat Immunol ***12***, 129-136.
23. Li, L., Lee, J.Y., ***et al.*** (2010). A requirement for Lim domain binding protein 1 in erythropoiesis. J Exp Med ***207***, 2543-2550.
24. Deng, W., Lee, J., ***et al.*** (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell ***149***, 1233-1244.
25. Eichmann, A., Marcelle, C., ***et al.*** (1993). Two molecules related to the VEGF receptor are expressed in early endothelial cells during avian embryonic development. Mech Dev ***42***, 33-48.
26. Eichmann, A., Corbel, C., ***et al.*** (1997). Ligand-dependent development of the endothelial and hemopoietic lineages from embryonic mesodermal cells expressing vascular endothelial growth factor receptor 2. Proc Natl Acad Sci U S A ***94***, 5141-5146.
27. Choi, K., Kennedy, M., ***et al.*** (1998). A common precursor for hematopoietic and endothelial cells. Development ***125***, 725-732.
28. van de Ven, A.L., Adler-Storthz, K., ***et al.*** (2009). Delivery of optical contrast agents using Triton-X100, part 1: reversible permeabilization of live cells for intracellular labeling. J Biomed Opt ***14***, 021012.
29. Rouf, M.A., Vural, I., ***et al.*** (2009). Development and characterization of liposomal formulations for rapamycin delivery and investigation of their antiproliferative effect on MCF7 cells. J Liposome Res ***19***, 322-331.
30. Magee, W.E., and Miller, O.V. (1972). Liposomes containing antiviral antibody can protect cells from virus infection. Nature ***235***, 339-341.
31. Meier, N., Krpic, S., ***et al.*** (2006). Novel binding partners of ***Ldb1*** are required for haematopoietic development. Development ***133***, 4913-4923.
32. Rodriguez, P., Bonte, E., ***et al.*** (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. EMBO J ***24***, 2354-2366.
33. Matthews, J.M., and Visvader, J.E. (2003). LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. EMBO Rep ***4***, 1132-1137.
34. Nishimura, K., Fukagawa, T., ***et al.*** (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods ***6***, 917-922.
35. Hofemeister, H., Ciotta, G., ***et al.*** (2011). Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. Methods ***53***, 437-452.
36. Perez-Pinera, P., Ousterout, D.G., ***et al.*** (2012). Advances in targeted genome editing. Curr Opin Chem Biol ***16***, 268-277.
37. Jinek, M., Chylinski, K., ***et al.*** (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science ***337***, 816-821.
38. Hwang, W.Y., Fu, Y., ***et al.*** (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol.
39. Cong, L., Ran, F.A., ***et al.*** (2013). Multiplex genome engineering using CRISPR/Cas systems. Science ***339***, 819-823.
40. Jinek, M., East, A., ***et al.*** (2013). RNA-programmed genome editing in human cells. Elife ***2***, e00471.
41. Holland, A.J., Fachinetti, D., ***et al.*** (2012). Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc Natl Acad Sci U S A ***109***, E3350-3357.

Chapter 4

Identifying direct target genes of transcription factors using Anchor Away

Jorna, R., Stadhouders, R., Bolkestein, M., Ten Hagen, T., Grosveld, F. and Soler, E.



Abstract

Anchor-Away (AA) is a technique that provides a fast (<4h), specific and almost complete removal of target nuclear proteins by sequestering them to the cytoplasm. Here we describe efficient development of AA with six transcription factors involved in hematopoiesis and show the system is functional for three of six factors (other three not tested). The AA system is ready-to-use for the X-linked factor GATA1 and for the general adaptor molecule LDB1. AA-mediated inactivation of specific nuclear factors followed by gene expression analysis should yield better insights in direct target genes of TFs compared to conventional RNAi knock down or (conditional) knock out of the same factors. The use of AA may pave the way toward clearer, more advanced gene regulatory networks for cellular development as a result of limited secondary effects.

Introduction

An important drawback of traditional knock out/down techniques is that they produce the desired primary effects but also substantial secondary effects. Typically, the target genes of a given TF become differentially regulated after knock out/down of the factor, but transcriptome measurements will also capture the differential regulation of secondary target genes which are result from misregulation of the first set of direct target genes. The problem is in timing: typically gene expression analysis takes place 48 to 72 hours after knock out/down, a time which is necessary to eliminate the endogenous protein from the cell. To achieve a faster way to deplete target proteins, new techniques may be focused on directly disrupting target protein function rather than its mRNA or gene. Several studies have been published that focus on post-translational modification of the target protein by addition of small molecules, e.g. the auxin-inducible degron system [1]. In chapter 3 we described Anchor-Away (AA) in mammalian cells as a fast and efficient alternative.

Anchor-Away (AA) in mammalian cells is based on the Anchor-Away technique in yeast [2] and works through rapalog-mediated ternary complex formation between an 11-kDa FRB* domain and a 12-kDa FKBP12 domain. The FRB* domain is fused to the target nuclear factor and the FKBP12 domain to the ribosomal protein L13 (rpL13). Once rapalog is taken up by the cells, a ternary complex is formed, i.e. the FRB*-TF and FKBP12-rpL13 fusion proteins together with the rapalog. After ternary complex formation the FRB*-fused TF is rapidly moved from the nucleus into the cytoplasm due to the large flow of ribosomal proteins through the nucleus towards the cytoplasm, disrupting nuclear function of the target TF. A schematic overview of AA and immunofluorescence pictures of the result is presented in figure 1. Here we describe data of AA-mediated protein disruption of three out of six mammalian transcription factors that were targeted to express the AA-tag from the 3' side. The transcription factors ***Myb***, ***Gata1***, ***Gata2***, ***Cbf-β*** and ***Tal1*** were gene

targeted at the C-terminus with the FRB* domain, combined with a V5 tag. rpL13 was expressed as a transgene driven from the general CMV/chicken β-Actin (CAG) promoter, fused with the FKBP12 domain and an HA tag. Both tags will be referred to as AA tags below. ***Ldb1-AA*** was expressed from a BAC-tagged transgene in an ***Ldb1-/-*** background.

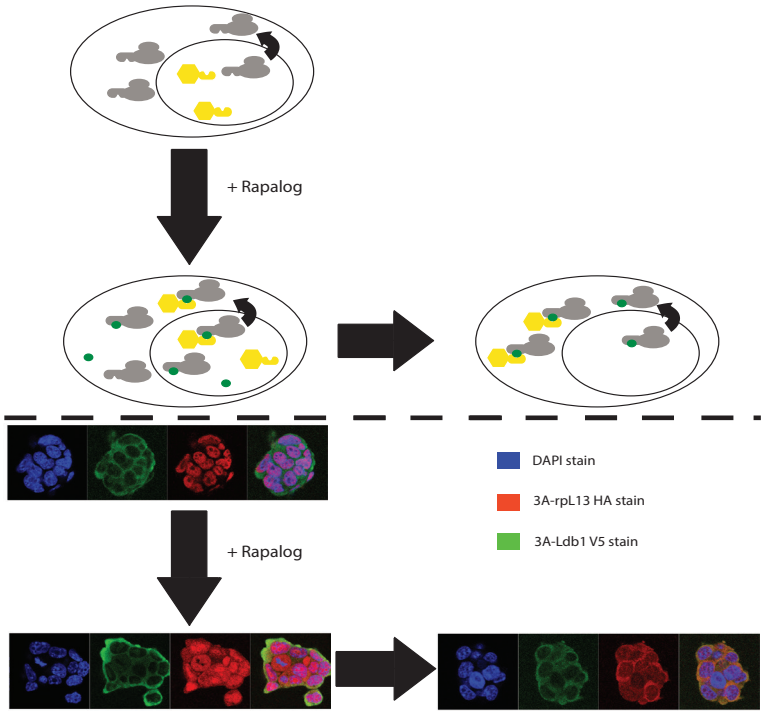


Figure 1. The AA technique. AA-tagged ribosomal protein rpL13 and AA-tagged target nuclear protein form a ternary complex upon addition of rapalog. The large flow of ribosomes from the nucleoli to the cytoplasm (and through the nucleus) drags along the target. Target nuclear protein is now sequestered in the cytoplasm. This is shown in the cartoon on top and immunofluorescence pictures of AA in mouse ES cells are shown below to illustrate.

The proteins were chosen because they are involved in hematopoiesis, and more particular in erythropoiesis. ***Ldb1*** (***Clim2, NL1***) as well as ***Tal1*** (***Scl***) play a role in the emergence of both the hemangioblast as primitive erythroid cells [3-7]. ***Gata2*** is essential for the emergence of the hematopoietic stem cell [8; 9], as well as ***Cbf-β***, the partner protein of RUNX1 [10].

Gata1-/- ES cells that are differentiated ***in vitro*** encounter a block in erythroid development, as they are not allowed to develop past the pro-erythroblast stage [11; 12]. Moreover, with the exception of MYB, all proteins investigated interact as protein partners in a transcription factor complex that controls expression of genes important for erythroid development [13-15]. The ***Myb*** gene itself, which plays a central role in HSCs and erythropoiesis is regulated by this complex, termed the LDB1-complex [16].

To investigate the influence of the factors on hematopoietic development, AA was performed in cells obtained after ***in vitro*** differentiation of genetically engineered ES cells. Two stages were chosen for further investigation, the ***Flk1***+ cell stage and the primitive erythroid stage. Both stages can be obtained by ***in vitro*** hematopoietic differentiation of ES cells [17-19]. A population of cells that express ***Flk1*** (***Kdr, Vegfr2***) can be observed after 3.25 days of hematopoietic differentiation of ES cells in culture. This population of cells is enriched for the so-called hemangioblast, an intermediate progenitor for hematopoietic, endothelial and smooth vascular muscle cells [20-22]. After six days of differentiation, primitive erythrocytes can be observed in blood islands of embryoid bodies [19].

Rapalog was loaded into the lipid bilayer of cationic liposomes in order to allow entrance of rapalog into the differentiated cells. EB cells derived from the targeted ES cells were AA proficient for all clones tested. Moreover, AA-mediated protein disruption in these cells happens fast, in less than four hours. This makes AA an important candidate technique when a thorough investigation for target genes of a specific protein is needed.

Material and Methods

Cell Culture

ES cells were cultured and differentiated as described in Chapter 3. ES cells were kept on irradiated feeders, unless mentioned otherwise. For gene targeting, 15µg of DNA was electroporated into $5*10^7$ cells using a Bio-Rad Gene Pulser II with the following settings: exponential wave, 800V, 10uF (Tc was 0.2 or 0.3 ms). Cells were seeded on gelatinized dishes in several densities and selection was started the next day. Medium-size round colonies were picked after seven days and allowed to expand.

Recombineering, generating targeting vectors

Recombineering plasmids were a kind gift of Francis Stewart's laboratory. The recombineering was performed as described [23]. A list of recombineering primers used is given in the supplementary table 1.

Southern Blotting

15µg of genomic DNA was cut with the appropriate restriction enzyme and run on a 0.7% weight:volume agaraso gel. The DNA was blotted to Hybond XL (Amersham, Cat. RPN203-S). Probes were made using 32P α-ATP random prime

labeling (Stratagene, Cat. 300385) of PCR amplicons indicated in figure 2 and supplementary figure 1. The primers used to generate the amplicons are shown in the supplementary table 1. Radioactive hybridization signals were imaged on a phosphor screen after washing of the Hybond filter.

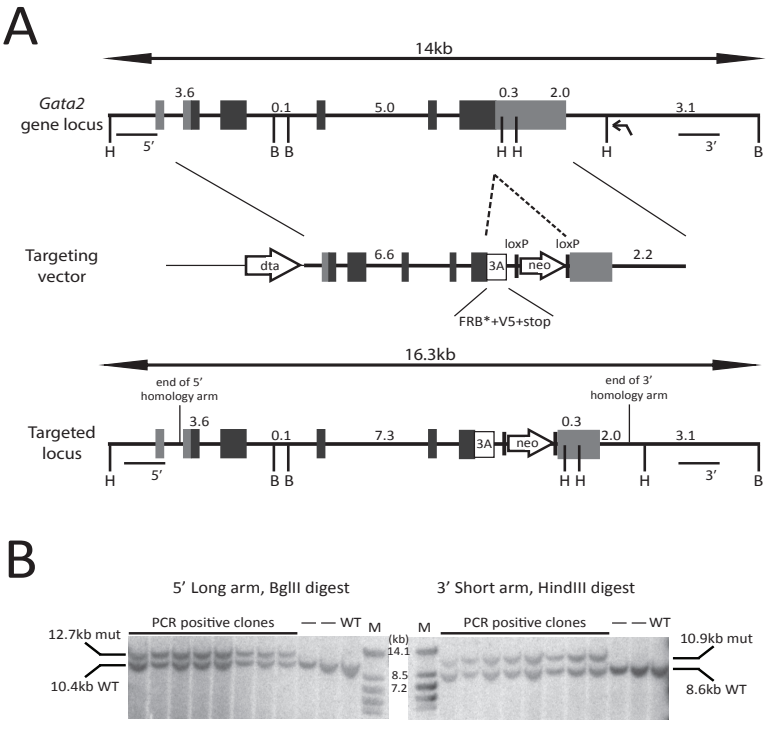


Figure 2. Knock in strategy on the ***Gata2*** gene locus. A) Overview of the (targeted) ***Gata2*** locus, exons are boxed, coding regions are in dark gray. Numbers depict the distance between different restriction sites in kbp. Restriction sites for HindIII (H) and BglII (B) are indicated, as are the positions for of the Southern Blot probes (5' and 3'). The arrow in the top view indicates the position for the primer that was used to screen for integrations. Dta = Diphteria toxin A; AA = AA tag; neo = neomycin resistance gene. B) Southern blot on eight clones that were positive for the integration on PCR, two clones that were negative (-) and wildtype DNA (WT). M = λ-BstEII marker

Results

Efficient targeting of C-termini of transcription factors with AA tags

The selected transcription factors were targeted at their C-termini with the AA tag. Targeting constructs were obtained through recombineering of BACs and gap repair in a suitable targeting vector. A screen for correct targeting events

was conducted using PCR over the short homology arm. Targeting events were confirmed by Southern blotting. An example of the strategy for ***Gata2*** is given in figure 2, an overview of the other loci can be found in supplementary figure 1. The loci were targeted with >20% efficiency for all genes with the exception of ***Myb*** (Table 1). The targeting efficiency observed for ***Myb*** is consistent with published data [24]. Targeting efficiencies may have suffered from the use of non-isogenic DNA as the BAC libraries used were obtained from C57/Bl6 DNA and the targeted IB10 cells are of 129P2 origin.

Table 1. Targeting efficiencies for all genes targeted. Correct targeting events of at least four clones per gene were confirmed by Southern Blot. For ***Myb*** all six clones were confirmed by Southern Blot, the number of the other genes correspond to the number of clones positive in the PCR screen.

Gene	Number of clones analyzed	Number of clones targeted	Targeting efficiency	Number of clones: differentiated / blood islands seen after 6 days
Gata2	24	14	58%	2 / 2
Gata1	179	42	23%	4 / 4
Tal1	192	49	26%	2 / 2
Myb	192	6	3%	2 / 2
Cbf-β	192	125	65%	2 / 2

All of the ES cell clones obtained have normal morphology and growth phenotypes, and are able to produce blood islands when differentiated towards the hematopoietic lineage (Table 1). For most factors a phenotype was not expected, as no phenotypes for ES cells haplo-insufficient for any of the factors have been reported. The X-linked gene ***Gata1*** however, is hemizygous for the insertion after the first targeting event in the male IB10 cells. ***Gata1***-/- ES cells have an inability to produce primitive erythrocytes, as cells are blocked at the pro-erythroblast stage [25]. The observation that primitive erythrocytes are formed in these EBs suggests that the AA-tag does not interfere with ***Gata1*** function. Expression of ***Gata1***, ***Tal1*** and ***Ldb1*** was imaged on at least two independent ES cell clones and again on day 3.25 and day six of differentiation (figure 3 and 4). All clones tested faithfully recapitulated the expression of the genes.

AA works in EB cells derived from the targeted ES cell clones

The AA system was tested in the targeted clones for ***Tal1***, ***Ldb1***, and ***Gata1***. RpL13-AA was stably expressed as a transgene from the CAG promoter in these clones. Proper expression of the ***rpL13-AA*** construct has been confirmed previously. The stable clones obtained with ***rpL13-AA*** were differentiated toward the hematopoietic lineage and disrupted after 3.25 days (***Tal1***, ***Ldb1***) or 6 days (***Gata1***). EBs were disrupted at indicated stages and cells were incubated in differentiation

medium. As described in Chapter 3, adding rapalog to the medium of EB cells will not induce AA. Instead, rapalog has to be delivered after incorporation in the lipid bilayer of liposomes. After four hours of incubation, cells were fixed on a slide and localization of the transcription factors judged by immunofluorescent staining. AA works for TAL1-AA, GATA1-AA, LDB1-AA (figure 4 and chapter 3, [Jorna et al., ***in preparation***]). This confirms that the cells generated are AA proficient.

Since endogenous protein (i.e. non AA-tagged) is still present in the ***Myb***, ***Gata2***, ***Tal1*** and ***Cbf-β*** targeted heterozygous ES cells, the neomycin resistance cassette will have to be removed by Cre expression and the second allele subsequently targeted to investigate phenotypes. The ***Ldb1-AA*** and ***Gata1-AA*** cells can be used directly for protein disruption studies.

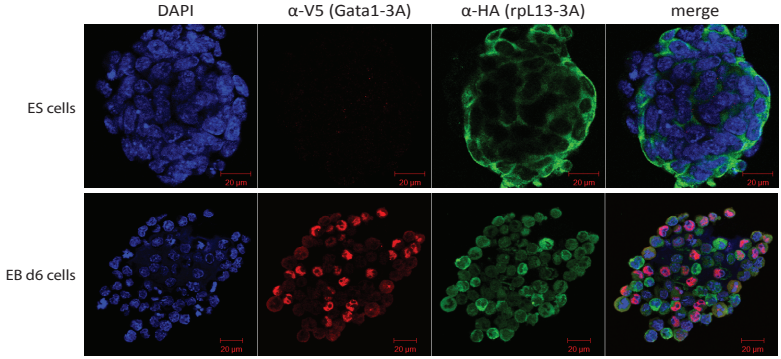


Figure 3. Imaging GATA1-AA at different stages of ES cell differentiation. Immunofluorescence pictures of GATA1-AA (red) and ***rpL13-AA*** (green) in ES cells and embryoid body (EB) cells at day 6 of differentiation. GATA1-AA is not yet expressed in ES cells (upper row) or EB cells at day 3.25 of differentiation (not shown), but it is in subsets of cells at day 6 (lower row.

Discussion

A comprehensive description of primary target genes after knock down of target TFs has been challenging due to an inability to dissect secondary from primary targets. By substantially reducing the time before gene expression analysis can be performed, AA may outperform traditional methods and substantially lower the amount of secondary targets. Moreover, we present a method to make use of AA in differentiated cells, making it possible to focus on developmentally relevant cells. A good measure to judge the rate of secondary effects is given by the overlap between genes showing a differential regulation after AA and genes that show binding of the transcription factor near their genomic locus as determined by ChIP-Seq. This overlap is typically 30% at best (e.g. [26; 27]) when using conventional knock down techniques (e.g. RNAi), and is expected to increase after AA. Cells generated in this work can be used to investigate and compare the techniques.

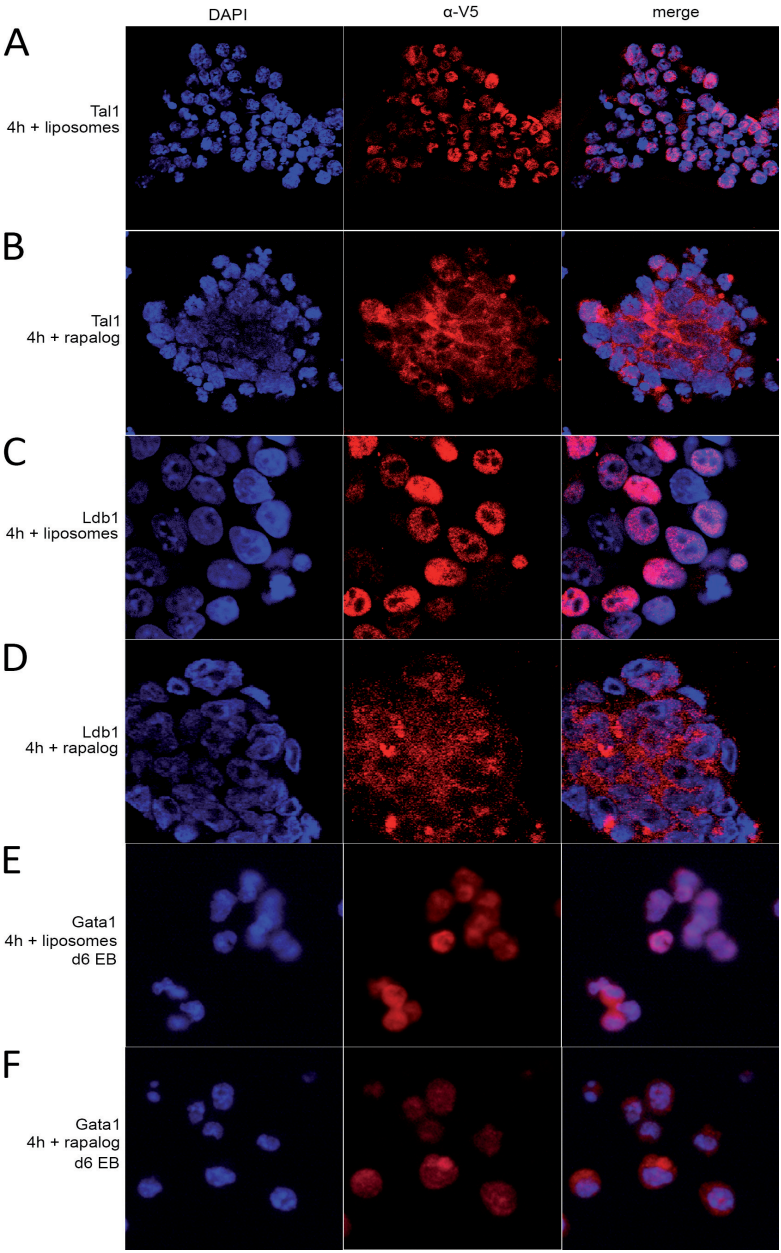

DAPI
α-V5
merge
A
Tal1
4h + liposomes
B
Tal1
4h + rapalog
C
Ldb1
4h + liposomes
D
Ldb1
4h + rapalog
E
Gata1
4h + liposomes
d6 EB
F
Gata1
4h + rapalog
d6 EB

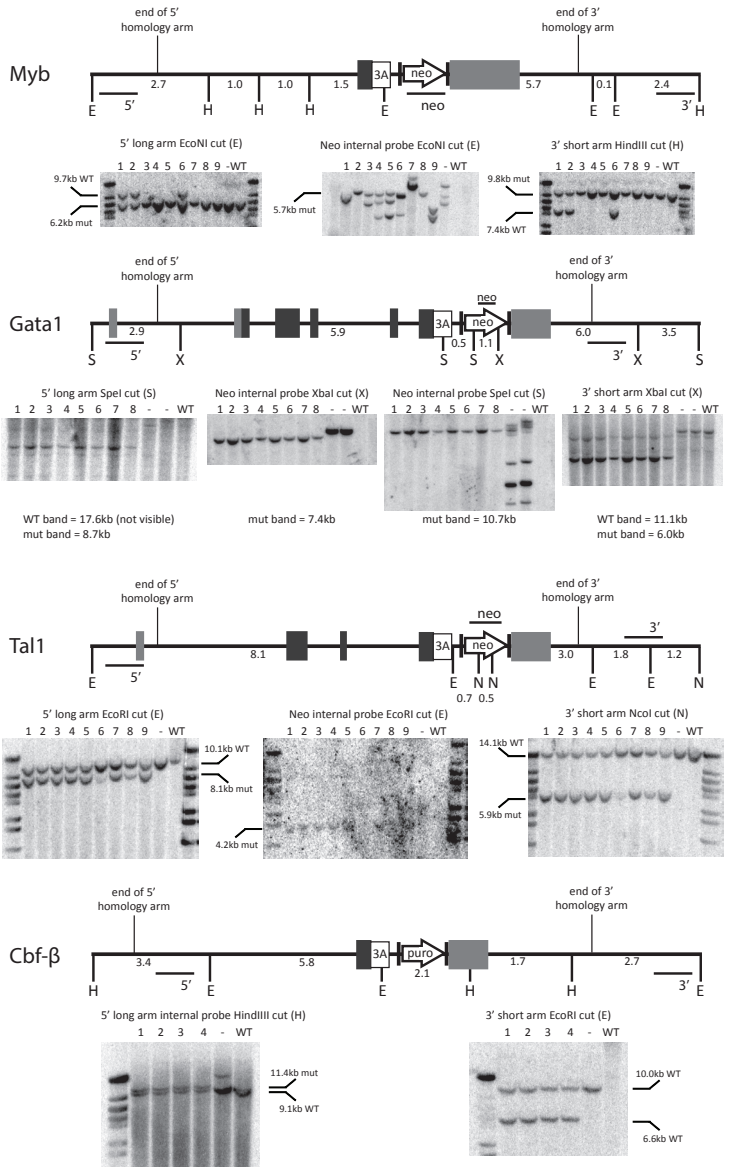
Figure 4. AA in differentiated EB cells. The C-termini of TAL1 and GATA1 were fused to an AA tag by gene targeting in mouse ES cells(supplementary figure 1) and ***rpL13-AA*** was transfected as a transgene.. ***Ldb1-AA*** was expressed in ***Ldb1-/-*** ES cells together with ***rpL13-AA***. ES cells were differentiated for 3.25 days (6 days for ***Gata1***) and resulting EBs were disrupted. Cells were fixed and stained four hours after liposomes were added. When empty liposomes were added AA fused proteins did not relocalize to the cytoplasm (panel A, C and E). When rapalog is loaded into the lipid bilayer of liposomes, the proteins do change localization (panel B, D and F). Pictures on panels A-D were taken on 40x magnitude using a confocal laser scanning microscope (Zeiss LSM 510 Meta), pictures on panels E and F were taken using a CCD camera coupled to a widefield microscope (Olympus IX70).

To allow entrance of rapalog into differentiated cells, the cells were incubated with rapalog-loaded cationic liposomes. We described earlier that rapalog was unable to enter differentiated cells, although uptake of the rapalog was possible in ES cells, HEK293T and HeLa (data not published) cells. The reason for this difference is not clear, although the most logical underlying explanation would be changes in the (composition of the) cellular membrane. It is not likely this is caused by expression of transmembrane proteins that are able to clear the cytoplasm of rapalog, since AA can be induced when rapalog is loaded by the use of the cationic liposomes.

Plasmids generated for this work have successfully been used to target AA tags to different proteins. These plasmids can be used to generate targeting plasmids to tag any protein of interest with the AA tag using recombineering [23]. A mouse ES cell line harboring an AA tag at the C-terminus of rpL13 already has been generated and used to generate chimaeric mice. The selectable cassette has been removed from this cell line by transient expression of Cre recombinase, making selection with G418 possible in these ES cells for additional gene targeting. It is also possible to generate AA mice using these cells, an exciting possibility that enables working with primary cells and tissues. Whether AA can be used in live animals will depend on the ability of rapalog to enter cells ***in vivo***. In this and previous work we used cationic liposomes to target rapalog to differentiated cells ***in vitro***. This is an avenue worth exploring ***in vivo*** as well, as liposomes have already been used to allow rapamycin to enter specific tissues both ***in vitro*** [28; 29] and ***in vivo*** [29; 30].

We conclude that AA may provide an important new way to investigate gene function, particularly for transcription factors. Transcription factors influence cell fate by tuning gene expression programs, and often exert their function in small developmental windows. AA provides a chance to disrupt specific factors within the functionally relevant time window during development, while leaving the factor untouched at other stages of development.

Supplementary table and figures for chapter 4



Supplementary figure 1. Knock in strategy on the ***Myb***, ***Gata1***, ***Tal1*** and ***Cbf-β*** locus. Indicated are only the targeted genomic loci, with distances in between restriction sites in kbp under the line. Exons are boxed, light gray indicates non-coding (parts of) exons. Southern blots that confirm correct integrations are shown. PCR-positive clones are numbered, negative clones are indicated with – and wildtype DNA with WT.

Supplementary Table 1. Primers used for recombineering, gap repair and Southern Blot analysis described. For ***Cbf-β*** the neomycin resistance gene was swapped for a puromycin resistance gene using recombineering (as described in [31]). To this end, first a counter-selectable cassette (cs) was brought in, and after that a puromycin selectable marker. The primers used for this are given as well.

Forward primer	Forward primer Sequence 5'>3'	Reverse primer Sequence 5'>3'
Gata2 AA tagging	GTCTCTCTTTTGGCCACCCCCACCCG TCCAGCATGGTGACTGCCATGG GCGC CAGCACCAAGGGCCC	GTTCTGCCAAACCACCCTTGATG CCATGTCTGTCCAGTGGGAGGCTT GC GGATCCCCTCGAGGGACC
Gata2 gap repair	CCAGGGCCGGCGGATTCACGGGATA GGGGTGGGGGACAGCGCGCGTCTC GATTTAAATCACCGGTGACCCGGGTC	CCTAGAGTCTGTCTCTCATCAAGGA GGTTTCTCATGGCCACTGTGTGGAC ATTTAAATCGCCGGCGACTTAAGTC
Gata2 5' probe	GTAGAACCTGTAGTTTAAACCC	AAAGACAGTGGTCACCCTGC
Gata2 3' probe	TGGACTCTTGGAGTTTAGGTC	ACTCCTGCACAGACGTGAAG
Myb AA tagging	GTCCGGCTCGGAAATACGTGAACGC GTTCTCAGCTCGAACTCTGGTCATG	AAGTCAACTTTTAAGTGTTCTGAAA ACCATAATGCTTTTCTGGAAATGTC
Myb gap repair	GCCAGCACCAAGGGCCCGGAAATT CTATTTTACAATGTGAAGATGTAATA AGACAGCGAGATGAATCATTTAAAT CACCGGTGACCCGGGTC	GGATCCCCTCGAGGGACCGTATCCCT TGCACAAATACTGATTGATTCTCACAA CTCAAATCCAAACTCATTTAAATCGCC GGCGACTTAAGTC
Myb 5' probe	CAGAGAGACTTGACAGTTGTC	GAATCCAGAGTAAGAATTCTCC
Myb 3' probe	CCATTGGTCTGCAACAGAACC	CTTCATATGTTCGAGGGTATCC
Gata1 AA tagging	GTCCTGCACCTACCACCAGCAGCACC AGCGTGATCGCCCCACTCAGTTCT GC CAGCACCAAGGGCCC	AAAAAGGGGGGGTCTCCTCTGCCA CAAGGTCAAGGCTATTCTGTGTACCT GGATCCCCTCGAGGGACC
Gata1 gap repair	TTATTGATTGTGAAAAGTCAGCTTGGT TTATATAAGAAGATCCTATCAGA ATTTA AATCACCGGTGACCCGGGTC	GTGGTGGCGCACACCTTTAAACCCAG CACTCAGGAGGCAGAGGCAGGCGC ATTTAAATCGCCGGCGACTTAAGTC
Gata1 5' probe	TTGGGGGACATGGTAAAATGC	CAAGGCCAGTGAGGACTCC
Gata1 3' probe	GACATGATTGAGGAAGAGGGA	ATTCTCCTCTACACACAGTAAC
Tal1 AA tagging	GCAGCCTCCATCCTGCCCTGCTGCCT GCCGCTGATGGGGCTGGCCCCCGG G CCAGCACCAAGGGCCC	CCAGGGGCCTAAAGGCCCCTGCCT GCTGGCCCTGGGCAGCCCCAGACGCA GGATCCCCTCGAGGGACC
Tal1 gap repair	CCTTTAAGTCCCTCGCGGCACCGC CCCCACTGGCAGGGCCG CCCCCC GCGCACCGGTGACCCGGGTCTTA ATTAATAAGATGATCTTCTTGAGA TCG	GATTATAAAATATTTTCCATTGAGTTA TGTAGATGCTAATTTCTCAAGATCGC CGGCGACTTAAGTCTTACCAATGCT TAATCAGTGAGG
Tal1 5' probe	CTGTGTATTGAGAATACCAAGG	TGAGCAGGACTAGGTGCGG
Tal1 3' probe	CCAAGAAAGGTGGTAAGGAAC	CTTTTCTTCCCTTGTGGAGTC

Cbf-β A A tagging	ACCCTAGTCCTGGTTCTAACTTAG GTGGCGGTGATGATCTCAAACTTC GT GCCAGCACCAAGGGCCC	CCAACTAGAAGCAGTGTGTGTAGA GATGGGGCACATAAGCTGTGCTCCAC GGATCCCCTCGAGGGACC
Cbf-β gap repair	CATGTTGATACCAGACCCATAAAG TACTAAGAATATACAGTCAAAGATA A ATTTAAATCACCGGTGACCCGG GTC	GAGGATCAAGAGTTCAAGGTCATTC TTGGCGAGTTCAAGGCCATCCTGG GATTTAAATCGCCGGCGACTTAAGTC
Cbf-β cs insert	TAGGTCTGAAGAGGAGTTTACGTC CAGCCAAGCTAGTTTGGCGCGCC TTT GAAGTGCATACCAATCAGGAC CCGC	CTAGAAGGCACAGTCGAGGCTGAT CAGCGAGCTCTAGAGAATTGATCCC C GTTCCTTCTTCACTGTCCCTTATTC
Cbf-β puro swap	TAGGTCTGAAGAGGAGTTTACGTC CAGCCAAGCTAGTTTGGCGCGCC TTT CACACATTCCACATCCACC	CTAGAAGGCACAGTCGAGGCTGAT CAGCGAGCTCTAGAGAATTGATCC CC TCAGGCACCGGGCTTGCG
Cbf-β 5′ probe	GACCCCTTGTTTTAATGCTAGC	TTTGCTTGGACCATACAGTCTC
Cbf-β 3′ probe	TACTTCCAGCTAGCAGTGGC	CAGTGCCTATACATTTCCCCC

References

1. Holland, A.J., Fachinetti, D., ***et al.*** (2012). Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc Natl Acad Sci U S A ***109***, E3350-3357.
2. Haruki, H., Nishikawa, J., ***et al.*** (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell ***31***, 925-932.
3. Mylona, A., Andrieu-Soler, C., ***et al.*** (2013). Genome-wide analysis shows that ***Ldb1*** controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis. Blood.
4. Li, L., Jothi, R., ***et al.*** (2011). Nuclear adaptor ***Ldb1*** regulates a transcriptional program essential for the maintenance of hematopoietic stem cells. Nat Immunol ***12***, 129-136.
5. Li, L., Lee, J.Y., ***et al.*** (2010). A requirement for Lim domain binding protein 1 in erythropoiesis. J Exp Med ***207***, 2543-2550.
6. Begley, C.G., Aplan, P.D., ***et al.*** (1989). The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc Natl Acad Sci U S A ***86***, 10128-10132.
7. D'Souza, S.L., Elefanty, A.G., ***et al.*** (2005). SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development. Blood ***105***, 3862-3870.
8. Lugus, J.J., Chung, Y.S., ***et al.*** (2007). GATA2 functions at multiple steps in hemangioblast development and differentiation. Development ***134***, 393-405.
9. Tsai, F.Y., and Orkin, S.H. (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood ***89***, 3636-3643.
10. Wang, Q., Stacy, T., ***et al.*** (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc Natl Acad Sci U S A ***93***, 3444-3449.
11. Weiss, M.J., Keller, G., ***et al.*** (1994). Novel insights into erythroid development revealed through ***in vitro*** differentiation of GATA-1 embryonic stem cells. Genes Dev ***8***, 1184-1197.
12. Gutierrez, L., Tsukamoto, S., ***et al.*** (2008). Ablation of ***Gata1*** in adult mice results in aplastic crisis, revealing its essential role in steady-state and stress erythropoiesis. Blood ***111***, 4375-4385.
13. Meier, N., Krpic, S., ***et al.*** (2006). Novel binding partners of ***Ldb1*** are required for haematopoietic development. Development ***133***, 4913-4923.
14. Rodriguez, P., Bonte, E., ***et al.*** (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. EMBO J ***24***, 2354-2366.
15. Soler, E., Andrieu-Soler, C., ***et al.*** (2010). The genome-wide dynamics of the binding of ***Ldb1*** complexes during erythroid differentiation. Genes Dev ***24***, 277-289.
16. Stadhouders, R., Thongjuea, S., ***et al.*** (2012). Dynamic long-range chromatin interactions control ***Myb*** proto-oncogene transcription during erythroid development. EMBO J ***31***, 986-999.
17. Lacaud, G., Gore, L., ***et al.*** (2002). ***Runx1*** is essential for hematopoietic commitment at the hemangioblast stage of development ***in vitro***. Blood ***100***, 458-466.
18. Lacaud, G., Kouskoff, V., ***et al.*** (2004). Haploinsufficiency of ***Runx1*** results in the acceleration of mesodermal development and hemangioblast specification upon ***in vitro*** differentiation of ES cells. Blood ***103***, 886-889.
19. Keller, G., Kennedy, M., ***et al.*** (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol Cell Biol ***13***, 473-486.
20. Eichmann, A., Marcelle, C., ***et al.*** (1993). Two molecules related to the VEGF receptor are expressed in early endothelial cells during avian embryonic development. Mech Dev ***42***, 33-48.

21. Eichmann, A., Corbel, C., ***et al.*** (1997). Ligand-dependent development of the endothelial and hemopoietic lineages from embryonic mesodermal cells expressing vascular endothelial growth factor receptor 2. Proc Natl Acad Sci U S A ***94***, 5141-5146.
22. Choi, K., Kennedy, M., ***et al.*** (1998). A common precursor for hematopoietic and endothelial cells. Development ***125***, 725-732.
23. Hofemeister, H., Ciotta, G., ***et al.*** (2011). Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. Methods ***53***, 437-452.
24. Mucenski, M.L., McLain, K., ***et al.*** (1991). A functional c-***myb*** gene is required for normal murine fetal hepatic hematopoiesis. Cell ***65***, 677-689.
25. Simon, M.C., Pevny, L., ***et al.*** (1992). Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. Nat Genet ***1***, 92-98.
26. Yuan, P., Han, J., ***et al.*** (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev ***23***, 2507-2520.
27. Zhao, L., Glazov, E.A., ***et al.*** (2011). Integrated genome-wide chromatin occupancy and expression analyses identify key myeloid pro-differentiation transcription factors repressed by ***Myb***. Nucleic Acids Res ***39***, 4664-4679.
28. Rouf, M.A., Vural, I., ***et al.*** (2009). Development and characterization of liposomal formulations for rapamycin delivery and investigation of their antiproliferative effect on MCF7 cells. J Liposome Res ***19***, 322-331.
29. Haeri, A., Sadeghian, S., ***et al.*** (2011). Use of remote film loading methodology to entrap sirolimus into liposomes: preparation, characterization and ***in vivo*** efficacy for treatment of restenosis. Int J Pharm ***414***, 16-27.
30. Iwase, Y., and Maitani, Y. (2012). Preparation and ***in vivo*** evaluation of liposomal everolimus for lung carcinoma and thyroid carcinoma. Biol Pharm Bull ***35***, 975-979.
31. Bird, A.W., Erler, A., ***et al.*** (2012). High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. Nat Methods ***9***, 103-109.

Chapter 5

Functional analysis of distal gene regulatory elements: Myb -81kb enhancer knock-out and inactivation of Myb intronic transcriptional attenuator element

Jorna R., Stadhouders, R., Soler, E. and Grosveld, F.

Abstract

MYB is an essential factor that controls proliferation/differentiation of hematopoietic stem/progenitor cells. It is known that an attenuator element within the first intron of *Myb* plays a role in controlling the expression of the *Myb* gene through regulation of transcriptional elongation. We have recently shown that the transition from initiation to elongation takes place around a conserved CTCF binding site and that the -81kb enhancer of the *Myb* gene may be involved as it was shown to interact with this region by 3C assays. Moreover, single nucleotide polymorphisms falling in the human equivalent of the murine -81kb enhancer affect *MYB* expression. Here I present the development of two mouse models, a conditional *Myb* -81kb enhancer knock-out and an inactivation of CTCF binding sequences within the first intron of *Myb*. The mouse models obtained allow further investigation of *Myb* expression within the relevant primary cells.

Introduction

Differentiation from the hematopoietic stem cell (HSC) to mature blood cell types occurs through a complex series of proliferation and differentiation steps. Progenitor expansion, proliferation arrest and terminal differentiation are balanced and tightly controlled in order to successfully populate the hematopoietic system. One of the players that has an important role in controlling the proliferation of stem and progenitor cells is the MYB transcription factor (TF), encoded by the ***Myb*** proto-oncogene [1-5]. Ablation of ***Myb*** is embryonic lethal around E15 due to a lack of terminal erythropoiesis [1], ***Myb*** also plays an essential role in the control of differentiation of lymphoid cells [6; 7] and the balance between proliferation and differentiation in adult HSCs [2]. Expression of ***Myb*** is tightly regulated and follows a pattern where it is highly expressed in immature progenitor blood cells, but decreases dramatically during terminal maturation of blood cells [8; 9]. MYB remains in leukemic cells, indicating ***Myb*** is important for rapidly proliferating cells [4]. And constitutive enforced ***Myb*** expression prevents differentiation and favours progenitor cell proliferation, further suggesting a role for MYB in proliferation to differentiation transition [10].

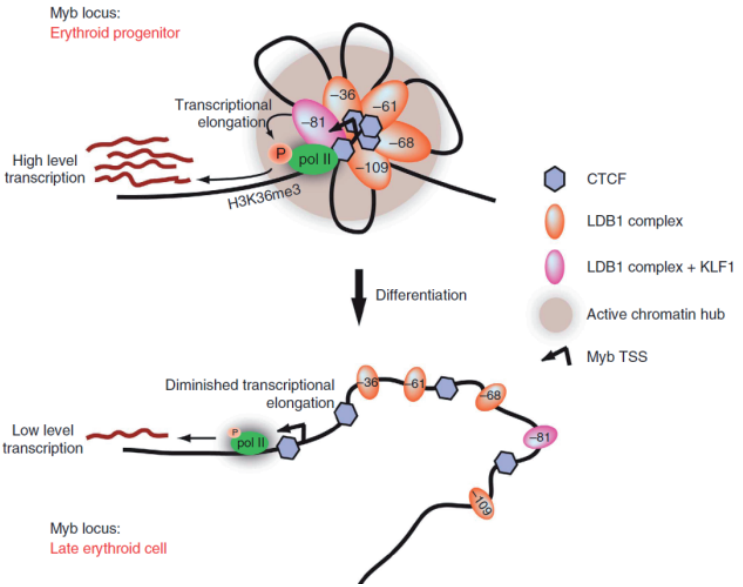


Figure 1. Model of the dynamic transcriptional regulation of ***Myb*** in differentiating erythroid cells. The ***Myb*** Active Chromatin Hub (ACH, grey sphere) is a structured nuclear compartment containing clustered ***cis***-regulatory elements enriched for activating transcription factor complexes containing transcription elongation factors (orange and pink ovals) and CTCF (blue diamonds). The ACH provides a local high concentration of RNA pol II, transcription and elongation factors around the ***Myb*** gene, allowing for high-level expression in erythroid progenitors. During differentiation, intergenic transcription factor occupancy decreases (small ovals) at the ***cis***-regulatory elements, leading to a destabilization of the ACH and a dramatic decrease of ***Myb*** transcription, allowing cells to terminally differentiate. Taken from [12].

Taken together, the data suggest that the integration of the transgene into the intergenic region disrupted ***Myb*** expression. The position where the transgene integrated, i.e. 4kb downstream of the -81kb enhancer, together with the size of the integration (>44kb) and the fact that several strong LDB1-complex binding places are present within the transgene are supportive of the idea that this enhancer plays critical role for proper ***Myb*** expression in erythropoiesis. This notion is further strengthened by the fact that single nucleotide polymorphisms falling in the human equivalent of the -81kb enhancer affect ***MYB*** expression in humans. To clarify the role of this enhancer and further investigate long-range genomic interactions within the ***Myb*** genomic locus, we deleted the -81kb enhancer and altered a double CTCF motif in the first intron of the ***Myb*** gene (figure 2A) in the mouse. The phenotypic analysis of the resulting animals is currently ongoing.

Materials and Methods

Cell culture

The conditions of mouse ES cell culture have been described in chapter 3. The specifics for gene targeting are described in chapter 4.

Recombineering, generating targeting vectors

Recombineering was performed as described in [17]. Recombineering plasmids were a kind gift of Francis Stewart (Dresden). A vector containing the mutated CTCF site was obtained using routine cloning techniques. A list of the recombineering primers used is provided in table 1.

Table 1. Recombineering primers used for generation of targeting vectors.

Primer	Sequence 5′ to 3′
F Gap repair -81	GTGCCTAATGCAGTCAGTAAACACCATCCTAAGGTAACTGAG TACAGGTATTTAAATCACCGGTGACCCGGGTC
R Gap repair -81	CAATGAATTAACTCATTAAAGATCTAATAGAGTGAGAAGAGCT GAAGTGGATTTAAATCGCCGGCGACTTAAGTC
F 3′loxp -81	GCGGTACTCGCTGGAGTGTGCTGAAGCTATTGGAGAGCATG GAGTTCTATAACTTCGTATAGCATACATTATACGAAGTTATTAGGTCT GAAGAGGAGTTTAC
R 3′loxp -81	CTCTAAGCCAGCCCATCCACTGTTGTTCTCTGGCATAACTGGAGT AAATTATAACTTCGTATAATGTATGCTATACGAAGTTATATTAAGGGTT CCGCAAGCTC
F 5′frt-neo-frt-loxp -81	GCCCTGGCCAGCTAGGGATGCACACCAAGACTCCACTTCTGGGCA CTTTTGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTAGGTCT GAAGAGGAGTTTAC
R 5′frt-neo-frt-loxp -81	AGTTTTTGCTAAAAAAATGGAGAGAATACAAAGTGTAAAAAGAA CTAAATATAACTTCGTATAATGTATGCTATACGAAGTTATGTCGA CGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCATTAAGGGTT CCGCAAGCTC
F Gap repair *CTCF*	TACTGAGCACCCTCCAAAGGCCGGATCATGAGCTCTCGAAT CCTCTGAAATTTAAATCACCGGTGACCCGGGTC
R Gap repair *CTCF*	CCCACCTGCTCCGGGATGCCTGGATCCCCGCGCGCCCTAGCCGAG GCCGATTTAAATCGCCGGCGACTTAAGTC
F intermediate	TGGAAAGTACCTTAAACATAGAATCCCCTCCCTAGTGTGTAAGATGG GATGAAGTGCATACCAATCAGGACCCGC
R intermediate	TTCCAAGCAAATTTTCCTCTGGTCAAAACTTAATCTAAAAACAAG CATCAGTTCCTTCTTCACTGTCCCTTATTC
F PmeI insert	TGGAAAGTACCTTAAACATAGAATCCCCTCCCTAGTGTGTAAGATGG GATATACGTTTAAACTGGGGGTTTG
R PmeI insert	TTCCAAGCAAATTTTCCTCTGGTCAAAACTTAATCTAAAAACAAG CATCACCCTCAGTCTATCAGTTTAAAC

F loxp-Neo-loxp	GCCAAAGTCTTGGGGTGAAATGTGGCGTGTGACTCTTGCAGATCT TGAAAATAACTTCGTATAGCATACATTATACGAAGTTATTAGGTC TGAAGAGGAGTTTAC
R loxp-Neo-loxp	CTTGTCAAAGCTGACAGAATCCTGGACTCGCCCAGCTCCCTCTCA CAGCCATAACTTCGTATAATGTATGCTATACGAAGTTATATTAAGGGT TCCGCAAGCTC

Southern blotting

Southern blot procedures are described in chapter 4. Primers used to make Southern blot probes are provided in table 2.

Table 2. Primers used to make Southern blot probes.

Primer	**Sequence 5' to 3'**
F -81 5' probe	GCAAAAACTTCCCCAAGTGAG
R -81 5' probe	CATAGCACACTGCAGGCAGCAG
F -81 3' probe	TTCTGCGAGGGAAGGCTACT
R -81 3' probe	GAGTTCTCTGTTGCTGTGGG
F ***CTCF*** 5' probe	TGGACATGCAGGAGCCAGC
R ***CTCF*** 5' probe	GAACCTCTTCTTTAAGAGTCCT
F Neo	GCTATTCGGCTATGACTGGG
R Neo	GAAGGCGATAGAAGGCGATG

Results

Targeting of *Myb* regulatory sequences and generation of a mouse model

We identified several strong candidate regulatory sequences within the ***Myb-Hbs1l*** genomic locus before [12]. The enhancer at 81kbp upstream of the ***Myb*** gene (-81kb enhancer) was specifically identified as a strong candidate, binding KLF1 and having differential dynamics upon differentiation of erythroid progenitor. The enhancer interacts with a region occupied by CTCF within the first intron of the ***Myb*** gene in erythroid progenitor cells [12]. Interestingly, the first intron of the ***Myb*** gene was known to harbor a transcriptional attenuator element [13; 14] regulating RNA pol II at the level of transcription elongation. We recently showed that the transition between transcription initiation to elongation occurs in the vicinity of a conserved CTCF site in the first intron of ***Myb***. The structural proteins CTCF and Cohesin have been identified as important factors in 3D genomic interactions [18], and CTCF may be able to interfere with transcription and regulate RNA pol II pausing [19]. To further investigate the role of these two ***Myb*** regulatory elements, the -81kb enhancer and the CTCF site, we targeted both elements by homologous recombination in mouse ES cells. The ***Myb*** genomic locus, binding by LDB1 and

CTCF and structure (ChIP-seq and 3C-seq profiles) in erythroid progenitors and control cells are shown in figure 2A. A schematic overview of the locus before and after targeting is also provided in figure 2A.

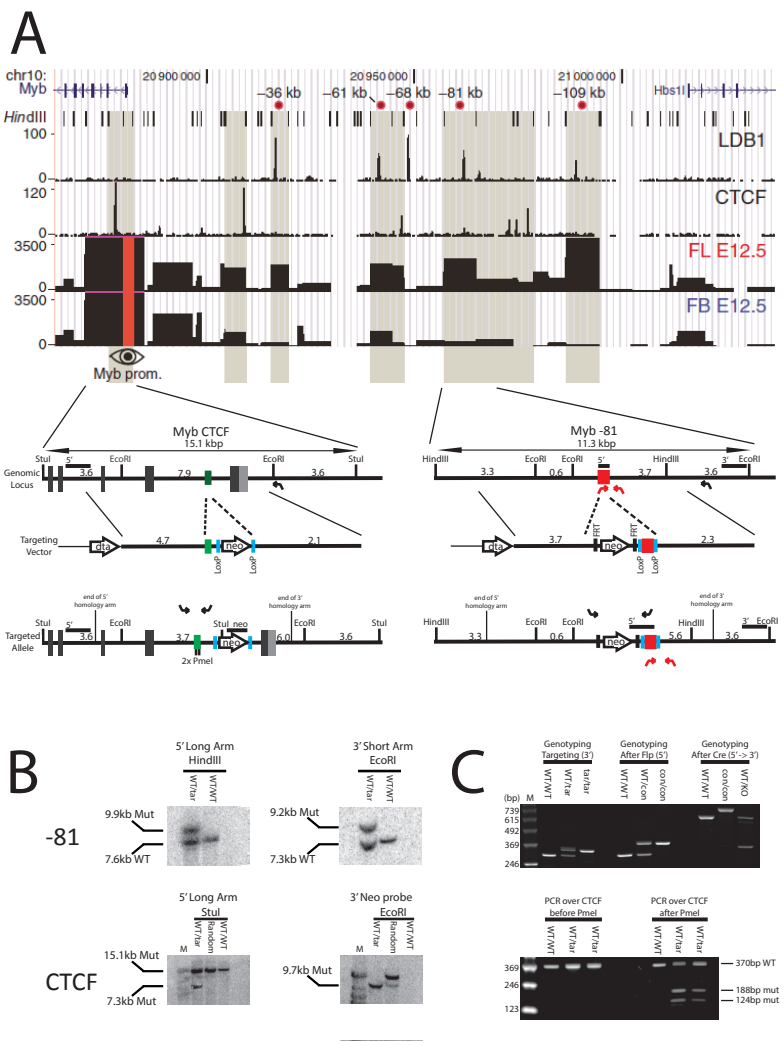


Figure 2. Long-range genomic interactions in the mouse ***Myb-Hbs1l*** locus on chromosome 10 and alleles constructed to investigate control of ***Myb*** gene expression.

A) ChIP-seq profiles for LDB1 and CTCF (MEL cells) are shown together with the 3C-seq signals obtained using the ***Myb*** promoter as viewpoint in fetal liver (FL) and fetal brain (FB, control). HindIII restriction fragments where TF binding and 3C-seq signal in fetal liver coincide are shaded grey. Position of the ***Myb*** and Hbs1l genes, HindIII restriction sites and intergenic enhancers are indicated at the top of the track.

Underneath the tracks, two separate strategies for targeting a mutation to the CTCF site in the first intron of ***Myb*** and a conditional knock-out of the -81 enhancer of ***Myb*** are shown. In dark green on the left are two tentative CTCF binding sites that are mutated to PmeI sites in the targeted allele. On the right the red box depicts the -81 enhancer 384bp sequence that is found underneath the peak in the ChIP-seq track. Distance between restriction sites in kbps is given in the numbers above the lines. The black arrow in the genomic loci indicates the place of the primer used in the screen for homologous recombination. Small flat black lines show the location of probes used in Southern Blot analysis. Other arrows depict primer pairs to identify germ line transmission in mice. Position of LoxP and frt sites are shown. Neo = G418 selectable marker, Dta = diphtheria toxin A.
B) Southern Blot analysis of homologous recombined clones of both targetings. On top is the -81 enhancer conditional knock-out and below the CTCF mutant analysis. WT = wildtype allele, mut = mutant allele, random = clone with random integration.
C) PCRs to genotype both constructs, -81 enhancer on top shows PCR identification of germ line transmission, neo removal after Flp expression and knock out of the enhancer after subsequent Cre expression. Below is the PCR analysis of the CTCF mutant: after targeting the CTCF binding sites are replaced with a PmeI restriction sites. When a PCR over the CTCF binding site/PmeI recognition site is performed, followed by PmeI digest, only a targeted allele can be digested, not the WT allele. WT = wild type allele, tar = targeted allele, con = conditional allele (after Flp expression), KO = Knock out allele (after Flp and Cre).

The -81kb enhancer was placed between loxp sites, next to a G418-resistance marker between frt sites. The resistance marker can be removed by expression of Flp recombinase and the -81kb enhancer itself by expression of Cre recombinase. We defined the -81kb enhancer of ***Myb*** as the 347bp sequence covered by the ***Ldb1*** ChIP-seq peak (see figure 2A), with the genomic coordinates ch10:21241829:21242175 (genome assembly GRCm38). Targeting of the ES cells was performed and confirmed by PCR and Southern blot (figure 2B). Two independent lines of ES cells were injected into blastocysts and one gave germline transmission. F1 mice were bred with mice that express Flp recombinase ubiquitously to remove the resistance marker. G418-resistance marker removal was monitored by PCR (figure 2C), these mice were crossed to mice harboring a Cre recombinase under the control of the interferon inducible Mx1 promoter or a ubiquitous promoter. Removal of the -81kb enhancer can also be monitored by PCR (figure 2C).

The double CTCF site within the first intron of ***Myb*** was also targeted in ES cells. The two potential CTCF DNA binding motifs were substituted for PmeI restriction sites, flanked at the 5' side by a G418 resistance marker in between loxp sequences. The G418 resistance marker is located at the 5' side of the genomic coordinates chr10:21158074:21158433 (genome assembly GRCm38), the two putative CTCF binding sites that are mutated are within these genomic coordinates. Targeting was confirmed by PCR and Southern blotting (figure 2B). An additional confirmation of proper targeting was performed using a PCR spanning the (former) CTCF site, followed by PmeI digest. The wild type genomic sequence cannot be cut by PmeI and yields a 370bp band, while the targeted sequence is cut twice and falls apart in three bands of 188, 124 and 58bp (see figure 2C, 188 and 124bp bands are visible).

Discussion

Enhancer knock-outs

Conditionally knocking out enhancers is a definitive way to determine the effect of a given enhancer on the transcription of a gene. Relatively few examples in the literature exist of how the knock out of a gene-specific enhancer, without disturbing the gene itself, can lead to a phenotype in mice [20-23]. However, one has to be careful extrapolating results obtained in cells or cell lines to whole organisms. Regardless of the strength of the evidence obtained with other experiments, it has been shown in some instances that knocking out an enhancer ***in vivo*** only has minor effects on gene expression. One of the most striking examples is the +19 enhancer of the ***Scl/Tal*** gene, which gives clear early hematopoietic LacZ staining in embryos when coupled to a SV40-LacZ construct [24], but the +19kb Scl enhancer knock out did not yield a striking phenotype [25]. The ***Myb*** enhancer knock out described here is ready to be tested in cells directly harvested from adult mice or mouse embryos and should give a definite view on the function of the ***Myb*** -81kb enhancer and the CTCF binding site within the first intron of ***Myb***. The choice of these elements was driven by observations that small genomic variations at the -81kb enhancer in humans (SNPs) lead to altered ***MYB*** expression in humans, suggesting that this enhancer plays a key role in ***Myb*** regulation. The intronic CTCF site was chosen based on early reports showing that ***MYB*** first intron harbors a transcriptional attenuator element, which is affected in human leukemia [13; 14], and based on our recent data [12] which showed that RNA pol II transcription elongation is switched on at the intronic CTCF site.

Model of *Myb* expression in erythroid progenitors

The model of ***Myb*** expression that was suggested in previous work (figure 1, [12]) can be improved with the data obtained with this mouse model. The first question addressed will be whether a phenotype is present within the erythroid compartment of mice without the -81kb enhancer or without a CTCF binding site within the first intron of ***Myb***. Mice that lack ***Myb*** seem to exhibit problems in the hematopoietic stem cell compartment, as only few blood cells with primitive appearance can be found at E14 throughout the embryo [1]. It will be interesting to see whether these new mouse models produce hematopoietic stem cells. The main questions on the molecular side will concern maintenance of ***Myb*** transcription in erythroid progenitor cells. Will stem/progenitor cells of the hematopoietic compartment still be able to express high levels of ***Myb***? If not, are all hematopoietic lineages expressing ***Myb*** affected by the enhancer deletion or will it affect the erythroid lineage only? What is the influence on the identity of cells and what happens if the -81kb enhancer is specifically deleted in erythroid progenitor cells? When ***Myb*** transcripts are no longer present or transcription is clearly inhibited after deletion of the -81kb enhancer, it will be interesting to see whether transcription is hampered during transcription initiation or elongation. The last questions will concern the 3D confirmation of the ***Myb-Hbs1l*** locus. Will the ACH still form when the -81kb enhancer or the CTCF binding sites are absent? Lastly, these mouse models could also be used to address whether loop

formation within the ***Myb*** locus is a cause or consequence of transcription of the ***Myb*** gene. It has been shown recently by an elegant study of the Blobel and collegues that transcription of the ***β-globin*** gene could be induced by tethering LDB1 with a zinc finger to an enhancer site where normally the LDB1-complex binds [26]. Their results suggest that loop formation causes gene transcription for this particular gene.

References

1. Mucenski, M.L., McLain, K., ***et al.*** (1991). A functional c-***myb*** gene is required for normal murine fetal hepatic hematopoiesis. Cell ***65***, 677-689.
2. Lieu, Y.K., and Reddy, E.P. (2009). Conditional c-***myb*** knockout in adult hematopoietic stem cells leads to loss of self-renewal due to impaired proliferation and accelerated differentiation. Proc Natl Acad Sci U S A ***106***, 21689-21694.
3. Vegiopoulos, A., Garcia, P., ***et al.*** (2006). Coordination of erythropoiesis by the transcription factor c-***Myb***. Blood ***107***, 4703-4710.
4. Ramsay, R.G., and Gonda, T.J. (2008). ***MYB*** function in normal and cancer cells. Nat Rev Cancer 8, 523-***534***.
5. Sandberg, M.L., Sutton, S.E., ***et al.*** (2005). c-***Myb*** and p300 regulate hematopoietic stem cell proliferation and differentiation. Dev Cell ***8***, 153-166.
6. Thomas, M.D., Kremer, C.S., ***et al.*** (2005). c-***Myb*** is critical for B cell development and maintenance of follicular B cells. Immunity ***23***, 275-286.
7. Bender, T.P., Kremer, C.S., ***et al.*** (2004). Critical functions for c-***Myb*** at three checkpoints during thymocyte development. Nat Immunol ***5***, 721-729.
8. Emambokus, N., Vegiopoulos, A., ***et al.*** (2003). Progression through key stages of haemopoiesis is dependent on distinct threshold levels of c-***Myb***. EMBO J ***22***, 4478-4488.
9. Gonda, T.J., and Metcalf, D. (1984). Expression of ***myb***, myc and fos proto-oncogenes during the differentiation of a murine myeloid leukaemia. Nature ***310***, 249-251.
10. Clarke, M.F., Kukowska-Latallo, J.F., ***et al.*** (1988). Constitutive expression of a c-***myb*** cDNA blocks Friend murine erythroleukemia cell differentiation. Mol Cell Biol ***8***, 884-892.
11. Soler, E., Andrieu-Soler, C., ***et al.*** (2010). The genome-wide dynamics of the binding of ***Ldb1*** complexes during erythroid differentiation. Genes Dev ***24***, 277-289.
12. Stadhouders, R., Thongjuea, S., ***et al.*** (2012). Dynamic long-range chromatin interactions control ***Myb*** proto-oncogene transcription during erythroid development. EMBO J ***31***, 986-999.
13. Bender, T.P., Thompson, C.B., ***et al.*** (1987). Differential expression of c-***myb*** mRNA in murine B lymphomas by a block to transcription elongation. Science ***237***, 1473-1476.
14. Reddy, C.D., and Reddy, E.P. (1989). Differential binding of nuclear factors to the intron 1 sequences containing the transcriptional pause site correlates with c-***myb*** expression. Proc Natl Acad Sci U S A ***86***, 7326-7330.
15. Suzuki, M., Yamazaki, H., ***et al.*** (2013). Disruption of the Hbs1l-***Myb*** locus causes hereditary persistence of fetal hemoglobin in a mouse model. Mol Cell Biol ***33***, 1687-1695.
16. Mukai, H.Y., Motohashi, H., ***et al.*** (2006). Transgene insertion in proximity to the c-***myb*** gene disrupts erythroid-megakaryocytic lineage bifurcation. Mol Cell Biol ***26***, 7953-7965.
17. Hofemeister, H., Ciotta, G., ***et al.*** (2011). Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. Methods ***53***, 437-452.
18. Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet ***12***, 283-293.
19. Shukla, S., Kavak, E., ***et al.*** (2011). ***CTCF***-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature ***479***, 74-79.
20. Juan, A.H., and Ruddle, F.H. (2003). Enhancer timing of Hox gene expression: deletion of the endogenous Hoxc8 early enhancer. Development ***130***, 4823-4834.
21. Berland, R., Fiering, S., ***et al.*** (2010). A conserved enhancer element differentially regulates developmental expression of CD5 in B and T cells. J Immunol ***185***, 7537-7543.
22. Pinaud, E., Khamlichi, A.A., ***et al.*** (2001). Localization of the 3' IgH locus elements that effect long-

distance regulation of class switch recombination. Immunity ***15***, 187-199.

23. Hug, B.A., Wesselschmidt, R.L., ***et al.*** (1996). Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. Mol Cell Biol ***16***, 2906-2912.
24. Sanchez, M., Gottgens, B., ***et al.*** (1999). An SCL 3' enhancer targets developing endothelium together with embryonic and adult haematopoietic progenitors. Development ***126***, 3891-3904.
25. Spensberger, D., Kotsopoulou, E., ***et al.*** (2012). Deletion of the Scl +19 enhancer increases the blood stem cell compartment without affecting the formation of mature blood lineages. Exp Hematol ***40***, 588-598 e581.
26. Deng, W., Lee, J., ***et al.*** (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell ***149***, 1233-1244.

Chapter 6

General Discussion

A key goal of cell biological research is to determine protein function and one of the most straightforward ways to do this is by depleting the target protein from a cell and determine its phenotype. The development of embryonic stem cells (ES) and genome editing technologies made it possible to obtain ES cells that lack a functional gene of a specific protein, some quarter of a century ago [1]. This development made it possible to obtain mouse models lacking a specific protein. Such mice are called knock out mice and have been an invaluable source of information for cell biologists. The process of gene targeting has been optimized continuously, but particularly the introduction of recombineering technology [2] has made it possible to obtain targeting constructs much quicker. Nevertheless targeting genes is still considered to be laborious.

The development of RNAi mediated knock down techniques provided a much less laborious way to investigate protein depletion [3; 4]. Transfection or viral transduction of esiRNA or shRNA suffices to knock down the level of RNA and hence the protein of a target gene. Libraries that contain sh/esiRNA of all genes were constructed to allow testing of the knock down of a specific gene fast, or set up high-throughput screenings of gene knock downs. Although RNAi mediated knock down provides an easy way to study protein disruption, the extent of the knock down differs per target and is often not complete. Moreover, both the knock out and knock down strategy rely on the cell's proteases to clear the protein, and hence the half-life of a protein is the key determinant of how fast the depletion occurs. The half-life of proteins differs widely, but in practice this means that gene expression analysis usually cannot be performed within 48 hours after initial knock down/out. Within this time the primary targets of a gene will be differentially regulated, but their targets (so-called secondary targets) will also be affected. This leads to a list with differentially regulated genes that has a substantial portion of genes that are there due to secondary effects, which may be as high as 70% (unpublished results, [5; 6]). In order to get a better view on the true primary targets of a target gene, a faster way to perform knock down/out is needed.

Anchor-Away as a new method to disrupt protein function

Chapter 3 and 4 describe the development of Anchor-Away, a new technique to disrupt protein function in mammalian cells. AA was developed in yeast [7] and provides a way to displace target nuclear proteins to the cytoplasm, where they are unable to perform their nuclear function. Target proteins of this study are transcription factors involved in hematopoietic development in general and erythropoiesis in particular. AA works through ternary complex formation between rapalog and two rapalog binding domains, one fused to the target nuclear factor (TF here) and the other one fused to the ribosomal protein L13 (rpL13). RpL13 is incorporated into maturing ribosomes and thus the TF is tethered to ribosomal subunits in the nucleus after addition of the rapalog. There is constant flow of maturing ribosomes from nucleoli to the cytoplasm through the nucleus [8], taking the TF to the cytoplasm. The system works in HEK293T within 45 minutes and in mouse ES cells within four hours by adding the rapalog to culture medium. In

differentiated ES cells AA also works within four hours, but the rapalog needs to be targeted to the cells by cationic liposomes (**chapter 3**) to allow rapalog to enter the cells.

Clarifying the reason rapalog is not able to enter differentiated cells, while it can enter ES cells has not been pursued in this work. It is a surprising observation, however. It is likely that the obstruction lies in entrance of rapalog into the cells since I was able to restore AA activity by targeting the rapalog to the differentiated cells using liposomes. This notion is further supported by AA experiments conducted for separate TFs (LDB1, GATA1, MYB); no AA activity was observed in differentiated cells unless the rapalog was targeted to the cells by liposomes for any of these targets. Experiments on the uptake of rapamycin in cells have been performed, and suggest this happens through passive transport (diffusion) [9]. Together these results suggest that there is a difference in the plasma membrane probably in the lipid composition rather than in the absence or presence transmembrane transporters within the membrane. This suggests that the composition of the cellular membrane changes in the course of differentiation, stopping rapalog to diffuse through the membrane. Indeed, some studies on lipid composition of the plasma membrane of human ES cells and differentiated cells have been published and a difference in lipid composition was reported [10; 11].

Chapter 4 describes work done to test AA for different targets by performing knock in of the AA tag to different TFs. An ES cell line with the AA tag targeted to rpL13 was also obtained for the purpose of generating a ***rpL13-AA*** mouse. Lastly, a basic vector that can be used to generate targeting vectors for C-terminal tagging of target factors has been produced. Generation of targeting constructs can be fast and straightforward by recombineering [12], a technique I introduced in our laboratory and this is an asset for anyone considering gene targeting. The resources developed during these studies shorten the time needed to obtain a homozygous target-AA ES cell and/or mouse model. Similar to a regular knock out, one round of targeting of the gene of interest will be necessary (two if a homozygous ES cell is needed). Another aim is to compare AA and RNAi mediated protein depletion.

Anchor-Away compared to other methods to study protein function

AA is a fast way to disturb protein function, faster than can be achieved by RNAi knock down or traditional knock out. Because of its speed AA may allow dissection of primary and secondary effects of the knock out for the first time. This will be especially interesting in the case of TFs, since they are able to quickly influence gene expression programs in small developmental windows. When interested in the effects of one or few nuclear factors on gene expression programs a rapid way to remove the factor is desirable, hence AA may be the best choice for such studies. AA may also be an interesting choice to obtain a conditional model to disrupt target protein function in primary cells, as these cells tend to be sensitive to viral transduction and transfection.

AA is not a technique amenable to high-throughput screening. As a first test for gene disruption or for high-throughput screenings, RNAi mediated knock down will be the method of choice. However, if more knowledge of a specific factor or few factors is required AA will be a good alternative. The amount of manual labor in the laboratory is comparable to what is needed to obtain a (conditional) knock out. This makes AA also a good alternative if primary cells are obtained from mice to perform knock down ***ex vivo***. In this way, cells that cannot be obtained from a conditional knock out mouse model, e.g. when primary cells are too delicate for viral transduction and an appropriate Cre deleter mouse is absent, can potentially be investigated as well. However, conditional knock out mouse models can be favorable when knock out within organs ***in vivo*** is the objective as at this moment it is unclear whether rapalog delivery in live animals can be achieved. Liposome delivery to specific cells in mouse models has been reported and may be an avenue worth exploring in future [13; 14]. Tables 1 gives an overview of the different aspects of AA and knock down/out.

Table 1. Comparison of AA with RNAi knock down and knock out

	Anchor-Away	RNAi knock down	(conditional) knock out
Influences	Protein	RNA	DNA
Prerequisite	Targeting of target gene in (ES) cells that are expressing rpL13-AA	siRNA/shRNA construct that must be transfected /transduced in relevant cells	Targeting of target gene in (ES) cells
Amount of protein depleted	To be determined for every cell, results generated until now indicate close to 100%	To be determined for every esiRNA/shRNA used, varies from almost nothing to almost complete	100%
Time to depletion	≤4 hours	≥48 hours	≥48 hours
Compatibility to primary cells	Only for cells that can be targeted by rapalog. Liposome delivery is in theory possible with all cells (but needs testing for every specific cell)	Only cells that are amenable to electroporation or viral transduction	Yes, as mice can be crossed with inducible Cre knock-ins (like ERt2-Cre)
Amenability to high-throughput	No	Yes	No
Key advantages	Specificity, fast knock-out enables in-depth analysis of the phenotype, possibility to perform experiments in short-lived cell populations (e.g. hemangioblast)	Easy to set up, fast answer, high-throughput screenings are possible	Specificity, large amounts of definitive knock-out cells can be acquired (no need to induce)
Key disadvantages	Laborious	Possible off-target effects, secondary effects, system may not reach enough knock down for phenotype	Laborious, secondary effects
Adviced when	'Deep' knowledge of a protein depletion or depletion in sensitive primary cells	High-throughput screens, confirmation of few genes is required	(Conditional) Knock out of genes in live mice

Further development of AA

AA as described here can be used to disrupt protein function of nuclear factors. However with relatively small adjustments AA could be used to disrupt cytoplasmic factors as well, the anchor would have to be a protein that is transported in a big flow to the nucleus from the cytoplasm. An example of such proteins could be individual members of the RNA polymerase II complex. Another avenue to be explored is the development of 3A (Advanced Anchor-Away), where the protein disrupt during AA can be restored through addition of 4-hydroxytamoxifen (4-OHT) to cells. In order to achieve 3A the target wild-type or mutant protein of AA is fused to ER^{T2}, a triple mutated estrogen receptor. ER^{T2} does not bind estrogen at physiological levels, however it does bind 4-OHT. Proteins fused to ER^{T2} are restricted to the cytoplasm because ER^{T2} interacts with a heat shock protein complex, however upon ligand binding this interaction is lost and ER^{T2} with its fused protein relocates to the nucleus (see e.g. [15; 16]). The opportunity to attach a mutant factor to ER^{T2} could make this system especially interesting: during development at any time the wild-type target could be disrupted and a mutant form of the factor could be put in place. 3A could allow study of factors that are slightly altered instead of completely disrupted in relevant cells. Some effort has been undertaken to develop 3A (data not shown), but the full development of this system will take more work.

Study of long-range chromatin interactions

Long-range chromatin interactions are interactions by distant sites on the chromatin. They take place to fold the genome into 3D topological domains, needed to compact the genome enough to fit into the nucleus of the cell, but also to regulate accessibility and activity of different regions of the genome. The proteins CTCF and Cohesin are indicated to play an important role in this process ([17-19]; personal communication K.Wendt). Long-range chromatin interactions can also take place within topological domains, for instance to regulate gene transcription. This is also referred to as (chromatin) looping. Looping typically occurs through physical interactions of ***cis***-regulatory elements (enhancers) with the gene body of a specific gene. It is mediated by TFs, although CTCF and Cohesin also play a role in these interactions [20-23]. There is evidence that the interactions of enhancers often take place with the first intron of genes, mediating transcription elongation rather than transcription initiation at the promoter [24-26]. These long-range chromatin interactions are reviewed in **chapter 2**. 3C-Seq is a good way to determine long-range chromatin interactions from a specific viewpoint [24; 27; 28] and gives a good indication of enhancer activity. However, to definitively determine the function of ***cis***-regulatory elements ***in vivo*** a (conditional) knock out or mutation strategy in mouse ES cells has to be taken. **Chapter 5** describes the development of an -81kb enhancer conditional knock out model for the ***Myb*** gene. The development of another mouse model where CTCF binding sites within the first intron of ***Myb*** are mutated is also described.

Apart from studying chromatin interactions within the ***Myb-Hbs1l*** genomic locus and ***Myb*** expression in particular, the mouse models may also be used to study more general questions about long-range chromatin interactions. One of the long-standing questions in the field has been whether interaction is a cause or consequence of gene expression. A study where the TF LDB1 was tethered to an enhancer using zinc fingers to artificially induce ***β-globin*** expression in erythroid cells suggests looping causes gene expression [29]. These mouse models can be used to add to this knowledge by looking at loss of an interaction. Moreover, by mutating the CTCF sites, this model can also be used to study the influence of long-range chromatin interactions on transcription elongation of the ***Myb*** gene.

Future work on long-range chromatin interactions

In **chapter 1** I state that a combination of recombineering techniques and linear integration of bacterial artificial chromosomes (BACs) by transposases into the genome of fertilized oocytes [30] or ES cells followed tetraploid complementation [31] harbors a promise of elegant experiments that can be done in future. This is especially true for study of long-range chromatin interactions: the presented enhancer knock out strategy is chosen to investigate the function of the -81kb enhancer of ***Myb*** in animals. In this strategy we choose to delete a stretch of 347bp, although we hypothesize that the core element vital for binding is the half E-box-GATA motif ((C)TGN$_{7\text{-}8}$WGATAR). The reason for this is that obtaining a mouse model is laborious: after successful targeting of the gene one has to go through an elaborate breeding scheme to obtain germ line transmitted animals, remove the selectable marker, cross in the right Cre and making sure all controls are also there (e.g. heterozygous animals or littermates without Cre). To maximize the chance that we are looking at complete knock out of enhancer function, we decided to delete the complete DNA sequence below the ChIP-seq peak of LDB1 in fetal liver cells (**chapter 5**).

Recombineering allows making (point) mutations to non-coding parts of genomic loci in BACs **and** fusion of the gene of interest with a reporter gene (e.g. fluorescent proteins). Transposons like PiggyBac have been shown to be able to integrate BACs linearly into a target genome efficiently. The combination with rapid ways to obtain mouse models of the genomic loci, tetraploid complementation of ES cells or mutation of the genome of fertilized oocytes, allows the possibility to test mouse models fast. In such a set-up one can investigate gene expression in relevant cells without the need to remove the selectable marker or breed homozygous animals. In this much shorter time line it becomes more attractive to look at the effects of subtle mutations on gene expression. New generation genome-editing techniques are interesting candidates in this respect as well: development of the zinc finger nuclease, TALE nuclease and CRISPR systems show that there is a potential to allow fast homozygous genome of ES cells or even fertilized oocytes. This is another avenue to be explored in the future.

References

1. Kuehn, M.R., Bradley, A., ***et al.*** (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature ***326***, 295-298.
2. Copeland, N.G., Jenkins, N.A., ***et al.*** (2001). Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet ***2***, 769-779.
3. Fire, A., Xu, S., ***et al.*** (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature ***391***, 806-811.
4. Elbashir, S.M., Harborth, J., ***et al.*** (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature ***411***, 494-498.
5. Yuan, P., Han, J., ***et al.*** (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev 23, 2507-2520.
6. Zhao, L., Glazov, E.A., ***et al.*** (2011). Integrated genome-wide chromatin occupancy and expression analyses identify key myeloid pro-differentiation transcription factors repressed by Myb. Nucleic Acids Res ***39***, 4664-4679.
7. Haruki, H., Nishikawa, J., ***et al.*** (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell ***31***, 925-932.
8. Alberts, B., Johnson, A., ***et al.*** (2002). Molecular biology of the cell, 4th edn (New York: Garland science).
9. Zhu, W., Masaki, T., ***et al.*** (2009). In-vitro Release of Rapamycin from a Thermosensitive Polymer for the Inhibition of Vascular Smooth Muscle Cell Proliferation. Journal of bioequivalence & bioavailability ***1***, 3-12.
10. Bosman, G.J., Boer, P., ***et al.*** (1982). Changes in the glycoprotein composition of plasma membrane during the differentiation of friend erythroleukemia cells. Biochimica et biophysica acta ***693***, 444-450.
11. Liang, Y.J., Yang, B.C., ***et al.*** (2011). Changes in glycosphingolipid composition during differentiation of human embryonic stem cells to ectodermal or endodermal lineages. Stem Cells ***29***, 1995-2004.
12. Hofemeister, H., Ciotta, G., ***et al.*** (2011). Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. Methods ***53***, 437-452.
13. Lasic, D.D. (1993). Liposomes from physics to applications (Amsterdam, The Netherlands: Elsevier).
14. Koning, G.A., Eggermont, A.M., ***et al.*** (2010). Hyperthermia and thermosensitive liposomes for improved delivery of chemotherapeutic drugs to solid tumors. Pharm Res ***27***, 1750-1754.
15. Vallier, L., Mancip, J., ***et al.*** (2001). An efficient system for conditional gene expression in embryonic stem cells and in their ***in vitro*** and ***in vivo*** differentiated derivatives. Proc Natl Acad Sci U S A ***98***, 2467-2472.
16. Glass, C.K., and Ogawa, S. (2006). Combinatorial roles of nuclear receptors in inflammation and immunity. Nature reviews Immunology ***6***, 44-55.
17. Phillips, J.E., and Corces, V.G. (2009). ***CTCF***: master weaver of the genome. Cell ***137***, 1194-1211.
18. Nativio, R., Wendt, K.S., ***et al.*** (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet ***5***, e1000739.
19. Garrick, D., De Gobbi, M., ***et al.*** (2009). ***CTCF***, cohesin and higher-order chromatin structure. Epigenomics ***1***, 232.
20. Wendt, K.S., and Peters, J.M. (2009). How cohesin and ***CTCF*** cooperate in regulating gene expression. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology ***17***, 201-214.
21. Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet ***12***, 283-293.
22. Ribeiro de Almeida, C., Stadhouders, R., ***et al.*** (2011). The DNA-binding protein ***CTCF*** limits proximal

Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus. Immunity ***35***, 501-513.

23. Splinter, E., Heath, H., ***et al.*** (2006). ***CTCF*** mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev ***20***, 2349-2354.
24. Stadhouders, R., Thongjuea, S., ***et al.*** (2012). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. EMBO J ***31***, 986-999.
25. Jing, H., Vakoc, C.R., ***et al.*** (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. Mol Cell ***29***, 232-242.
26. Krumm, A., Hickey, L.B., ***et al.*** (1995). Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. Genes Dev ***9***, 559-572.
27. Ribeiro de Almeida, C., Stadhouders, R., ***et al.*** (2012). DNA-binding factor ***CTCF*** and long-range gene interactions in V(D)J recombination and oncogene activation. Blood ***119***, 6209-6218.
28. Stadhouders, R., Kolovos, P., ***et al.*** (2013). Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nature protocols ***8***, 509-524.
29. Deng, W., Lee, J., ***et al.*** (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell ***149***, 1233-1244.
30. Rostovskaya, M., Naumann, R., ***et al.*** (2013). Transposon mediated BAC transgenesis via pronuclear injection of mouse zygotes. Genesis ***51***, 135-141.
31. Misra, R.P., Bronson, S.K., ***et al.*** (2001). Generation of single-copy transgenic mouse embryos directly from ES cells by tetraploid embryo complementation. BMC biotechnology ***1***, 12.

Chapter 7

Summary / Samenvatting

Summary

Mature lineages of cells are produced from stem/progenitor cells through multiple steps of differentiation and proliferation. This is an important concept in the development of multicellular organisms, but also in adult life. An example of such a process is hematopoiesis, blood or hematopoietic stem cells (HSCs) that reside in the bone marrow in adult life are continuously differentiated to generate mature blood cell lineages. Examples of these mature lineages are erythrocytes, megakaryocytes and lymphocytes. Transcription factors (TFs) are proteins that are able to influence gene expression programs by binding to genomic sequences and interacting with the transcription machinery, inhibiting or enhancing transcription on target genes.

These so-called long-range chromatin interactions mediated by TFs are reviewed in **chapter 2**. Several experiments combined suggest that long-range chromatin interactions take place on a specific genomic locus:

1. TF binding is confirmed by ChIP(-seq) on target enhancer sequence.
2. Confirmation of enhancer activity in a luciferase assay.
3. Gene expression analysis after knock-out or knock-down of the TF influences transcription of the specific gene where binding occurs.
4. 3C(-seq) analysis confirms a higher than expected physical vicinity of an enhancer sequence and the gene body (promoter or 1st intron of the gene mostly).

Based on the experiments we and other laboratories have conducted, we suggest a model that enhancer interaction of at least a sub-class of enhancers with the target gene is specifically enhancing transcription elongation.

A way to assess enhancer activity in a more definitive way is described for the ***Myb-Hbs1l*** genomic locus in **chapter 5**. A candidate enhancer, the mouse -81kb enhancer of the ***Myb*** gene, and a CTCF binding site predicted to function as an attenuator element within the 1st intron of the ***Myb*** gene are knocked out and mutated, respectively. Knocking out the -81kb enhancer of ***Myb*** is predicted to have an inhibiting effect on expression of the ***Myb*** gene in erythrocyte progenitors. This study will dig further into the effect of removal of regulatory sequences within the ***Myb-Hbs1l*** genomic region. We aim to disturb TF dynamics in relevant cells and strictly define functions of TF binding on the -81kb enhancer and ***CTCF*** binding to the first intron of ***Myb***. Although knocking out enhancers is the most definitive way to determine enhancer function, obtaining mouse models of enhancer knock-outs is time consuming and expensive. Therefore it is important to try and improve the results obtained with the other experiments. Current available knock down or out techniques in mammalian cells take at least 48 hours to eliminate gene function, leading to many genes being differentially regulated by secondary effects.

In **chapter 3** we describe the Anchor-Away (AA) technique in mammalian cells, a novel way to disturb TF function. AA was earlier developed in yeast and aims to disrupt TF function on the protein rather than the DNA (knock out) or RNA (knock down) level. One rapalogue-binding domain is fused to the ribosomal protein L13 (rpL13), the anchor, and another rapalogue-binding domain is fused to the target TF. The domains are confirmed not to interfere with the normal TF and rpL13 localization in regular medium. After addition of rapalogue to the medium, but not the solvent ethanol, the TF shifts localization from the nucleus to the cytoplasm within 45 minutes in HEK293T cells. The reason for this is that ternary complex formation between rpL13 and the TF has taken place. The rpL13 is incorporated into a maturing ribosome that is transported from the nucleus to the cytoplasm. The TF is captured by the anchor and dragged along with the ribosome to the cytoplasm. The TF is no longer capable of exerting its normal function in the cytoplasm, and so is functionally knocked out.

That AA also works in relevant cells is described in **chapter 4**. By either knocking in AA tags to TFs in the genome or re-expressing the specific TF in a knock-out background we obtained mouse ES cells that express only ***Ldb1-AA*** or ***Gata1-AA***, along with the ***rpL13-AA***. AA activity in these ES cells was confirmed by adding rapalogue to the medium. The targeted ES cells were differentiated toward hematopoietic lineages in culture and it was found that AA did not work after adding rapalogue to the medium. However, we were able to restore AA after rapalogue was delivered to the cells by liposomes. These cells are now ready to be tested and compared with the traditional ways of knocking-down or knocking-out genes.

Samenvatting

Stam/voorlopercellen kunnen via verschillende stappen van differentiatie en proliferatie definitieve celtypes vormen. Dit is een belangrijk principe in de ontwikkeling van multicellulaire organismen, maar ook tijdens het volwassen leven van deze organismen. Een voorbeeld van zo'n proces in het volwassen leven is het hematopoëtisch systeem. De bloed- of hematopoëtische stamcellen (HSCs) bevinden zich in het beenmerg en moeten continue gedifferentieerd worden om nieuwe definitieve bloedcellen te blijven vormen. Voorbeelden van definitieve bloedcellen zijn erytrocyten (rode bloedlichaampjes), megakaryocyten (bloedplaatjes) en lymphocyten (witte bloedlichaampjes). Transcriptiefactoren (TFs) spelen een belangrijke rol tijdens de differentiatie/proliferatie stappen die nodig zijn om definitieve bloedcellen te maken van HSCs. TFs zijn eiwitten die snel genexpressie programma's aan of uit kunnen zetten. Ze doen dit door aan hun herkenningssequenties op het DNA te binden en interacties aan te gaan met de transcriptie machinerie in de cel. Daardoor stimuleren of inhiberen TFs transcriptie van hun target genen.

De herkenningssequenties van TFs kunnen duizenden baseparen weg liggen van waar de transcriptie machinerie gebonden is en genen transcribeert. Dit betekent dat delen van het genoom die relatief ver uit elkaar liggen contact moeten maken, zodat de TFs hun invloed kunnen uitoefenen. Deze interacties worden chromatine interacties over lange afstand genoemd en zij kunnen met behulp van TFs tot stand komen. Dit soort chromatine interacties over lange afstand zijn het onderwerp in **hoofdstuk 2**. Verschillende experimenten kunnen gezamenlijk een sterke aanwijzing vormen dat chromatine interacties over lange afstand plaats vinden binnen een specifiek genomisch locus:

1. De binding van TFs wordt bevestigd door ChIP(-seq) op een specifieke enhancer.
2. Bevestiging van enhancer activiteit in een luciferase test.
3. Analyse van gen expressie na verwijdering van de specifieke TF die op de enhancer bindt, wijst uit dat transcriptie van het gen waar de TF bindt beïnvloed is.
4. 3C(-seq) analyse bevestigt dat de enhancer en een deel van het gen (meestal promotor of het eerste intron van het gen) vaker dicht bij elkaar in de buurt zijn dan verwacht kan worden op basis van de afstand.

Gebaseerd op deze experimenten die in ons laboratorium gedaan zijn en andere experimenten die in andere laboratoria werden gedaan, stellen wij een model voor. Dit model behelst dat op zijn minst een deel van de chromatine interacties over lange afstand tussen enhancers en het eerste intron van een gen specifiek transcriptie elongatie stimuleert tijdens het contact dat middels TFs tot stand komt.

Een manier om een meer definitief antwoord te krijgen op wat voor invloed enhancer activiteit op transcriptie heeft van het ***Myb*** gen in de ***Myb-Hbs1l*** genomisch locus is beschreven in **hoofdstuk 5**. Een mogelijke enhancer in de muis, de -81kb enhancer van het ***Myb*** gen, en een CTCF bindingsplaats waarvan voorspeld wordt dat deze als transcriptie-dempend element werkt in het eerste

intron van ***Myb*** zijn respectievelijk verwijderd en gemuteerd in embryonale stamcellen. Vervolgens werden muismodellen van deze embryonale stamcellen gemaakt. Wij voorspellen dat het verwijderen van de -81kb enhancer een negatieve invloed heeft op het transcriptie niveau van ***Myb*** in voorlopercellen van de erytrocyt. In ieder geval zal deze studie kennis van het verwijderen van regulerende sequenties in de ***Myb-Hbs1l*** locus verdiepen. Het doel is om door de verwijdering van de -81kb enhancer en de CTCF bindingsplaats de dynamiek van TFs in bloedcellen te verstoren en daardoor een striktere definitie van de functie van deze regulatoire sequentie te kunnen geven. Dit soort studies, waarbij de specifieke sequenties uit het genoom verwijderd worden, geven het meest definitieve antwoord. Maar muismodellen maken en bijhouden kost veel tijd en geld. Daarom is het belangrijk om andere experimenten te verbeteren. De huidige manieren om dit te doen, conditionele knock out (het verwijderen uit het genoom) of knock down (het mRNA van een specifiek gen verwijderen met RNAi technieken), hebben als nadeel dat analyse van gen expressie op zijn vroegst 48 uur na het begin van de techniek plaats kan vinden. Dit heeft tot gevolg dat niet alleen de expressie niveaus van de target genen van een TF beïnvloed zijn, maar ook de expressie niveaus van andere genen die beïnvloed werden door de target genen in plaats van de specifieke TF. Dit heten secundaire effecten van de knock down of knock out.

In **hoofdstuk 3** beschrijven we de Anchor-Away (AA) techniek in zoogdiercellen, een nieuwe manier of de functie van TFs te verstoren. AA werd eerder ontwikkeld voor gisten en heeft als doel de functie van TFs uit te schakelen op eiwit niveau, in plaats van op DNA (knock out) of RNA (knock down) niveau. Eén rapaloog-bindend domein wordt gefuseerd aan het ribosomale eiwit L13 (rpL13), dit fungeert als het anker. Een ander rapaloog-bindend domein wordt gefuseerd aan de TF die men verwijderen. De eerste stap was bevestigen dat deze domeinen geen invloed uitoefenen op de normale werking van de TF en rpL13 in zoogdiercellen (HEK293T). Daarna werd aan het medium rapaloog of ethanol (het solvent) toegevoegd. De lokalisatie van de TF veranderde niet onder invloed van ethanol, maar als rapaloog werd toegevoegd verschoof de TF van de nucleus naar het cytoplasma binnen 45 minuten. De reden hiervoor is dat er ternaire complex-formatie heeft plaats gevonden tussen het rapaloog en de twee domeinen, gefuseerd aan de TF en rpL13. Het ribosomale eiwit rpL13 werd ingebouwd in het ribosoom, dat getransporteerd wordt van de nucleoli naar het cytoplasma. De TF werd gevangen door het anker en met het ribosoom meegesleept naar het cytoplasma. Eenmaal in het cytoplasma kan het zijn normale functie, binden aan DNA, niet meer uitvoeren en is functioneel verwijderd.

Hoofdstuk 4 beschrijft het werk dat gedaan is om AA niet in relevante cellen te laten werken. Om dit te doen moesten de rapaloog-bindende domeinen (AA tags) in het genoom van ES cellen gefuseerd worden aan de specieke TFs, in dit geval ***Gata1***, dit wordt knock-in genoemd. Er waren al ES cellen die geen ***Ldb1*** expresseren, in die cellen kon ***Ldb1-AA*** transgeen tot expressie worden gebracht. In beide cellijnen werd ***rpL13-AA*** nog transgeen tot expressie gebracht. Activiteit van AA werd geconformeerd in ES cellen door rapaloog aan het medium toe te

voegen. Functionaliteit van de TFs werd geconformeerd door de ES cellen naar voorlopercellen van erytrocyten te differentiëren in kweek. Voor beide TFs geldt dat dit niet lukt als het eiwit niet functioneel is. Vervolgens werd AA getest in deze voorlopercellen door rapaloog aan het medium toe te voegen, maar het bleek dat dit niet werkt. AA in deze cellen werkt wel wanneer rapaloog aangeboden wordt aan de cellen door middel van liposomen. Deze cellen zijn nu gereed om AA op te testen en te vergelijken met traditionele manieren van het verwijderen van TFs (knock down of knock out).

Chapter 8

Curriculum vitae and publication list

Curriculum Vitae

Personal details

Name	Ruud Jacobus Johannes Jorna
Birth date	February 17th, 1983
Birth place	Made en Drimmelen, The Netherlands

Education

2001 – 2008	BSc and MSc Life Science & Techonology University of Leiden and Delft University of Technology, The Netherlands
1995 – 2001	VWO Dongemond College, Raamsdonksveer, The Netherlands

Research experience

2008 – 2013	PhD research Department of Cell Biology, Theme Biomedical Sciences, Erasmus Medical Centre, Rotterdam, The Netherlands ***Prof. dr. F.G. Grosveld (promoter) & Dr. E. Soler (copromoter)***
2008	Summer School international Genetically Engineered Machines (iGEM) Team: Delft University of Technology, Jamboree at Massachusetts Institute of Technology ***Dr. D.Bellomo, Dr. M.J.L. De Groot & Dr. E. Nikerel***
2007 – 2008	Internship to obtain MSc Department of Cell Biology, Erasmus Medical Centre, Rotterdam, The Netherlands ***Prof. Dr. E. Dzierzak & Dr. E. Haak***
2006	Internship to obtain BSc Leiden Amsterdam Centre of Drug Research, University of Leiden, Leiden, The Netherlands ***Prof. dr. B. Van de Water & Dr. S. Le Dévédec***

Publication List

Stadhouders, R., Jorna, R., Soler, E. and Grosveld, F. (in prep) **Functional analysis of distal gene regulatory elements: Myb -81kb enhancer knock-out and inactivation of Myb intronic transcriptional attenuator element.**

Jorna, R., Stadhouders, R., Bolkestein, M., Ten Hagen, T., Grosveld, F. and Soler, E. (in prep) **Fast knock-out of target nuclear factors in mammalian cells by Anchor-Away.**

Mylona, A., Andrieu-Soler, C., Thongjuea, S., Martella, A., Soler, E., Jorna, R., Hou, J., Kockx, C., van Ijcken, W., Lenhard, B., ***et al.*** (2013). **Genome-wide analysis shows that *Ldb1* controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis. Blood 121, 2902-2913.**

Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F., and Soler, E. (2012). **Transcription regulation by distal enhancers: who's in the loop? Transcription 3, 181-186.**

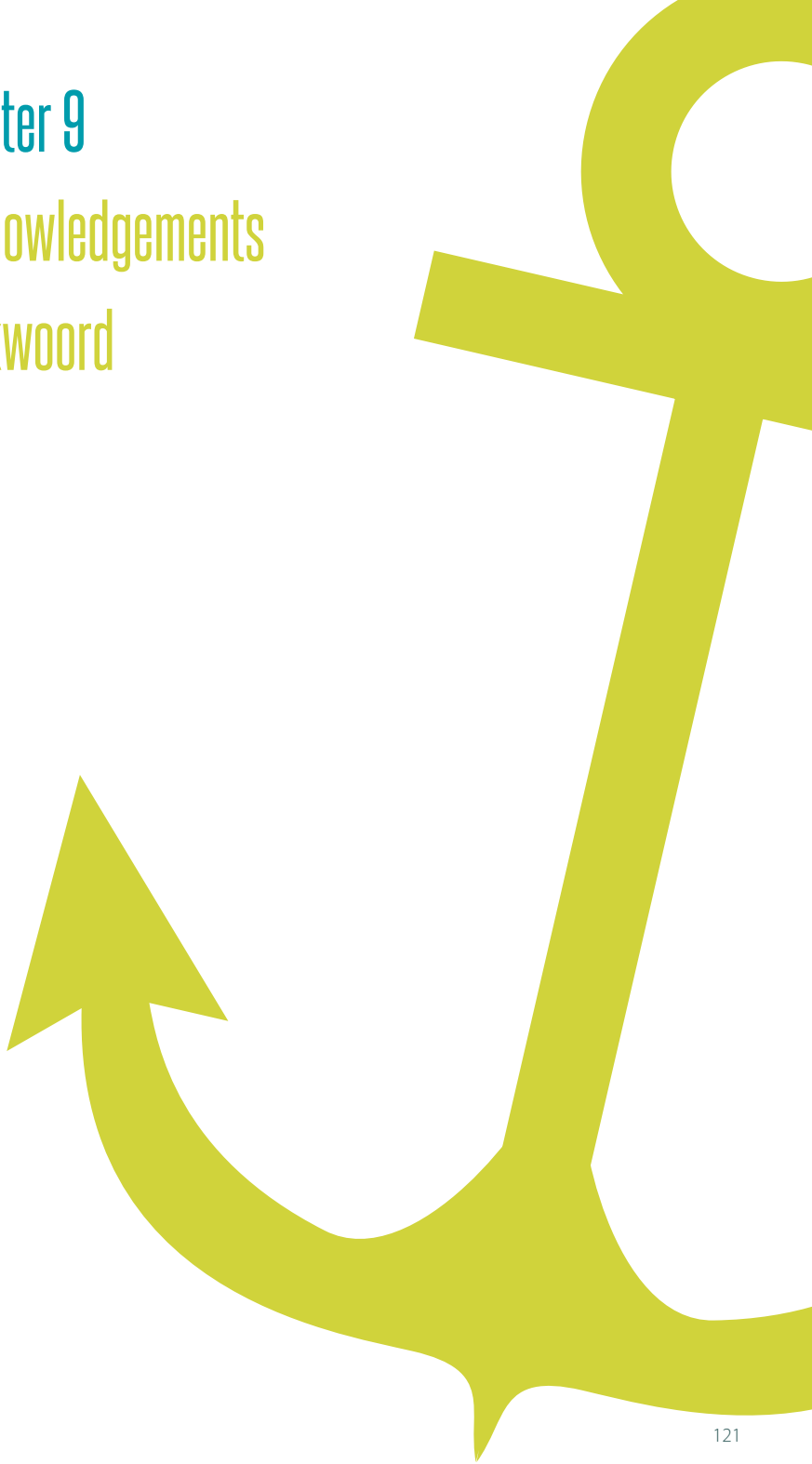
van Riel, B., Pakozdi, T., Brouwer, R., Monteiro, R., Tuladhar, K., Franke, V., Bryne, J.C., Jorna, R., Rijkers, E.J., van Ijcken, W., ***et al.*** (2012). **A novel complex,** RUNX1-**MYEF2, represses hematopoietic genes in erythroid cells. Mol Cell Biol 32, 3814-3822.**

Robin, C., Bollerot, K., Mendes, S., Haak, E., Crisan, M., Cerisoli, F., Lauw, I., Kaimakis, P., Jorna, R., Vermeulen, M., ***et al.*** (2009). **Human placenta is a potent hematopoietic niche containing hematopoietic stem and progenitor cells throughout development. Cell Stem Cell 5, 385-395.**

Chapter 9

Acknowledgements

Dankwoord



Acknowledgements

It is finished! After four-and-a-half years of work and some more months writing, interspersed the trivialities of life like getting a new job and taking care of our newborn baby, finally the work has been done. I am happy I was able to share this process with new and former colleagues, family and friends. Thanking everybody that has helped me during this time, some through suggestions and critical reading, others through helping me focus and yet others by helping me relax, would require me to write another 100 pages. That is to say: most of you will have to do with this general, but heartfelt thank you! I am excited to have taken a new path in my career and may be able to meet up with you in my new position at Luminex. But of course, there are some people whom I would like to mention personally for their big help in the realization of the work described in this thesis.

First of all my promoter, Frank, I would like to thank you for giving me a chance to work in your laboratory on one of your far-fetched projects. Your drive, genuine enthusiasm and curiosity about every result are reasons I could keep going on the project, also when results did not look particularly promising. Thank you as well for critically reading my thesis and helping to make it into the work it has become. I hope your renewed period at the bench will be as successful as your work from behind the desk has been!

Second, and not far behind, I would like to thank Eric. As my co-promoter and supervisor throughout the project you have been an invaluable source of information about literature and laboratory techniques. I would also like to give you a big 'thank you' for reading and correcting the texts in this thesis, even on holiday with the two new additions to your family. I hope Charlotte, you and the kids will be doing great in Paris, all the best in your future as a family and in your work! I sincerely hope and believe our paths will cross again!

Sjaak, Raymond and Jeroen, I am very happy you are willing to be a part of my small committee and comment on the pieces I have written. Your comments have improved this document considerably. Danny and Marieke, thank you for being on my committee as well. I feel honoured you make time to be a part of my thesis defense.

Huub and Ralph, thank you very much for being paranymph. Both of you have been a source of fun and scientific discussion during my time at the Erasmus MC. Huub, I am lucky to have a brother like you, and not only because you are the most experienced paranymph that is available at the Erasmus MC. Your contagious good mood and inventiveness make you one of the best people to spend a coffee break or lunch with. Ralph, good luck with the remainder of your studies and I wish you all the best in your scientific career. You are definitely one of the most talented young researchers I have come to know.

Next I would like to thank the people from the lab for their discussions (scientific and trivial), lessons and company. Rien and Rick, when it comes to high-throughput and experience with laboratory techniques you guys are the best! Petros, Anita, Xiao, Andrea and İrem: good luck with your own promotion, I hope I can be there when you defend your thesis. Guillaume: all the best with your post-doc, I hope you manage to get that great position in France. Alex: thanks for your help on the mouse work, I hope for you that PSV gets to be Dutch champions sooner rather than later!

I would also like to thank my parents for supporting me, financially and mentally, throughout my studies even though it must have been hard for you to really understand what I was doing. The days of studying are really over now, the time has come to stand on my own two legs. Knowing I can always call on you for advice and help is one of the greatest treasures I possess. Lastly I thank Hilde for being my dearest friend and for sharing the biggest wonder in my life yet. I do not know how I would have managed without you and I am looking forward to sharing our future as a family.

Dankwoord

Het is af! Na vier-en-een-half jaar werk en nog een paar maanden schrijven, onderbroken met de triviale dingen des levens zoals het beginnen aan een nieuwe baan en zorgen voor onze pasgeboren baby, zit dit werk er eindelijk op. Ik ben erg blij dat ik het proces van schrijven kon delen met zoveel nieuwe en oude collega's, familie en vrienden. In deze tijd hebben sommigen mij geholpen door teksten of resultaten te bekijken en te bekritiseren, anderen door mij te helpen focussen op wat belangrijk is en weer anderen hebben me geholpen ontspanning te vinden. Iedereen persoonlijk bedanken die mij geholpen heeft in de afgelopen jaren zou nog 100 pagina's tekst opleveren en daarom zullen de meesten het moeten doen met deze algemene, maar oprecht gemeende 'Dank je wel!' Ik ben blij dat ik een nieuwe weg heb kunnen inslaan en ben ervan overtuigd dat ik vele van jullie nog een keer tegen zal komen in mijn nieuwe baan bij Luminex. Toch zijn er enkele mensen die ik graag persoonlijk wil bedanken voor het grote aandeel dat ze hebben gehad in de realisatie van dit werk.

Als eerste wil ik graag mijn promotor bedanken. Frank, ik vind het geweldig dat je mij de kans hebt gegeven in je laboratorium te werken op een van je vergezochte projecten. Je gedrevenheid, oprechte enthousiasme en nieuwsgierigheid naar ieder resultaat zijn grote drijfveren om de moed niet op te geven, ook niet als de resultaten niet veelbelovend zijn. Ik wil je ook bedanken voor het lezen en verbeteren van mijn scriptie, zonder jou had het resultaat niet hetzelfde geweest. Ik hoop dat je werk aan wanneer je weer aan de labtafel staat net zo succesvol is als dat je werk van achter het bureau is geweest!

Als eervolle tweede wil ik graag Eric bedanken. Als mijn copromotor ben je een zeer waardevolle bron van informatie geweest van zowel literatuur als laboratorium technieken. Ook jou moet ik zeer dankbaar zijn voor het lezen en verbeteren van de teksten in deze scriptie, zelfs toen je met je familie (inclusief de twee nieuwe aanwinsten) op vakantie was. Ik hoop dat Charlotte, jij en de kinderen een fantastische toekomst in Parijs voor de boeg hebben, zowel persoonlijk als in het werk! Ik denk en hoop dat we elkaar in de toekomst nog wel zullen ontmoeten!

Sjaak, Raymond en Jeroen, ik ben erg blij dat jullie deel willen uitmaken van mijn kleine commissie en ook jullie bedankt voor jullie commentaar op mijn scriptie. De op- en aanmerkingen die ik van jullie ontvangen heb, hebben deze scriptie aanzienlijk verbeterd. Danny en Marieke, bedankt dat jullie zitting wilden nemen in mijn commissie. Ik voel me vereerd dat jullie tijd vrij willen maken om deel uit te maken van mijn verdediging.

Huub en Ralph, ik vind het geweldig dat jullie mijn paranimfen willen zijn. Jullie zijn beiden een grote bron van alledaagse en wetenschappelijke discussie geweest in mijn tijd bij het Erasmus MC. Huub, ik heb geluk dat ik iemand als jij als broer heb, en niet alleen omdat je waarschijnlijk de meest ervaren paranimf van het Erasmus MC bent. Je aanstekelijk goed humeur en inventiviteit zorgen dat je degene bent

waar ik het liefst een koffiepauze of lunch mee doorbreng. Ralph, veel geluk met de rest van je promotie traject en al het goeds toegewenst voor je wetenschappelijke carrière. Je bent een van de meest getalenteerde jonge wetenschappers die ik heb leren kennen.

Ik wil ook graag iedereen van het lab bedanken voor hun discussies (wetenschappelijk of triviaal), lessen en gezelschap. Rien en Rick, als het op high-throughput of ervaring met uiteenlopende technieken aankomt zijn jullie de beste! Petros, Anita, Xiao, Andrea en İrem: veel geluk met jullie eigen promotie, ik hoop dat ik aanwezig kan zijn wanneer jullie verdedigen. Guillaume: het beste voor met je postdoctorale onderzoek, ik hoop dat het je lukt om zo'n felbegeerde positie in Frankrijk te bemachtigen. Alex: bedankt voor je hulp met het muizenwerk, voor jou hoop ik dat PSV dit jaar (eindelijk weer eens) kampioen wordt!

Ik zou graag ook mijn ouders willen bedanken voor hun financiële en mentale ondersteuning tijdens mijn studies. Het moet soms lastig zijn geweest omdat het voor jullie niet altijd even begrijpelijk is waar ik precies mee bezig ben. Ik hoop dat dit tijdens mijn verdediging weer wat duidelijker is geworden. De dagen van studeren zijn nu echt voorbij, het wordt tijd dat ik (echt) op eigen benen ga staan! Ik weet dat ik altijd bij jullie kan aankloppen voor advies, dit is een van de grootste steunpilaren die ik heb. Als laatste wil ik Hilde bedanken: je bent mijn beste vriend en het grootste wonder in ons leven delen we. Ik weet niet hoe ik hier zonder jou doorheen gekomen was en ik kijk uit naar de toekomst met ons gezin.

Chapter 10

PhD portfolio

PhD Portfolio

ERASMUS UNIVERSITEIT ROTTERDAM

Name PhD student: Ruud Jorna
Erasmus MC Department: Cell Biology
Supervisor: Dr. E. Soler

PhD period: November 2008 – April 2013
Promotor: Prof. Dr. F.G. Grosveld

Activity	Year
General courses	
Safely working in the laboratory	2009
Molecular Cell Biology	2009
Analysis of microarray gene expression data	2009
Radiation course 5B	2010
Laboratory animal science, article 9	2010
Specific courses	
Genetic Engineering of Mammalian Stem Cells (oral presentation)	2011
Seminars and workshops	
Monday morning meeting presentations (10 oral presentations total)	2008-2013
PhD workshop MGC 2009, Brugge	2009
PhD workshop MGC 2010, Köln (oral presentation)	2010
PhD workshop MGC 2011, Maastricht (oral presentation)	2011
International conferences	
Second annual EUTRACC meeting, Prague	2009
Systems Biology of Stem cells, SyboSS meeting in Kirchberg (poster and oral presentation)	2013
Supervising master thesis	
İrem Baymaz – "A study on the interaction of the transcription factor ETO2 and the chromatin-modifier LSD1"	2010-2011
Extracurricular activities	
Organization of PhD workshop in Köln	2010
Participation in 'Wetenschapscafé'	2009-2012

Disturbance of Transcription Factor Dynamics in Mammalian Cells: Knock-In, Knock-Down, Knock-Out or Anchor-Away

The studies presented in this thesis were performed in the Department of Cell Biology of the Erasmus MC in Rotterdam, The Netherlands. The department is a member of the Medisch Genetisch Centrum Zuid-West Nederland (MGC).

ISBN: 978-94-6228-243-8