Chapter 1
Introduction

The human genome is regarded as the blueprint of life and the completion of its entire genomic sequence was a milestone in understanding the functions encoded in our genetic material. The genome contains of all of the coding and non-coding DNA sequences which control all of the functions within all cell types in our body. It is estimated that the human genome contains approximately 20,000 to 25,000 genes representing only 2% of genomic sequence [IHGSC (2004b) (Shabalina and Spiridonov, 2004)], while 98% of the genome is non-coding. The genes encode proteins controlling all of the various biological processes as well as ribosomal RNAs and proteins. The non-coding regions include maintenance elements, such as centromeres, telomeres and origins of replication which control DNA replication and repair, and elements such as promoters, enhancers/repressors, insulators, and regulatory RNAs (micro-RNAs) which regulate the spatial and temporal expression of coding genes.

Expression of eukaryotic genes is a tightly regulated process. It is crucial for genes to be expressed in the correct cell type to an appropriate level and at the correct time during cell differentiation and development in response to internal and external signals. Failure to regulate gene expression patterns can lead to serious consequences in genetic diseases. In the post-sequencing genomics era, with advances in both computational methods and genome-wide experimental approaches, it is important for us to study how different regulatory sequences and proteins interact to control gene expression, not only at a single gene locus, but globally across the genome within complex biological and transcriptional programmes. Understanding how gene expression is regulated is essential for us to fully delineate the function of our genome as well as to search for therapeutic remedies for genetic diseases.

1.1 Regulation of gene expression

Gene expression regulation can occur in different ways: during transcription, mRNA processing, and translation and at the level of protein stability. It is believed, however, that regulation occurs primarily at the transcriptional level. The transcriptional machinery of eukaryotes consists of two complimentary regulatory components: the *cis*-acting elements and the *trans*-acting elements.

The *cis*-acting elements are DNA sequences in the coding or non-coding regions of the genome. Epigenetic information can also be overlaid onto the *cis*-acting elements. This involves chromatin remodelling and modifications (histones or the DNA sequence itself) to create an accessible region

in the DNA for *trans*-factors to bind to initiate transcription. Conversely, some of these processes prevent *trans*-acting factors from binding to DNA by creating inaccessible chromatin environments.

The *trans*-acting elements are transcription factors or other DNA-binding proteins which recognise and bind to specific sequences in the *cis*-acting elements to initiate, enhance or suppress transcription. A transcription factor may regulate multiple genes or they may work in a combinatorial or complex manner to bind to the *cis*-regulatory elements at multiple transcription factor binding sites to generate a huge repertoire of unique and precise control patterns. It is estimated that the human genome encodes approximately 1800 transcription factors (Venter et al., 2001).

1.1.1 *Cis*-acting regulatory elements

Cis-regulatory DNA sequences include two distinct elements: promoters/proximal elements and the distal regulatory regions including enhancers, silencers or repressors, insulators and locus control regions (LCRs). These elements act in co-operation with one another to govern a co-ordinated expression pattern of a gene. They are summarised in Figure 1.1 and described in details below.

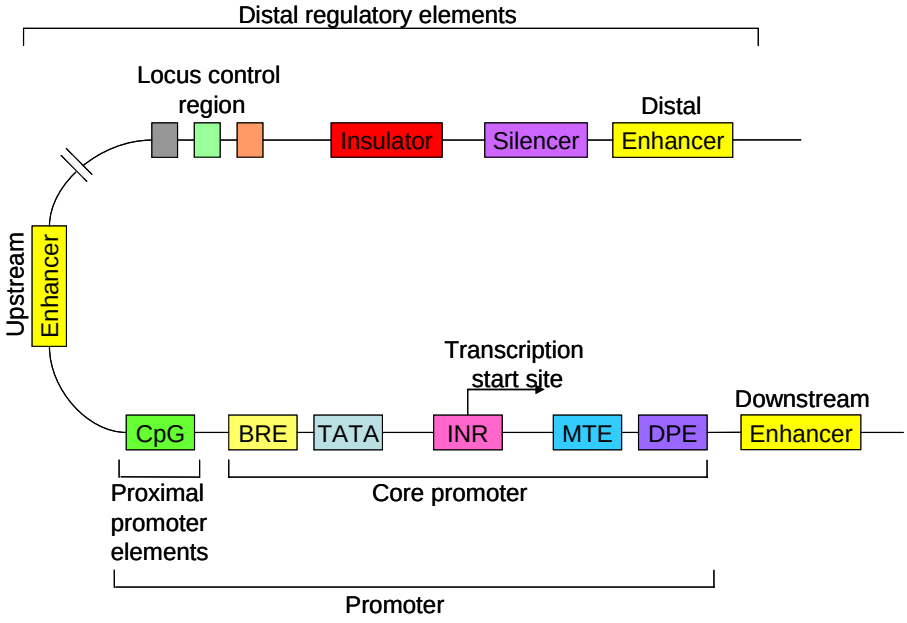

Figure 1.1. A schematic diagram of the types of *cis*-regulatory elements involved in regulation of gene expression. A typical promoter comprises a core promoter and proximal promoter elements such as CpG islands spanning about 1 kb around the transcription start site. The core promoter contains a TATA box (TATA), an initiator element (INR), a downstream promoter element (DPE), a motif ten element (MTE) and a TFIIB recognition element (BRE). Distal regulatory elements such as enhancers, silencers, locus control regions and insulators can be located upstream or downstream or even distant from the transcription start site. Various enhancers, silencers and locus control regions act together to activate or repress promoter activity while insulators prevent inappropriate regulation by regulatory signals from neighbouring genes.

1.1.1.1 Promoters

The RNA polymerase II (Pol II) promoter regions comprise the core promoter and the proximal promoter elements. Pol II promoters transcribe DNA to messenger RNA and small nuclear RNA (Section 1.1.2.1). The core promoter is located approximately 35 base pairs (bp) upstream or downstream of the transcription start site (TSS) and serves as the binding site of factors for assembly of the preinitiation complex (PIC). The core promoter contains a number of elements (Figure 1.2). The TATA box possesses the consensus sequence of TATAAAA located 25 to 30 bp upstream of the TSS. However, this consensus sequence may vary (Wong and Bateman, 1994; Zenzie-Gregory et al., 1993). Although the TATA box was believed to be a fundamental component of the core promoter, it was revealed that only 32% of the potential human core promoters contain the TATA box (Suzuki et al., 2001). The initiator element (INR) is located across the transcription start site (denoted as +1) from -3 to +5 having the consensus sequence of Py Py A(+1) N T/A Py Py. Downstream of the TSS, the downstream promoter element (DPE) functions in conjunction with the INR in TATA-less promoters and is located at +28 to +32 relative to the TSS and possesses the consensus sequence of A/G GA/T C/T G/A/C (Hahn, 2004; Smale and Kadonaga, 2003). Also located downstream of the TSS, the downstream core element (DCE) was first identified in the human β-globin promoter (Lee et al., 2005a). It is located at +10 to +45 relative to the TSS and acts distinct from the DPE. The motif ten element (MTE) is another newly defined element located at +18 to +27 relative to the TSS. It functions in a cooperative manner with the INR but independently from the TATA box and the DPE (Lim et al., 2004). All the core elements (TATA box, INR, DPE, DCE and MTE) initiate the recruitment of TFIID (Transcription factor IID) initiation complex to the promoter for transcription of gene to take place. Another core promoter element is the TFIIB recognition element (BRE) which is recognised by TFIIB instead of TFIID. It is located 3-6 bp upstream of the TATA box with the consensus sequence of G/C G/C G/A C G C C. BRE functions as a repressor of basal transcription whose repression is released upon the binding of activators. The existence of the core elements is not entirely universal (Gershenzon and Ioshikhes, 2005) and it is believed that other core elements may still remain to be discovered. Higher order structural properties of the DNA sequence are also involved in the recruitment of the PIC (Hahn, 2004).

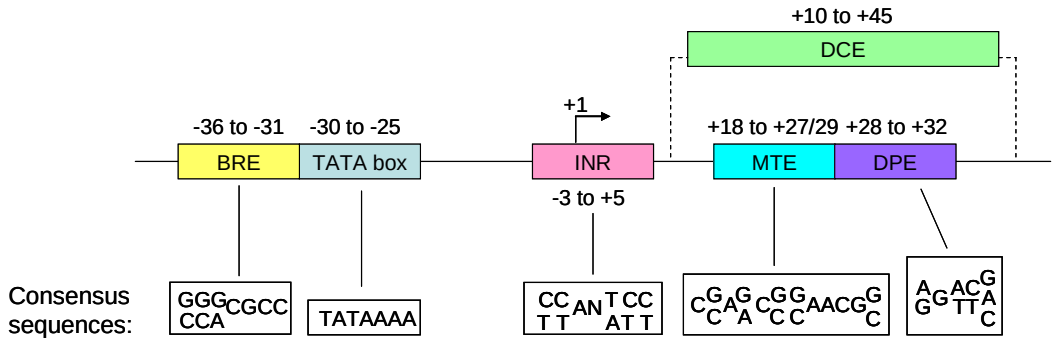

Figure 1.2. The RNA polymerase II core promoter. The locations of the core promoter elements TATA box, initiator element (INR), downstream promoter element (DPE), motif ten element (MTE), downstream core element (DCE) and TFIIB recognition element (BRE) corresponding to the transcription start site (+1) are shown. The consensus sequences of these elements are shown in the white boxes underneath each element. The joint arrow indicates the transcription start site (+1).

The proximal promoter elements refer to sequences upstream of the core promoter which can span up to a few hundred base pairs and can be involved in altering the rate of transcription. An example, of a proximal element is the CpG island which is 500 bp to 2 kilobase pairs (kb) in length and is highly GC rich (Smale and Kadonaga, 2003). They are associated with approximately 60% of human promoters. The core elements in the CpG islands have not been identified but CpG islands contain multiple binding sites for the transcription factor Sp1. CpG dinucleotides are the substrate of methylation by DNA methyltransferases and are normally underrepresented in the human genome as the methylated cytosine can undergo deamination to form thymine. However, CpG islands in the proximal promoters are not methylated in active genes. DNA methylation at CpG islands results in silencing of transcription and is implicated in epigenetic imprinting.

1.1.1.2 Enhancers

Enhancer elements increase the activities of promoters and thus facilitate the transcription of target genes in specific cell types during particular stages in development. Some promoters may be activated by a large repertoire of enhancers in different spatial and temporal environments or in response to different stimuli. An enhancer was first identified in the tumor virus SV40 and was found to increase transcriptional activities of heterologous genes in the host genome (Banerji et al., 1981). Soon after the discovery of the viral enhancer, the first endogenous enhancers in mouse and human were found to activate the immunoglobulin heavy chain gene in a tissue-specific manner (Banerji et al., 1983). A typical enhancer is approximately 50 bp to 1.5 kb in size and contains multiple transcription factor binding sites (TFBS) which are often conserved sequences with a certain degree of degeneracy which transcription factors recognise and bind. Different TFBS are arranged in a particular orientation to control the specificity of the enhancer. However, enhancer

elements *per se* are orientation and distance independent and can be located several kb upstream of the promoter, downstream of the promoter in intronic regions, or at/distal to the 3' end of the gene.

How does an enhancer mediate activation of its corresponding promoter? There currently are several models for its mode of action. Firstly, the proteins bound to enhancers and promoters may interact with each other by looping out the DNA sequence in between (Ptashne and Gann, 1997; Rippe et al., 1995; Vilar and Saiz, 2005). This results in the formation of a multi-protein complex for transcription to occur. Secondly, the enhancer and promoter may not come in contact with one another. Instead, the enhancer may direct the DNA element to localise into specific regions in the nucleus where high concentrations of transcription factors facilitate transcription (Lamond and Earnshaw, 1998). Alternatively, enhancers may act via supercoiling of DNA, nucleosome remodelling and altering chromatin structure to create an accessible structure for recruitment of regulatory proteins to initiate transcriptions (Freeman and Garrard, 1992). This will be discussed in more details in section 1.1.2.5. More recent studies have also demonstrated that RNA polymerase II (PolII) binds to distal enhancers and the PIC is assembled at the enhancer to promote formation of regulatory factor – promoter complexes for transcription (Louie et al., 2003; Spicuglia et al., 2002)

1.1.1.3 Silencers

In contrast to enhancers, silencers result in transcriptional repression rather than activation. Similar to enhancers, they are distance and orientation independent of gene structures. They can be located in the proximal promoter, as part of a distal enhancer, or occur independently in distal regions upstream or downstream of the gene they are regulating. Silencers are bound by repressor proteins to mediate repressions. These repressors may work independently, in cooperation with themselves (Harris et al., 2005) or other repressors (Sertil et al., 2003), or through the binding of a co-repressor (Chen and Evans, 1995).

There are two known mechanisms by which the association of repressors and silencers mediates transcriptional repression. The repressors may localise in the silencers preventing the access of an activator protein to their enhancers (Harris et al., 2005) or by preventing the binding of PolII or other basal transcription factors to the core promoter (Chen and Widom, 2005). Alternatively, the repressors may compete with activators for the same binding site to repress activation (Hoppe and Francone, 1998). Repressors may also recruit chromatin-remodelling enzymes or chromatin modifiers to create a chromatin structure which is unfavorable for the assembly of the transcriptional machinery (Heinzel et al., 1997).

1.1.1.4 Insulators

Insulators function in the genome to prevent genes from being incorrectly transcribed by the regulatory elements of the neighbouring genes. They are typically 500 bp to 3 kb in size. There are two main mechanisms for their function. Firstly, they may be present in the genome to block enhancer activity by inhibiting the interaction of promoters and enhancers (Zhao and Dean, 2004). Secondly, they may act by blocking the spread of repressive chromatin marks into regions containing transcriptionally active genes (West et al., 2002). Insulators are sometimes bound by trans-acting proteins to mediate their function. CTCF is one well-studied example which was found to bind to insulators at the β-globin locus (Bell et al., 1999) and to all known vertebrate insulators.

1.1.1.5 Locus control regions

The first locus control region (LCR) in mammals was discovered in the β-globin locus (Grosveld et al., 1987). LCRs are clusters of *cis*-regulatory elements such as enhancers, silencers and insulators where the collective action of these elements results in the overall control of gene expression. Similar to other *cis*-regulatory elements, LCRs can be located at upstream regions, downstream regions or within the introns of the gene they regulate. However, unlike normal enhancers or silencers, LCRs function in a copy number dependent manner and create an open chromatin structure for linked genes (Li et al., 2002).

1.1.1.6 Scaffold/Matrix attachment regions (S/MARs)

The eukaryotic genome is functionally compartmentalised into chromatin domains by attachment to nuclear matrixes or nuclear scaffolds which are protein-RNA structures within the nucleus. Such chromatin domains define gene transcriptional signatures and insulate the effects from adjacent genes. This is required for various biological functions to take place such as transcription and DNA replication. Scaffold/Matrix attachment regions (S/MARs) are DNA elements in the genome which mediate the attachment of chromatin loops to the nuclear matrix or nuclear scaffold. S/MARs are thus regarded as the borders of chromatin domains which range from 4 kb to 200 kb (Bode et al., 2003). S/MARs can function to insulate genes from any negative effects of the surroundings chromatin (Antes et al., 2001) or to increase transcription initiation rate even in the absence of an enhancer (Bode et al., 2000).

1.1.2 *Trans*-acting proteins involved in transcriptional regulation

In order for transcription to take place, various proteins and *trans*-acting elements are required for the assembly of the complete transcriptional machinery onto the various *cis*-acting elements. These proteins can be summarised as follows.

1.1.2.1 RNA polymerase

Transcription of genes from DNA to RNA is a three-step process involving initiation, elongation and termination. Initiation requires the association of RNA polymerase and general transcription factors to form a pre-initiation complex (PIC) at the promoter regions of genes. In eukaryotes, RNA polymerases are divided into three classes (RNA Pol I, II and III) according to the products they generate. RNA Pol I transcribes DNA to ribosomal RNAs (rRNAs) including the 28S, 18S and 6S subunits. RNA Pol II transcribes DNA to messenger RNA (mRNA) and small nuclear RNAs (snRNAs). RNA Pol III transcribes DNA to transfer RNA (tRNA) and 5S rRNA. The structure and transcriptional machinery of RNA Pol II is the most complicated among the three and the discussion below is focused on RNA Pol II.

The human RNA Pol II comprises 12 subunits, Rpb1 to Rpb12. Rpb1, 2, 3 and 11 have homologous counterparts in bacterial Pol whereas Rpb5, 6, 8, 10 and12 are common in all the three classes. Rpb4, 7 and 9 are unique components of RNA Pol II. Each of these subunits plays specific roles in transcription start site selection, alteration of elongation rate, interaction with activators and stability of RNA Pol (Lee and Young, 2000). Rpb1 contains a carboxyl-terminal repeat domain (CTD) which possesses repeats of the consensus sequence of Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The CTDs are phosphorylated during the switch from initiation to elongation and this phosphorylation is facilitated by protein kinases (Dahmus, 1995).

1.1.2.2 Basal/General transcription factors

The formation of a PIC at the core promoter is a stepwise process which requires assembly of general transcription factors (GTFs) and RNA Pol II. The core promoter is first bound sequentially by TFIID, TFIIA, TFIIB, RNA pol II and TFIIF to form the PIC, followed by binding of TFIIE and TFIIH. This complex then unwinds 12-15 bp of DNA at the transcription start site of the promoter to create an open structure for the formation of mRNAs. Different GTFs interact with the promoter at different regions and have various functions (Table 1.1).

GTF	Functions
TFIID	• Contains two subunits, TATA-binding protein (TBP) and TBP-associated factor (TAF). TBP binds to the TATA box of the core promoter in an orientation-independent manner while TAF binds to INR and DPE and is required for promoter selection and transcriptional activation. • Eukaryotes also encode a TBP-related factor (TRF) which recognises other DNA sequences in TATA-less promoters for the initiation of transcription.
TFIIA	• A heterodimer which interacts with TBP and stabilises the TBP-DNA interaction. • Promotes binding of TFIID to DNA. • Involved in the activation of transcription by binding to activators.
TFIIB	• Interacts with TFIID and RNA Pol II and is required for transcription start site selection. Binds to BRE and downstream sequences of the TATA box. • A direct interacting partner of activators which may promote the recruitment of TFIIB to the promoter.
TFIIF	• A heterodimer containing two subunits, TFIIFβ and TFIIFα. TFIIFβ binds to either upstream or downstream of TATA box while TFIIFα binds to regions downstream of the TATA box. • Binds RNA Pol II tightly and is involved in avoiding non-specific DNA binding and stabilisation of the PIC.
TFIIE	• Binds to DNA sequences directly upstream of the transcription start site after the formation of the PIC. • Promotes the recruitment of TFIIH and stimulates the CTD kinase and helicase activities of TFIIH.
TFIIH	• Contains two subunits, a core subunit with helicase activities (XPD and XPB) and a kinase subunit (Cdk7). The helicase subunit is required in the unwinding of DNA to create an open structure while the kinase subunit phosphorylates the CTD in RNA pol II during the transition to elongation stage.

Table 1.1. General transcription factors and their functions.

1.1.2.3 Sequence-specific transcription factors

RNA polymerase and general transcription factors account for the basal activity of the transcriptional machinery. In order to fully turn on or off the transcription of a gene, sequence-specific transcription factors are required. They bind to *cis*-regulatory regions such as promoters, enhancers and silencers to exert their activation or repression functions. These transcription factors recognise and bind to transcription factor binding sites (TFBS) which are often conserved sequences with a certain degree of degeneracy. Some of the properties of sequence transcription factors are outlined below.

- Modular nature

A sequence-specific transcription factor may be composed of several modules: the DNA-binding module, the dimerisation module, the activation or repression module and the regulatory module. This multi-module property was first observed in the yeast GAL4 transcription factor where the GAL4 protein binds to LexA binding sites through the fusion to the DNA-binding protein LexA (Brent and Ptashne, 1985). Many families of DNA-binding modules have been identified. These include the helix-loop-helix motif which was first discovered in prokaryotes, the homeodomain, zinc finger motif, leucine zipper motif etc (Pabo and Sauer, 1992). Different types of activation modules are present. The activation module can be an acidic or negatively charged alpha helix

(Hope et al., 1988), glutamine or proline-rich regions (Courey and Tjian, 1988; Mermod et al., 1989) or hydrophobic beta sheets (Leuther et al., 1993). In some transcription factors, a regulatory module is required for its activity or sub-cellular localisation. These modules may be present on the same polypeptide or they can be distinct subunits which are detachable and function in a *trans* manner (Baeuerle and Baltimore, 1988).

- Recruitment of coactivators or corepressors

Some sequence-specific transcription factors require the recruitment of coactivators or corepressors by protein-protein interaction to carry out their functions. Some co-factors, such as TAF, act as bridging molecule to bring the sequence-specific transcription factors and the general transcriptional machinery together. Other co-factors, such as chromatin-remodelling factors or histone-modifying enzymes, are recruited to alter chromatin structure, thereby initiating the activation or repression effect.

- Combinatorial effects

Activation or repression by sequence-specific transcription factors is tightly controlled and specific so that transcription of their target genes is regulated in a temporal and/or spatial manner. However, the binding of a single transcription factor may not be sufficient to exert tight regulatory control on the gene of interest. In many cases, clusters of various transcription factor binding sites are located in the *cis*-regulatory element to generate a unique motif for a combination of sequence-specific transcription factors to bind. Such a cluster of transcription factors often function synergistically where the combined activation is greater than with any one factor working alone.

- Posttranslational modification

The activity of sequence-specific transcription factors can also be controlled by post-translational modifications. One example is the phosphorylation of the cyclic AMP response element binding protein (CREB). When this protein is phosphorylated by protein kinase A upon cyclic AMP stimulation, it is activated and initiates the transcriptional activation at the target promoter (Gonzalez and Montminy, 1989). Other examples of post-translational modification include acetylation of p53 which increases its DNA binding affinity (Gu and Roeder, 1997) and ubiquitylation of LexA-VP16 (Salghetti et al., 2001).

- Multiprotein families

One additional property of sequence-specific transcription factors, which provides even more unique and complex regulatory patterns, is that many are members of multiprotein families. Examples of transcription factor family includes the Sp family, the AP-1 family and the GATA

family. Family members are closely related and share the same or very similar DNA binding motifs. In spite of this, they play different roles in transcriptional activation or repression and control the expression of their own set of target genes at certain stage of differentiation or development or in certain cell types. The GATA family of transcription factors is one classical example. The GATA family includes GATA 1→6. They are divided into two sub-families: the haematopoietic sub-family GATA1, 2 and 3 and the non-haematopoietic sub-family GATA4, 5 and 6. GATA1, 2 and 3 are expressed in various haematopoietic and neuronal cell lineage to control lineage commitment and specification whereas GATA4, 5 and 6 are expressed in the heart and digestive organs controlling cardiac-specific gene expression and epithelial cell differentiation in the gut (Ferreira et al., 2005; Molkentin, 2000).

1.1.2.4 Coactivators/ Corepressors

Coactivators and corepressors are important regulators of gene expression although they appear to have no DNA-binding properties. Instead, to exert their function, they interact with other general or sequence-specific transcription factors, and can modify histones/DNA or remodel chromatin.

As mentioned preciously, TBP-associated factors (TAFs) are part of the TFIID complex. Although some TAFs may bind to promoter DNA directly, others may bind to activators and general transcription factors transmitting information between the two. Examples are TAFII40 and TAFII60 which act as bridges between the p53 activator and the initiation complex (Thut et al., 1995).

Mediators, another class of coactivators, first identified in yeast, are multisubunit complexes which activate transcription stimulating the phosphorylation of CTD of RNA pol II. They also interact with activators and transmit positive or negative signals to the promoter (Myers and Kornberg, 2000). Seven mediator subunits have been discovered in human so far.

Certain coactivators or corepressors act as docking molecules on activators or repressors. Instead of having intrinsic enzymatic activities, they recruit other necessary factors for binding to the initiation complex or chromatin remodelling factors for transcriptional activation or repression. One example is OCA-B, a coactivator of the activator octamer binding protein (OCT). It recruits some TAFs for the activation of immunoglobin genes in B cells (Wolstein et al., 2000). Another example is the nuclear receptor corepressor (NcoR) which recruits histone deacetylases (HDACs) to produce an inactive chromatin structure to repress expression of nuclear receptors (Privalsky, 2004).

Some co-factors may possess both activating and repressive functions. Friend of GATA1 (FOG-1), a cofactor of GATA1, can promote or inhibit transcription by directly recruiting histone acetyl transferases (HATs) or histone deacetylases (HDACs) at specific sites (Letting et al., 2004). FOG-1

also functions as a chromatin occupancy facilitator, a possibly new class of cofactor, where it facilitates the binding of GATA1 to sites originally bound by GATA2 (Pal et al., 2004).

1.1.2.5 Chromatin modifying factors

Epigenetic regulation by modification of chromatin plays a crucial role in regulating gene expression. Nucleosomes are the basic subunits of chromatin where DNA is packaged with histone proteins. The core histone proteins H2A, H2B, H3 and H4 bind to one another to form a protein octamer wrapping the DNA whereas the linker histone H1 binds to the outside of the nucleosome which stablisises the folding of the nucleosome. Nucleosomes have dynamic properties which are governed by a specific class of co-factors - chromatin modifying factors - which include chromatin remodelling complexes and histone modifying enzymes. These co-factors modify the structure of chromatin to facilitate or interfere with the recruitment of PICs and transcription factors to promoter regions or other regulatory elements.

A. Chromatin-remodelling complexes

There are at least five families of chromatin-remodelling complexes in eukaryotes: SWI/SNF, ISWI, NURD/Mi2, INO80 and SWRI families (Saha et al., 2006). All families contain an ATPase subunit where they use ATP-hydrolysis to modify chromatin structure and remodel nucleosomes. Other subunits in the complex may be involved in the modulation of ATPase activity and the targeting to specific regions of chromatin.

Two mechanisms by which chromatin-remodelling complexes function to modify chromatin structure to increase accessibilities of nucleosomal DNA have been described. The sliding of DNA with respect to the histone proteins is the most widely studied mechanism (Meersseman et al., 1992). The result is that the histone octamer is re-positioned to interact with different DNA elements instead of the original DNA elements. Another possible mechanism of nuclesomal re-positioning involves conformational changes (Lorch et al., 1999; Studitsky et al., 1994). Such conformation changes may result in the collapse of the altered nucleosome to a canonical nucleosome in contact with a different DNA segment. It is also possible that the histone proteins are released and interact with a new segment of DNA following conformation change.

B. Histone-modifying enzymes

Histone modifying enzymes promote the covalent modifications of the histone proteins. These covalent modifications include acetylation, methylation, phosphorylation, ubiquitylation, ADP ribosylation, sumoylation and isomerisation. Such modifications either affect the higher-order chromatin structure by disrupting histone-DNA interactions or recruit chromatin remodelling complexes and other proteins.

Histone acetylation was the first post-translational modification identified on histone proteins (Allfrey et al., 1964). This modification is characterised by the addition of an acetyl moiety to the ε-amino group of the lysine residue and is associated with transcriptional activation. Acetylation is a dynamic and reversible process controlled by histone acetyltransferases (HATs) and histone deacetylases (HDACs).

HATs are categorised into three families: the Gcn5-related N-acetyltransferase (GNAT) superfamily, the MYST family and the p300/CBP family. HATs do not bind DNA directly and they are usually components of multiprotein complexes which are recruited to promoter regions or other regulatory regions by interaction with DNA-bound activators (Utley et al., 1998). These complexes contain several subunits which carry out distinct functions including interaction with different types of transcription factors. In addition to the modular nature of HAT complexes, the combination of various components in the complexes also dictates the recognition site specificity (Grant et al., 1999).

While HATs confer transcriptional activation, HDACs correlate with repression. There are three classes of HDACs: class I, II and III. Class I includes HDAC 1, 2, 3 and 8 while class II includes HDAC 4, 5, 6, 7, 9 and 10. Members of these two classes share sequence similarities and require Zn^{2+} for their function. Class III HDACs are also called the Sir2 family and includes SIRT1-7. They have low amino acid sequence homology to class I and II and they are nicotinamide adenine dinucleotide (NAD^{+})-dependent. Similar to HATs, HDACs are found in protein complexes. For example, both HDAC 1 and 2 are found in the Sin3, NuRD and CoREST complexes which contain other subunits required for protein-protein interaction and chromatin remodelling.

Histone methylation, including the addition of methyl group in a mono-, di- or tri- manner, has been shown to occur at both lysine and arginine residues. Methylation of lysine on histone subunits H3 and H4 is catalysed by histone methyltransferases (HMTs) which share the common catalytic 130-amino-acid SET domain, except in the modifier Dot1 (Rea et al., 2000). Methylation of arginines is catalysed by the protein arginine methyltransferases (PRMT) family. Members of this family share a highly conserved core AdoMet binding region which forms the protein substrate binding cleft and has methyltransferase activity (Lee et al., 2005b).

The lysine demethylase (LSD1) demethylates H3K4 mono- or di-methylation by means of an amine oxidase reaction and mediates transcriptional repression (Shi et al., 2004). LSD1 associates with Co-REST, a transcriptional co-repressor, to demethylate nucleosomal substrates (Shi et al., 2005). LSD1, when present in an androgen receptor complex, also demethylates H3K9 methylation and activates transcription (Metzger et al., 2005). A distinct class of lysine demethylases is the Jumonji C (JmjC)-domain-containing family where the JmjC domain is the core catalytic domain. A number

of members have been identified for this class and they target different lysine residues. Reversal of arginine methylation involves deimination which is the process of converting a methyl-arginine to citrulline by the enzyme peptidylarginine deiminase 4 (PADI4) (Cuthbert et al., 2004; Wang et al., 2004b). This antagonises the effect of arginine methylation. However, only mono-methylated arginine residues have been demonstrated to undergo deimination.

Phosphorylation of histone H3 subunit at serine 10 has been shown to be associated with transcriptional activation of the immediate early genes in human such as *c-jun* (Mahadevan et al., 1991). MSK1/2 and RSK2 kinases mediate this phosphorylation function (Sassone-Corsi et al., 1999; Thomson et al., 1999).

Ubiquitylation of histone subunits involve the addition of a 76-amino-acid ubiquitin protein. H2A is ubiquitylated at a lysine residue by the Bmi/Ring1A-containing human Polycomb repressive complex 1-like complex (hPRC1L) which mediates transcriptional repression (Wang et al., 2004a). H2B lysine ubiquitylation is catalysed by RNF20/RNF40 and UbcH6 in human (Zhu et al., 2005) and by Rad6 and Bre-1 in yeast (Robzyk et al., 2000; Wood et al., 2003). De-ubiquitylation of H2B is carried out by Ubp8 in the SAGA or SILK complexes, while de-ubiquitylation of H2A requires 2A-DUB (Daniel et al., 2004; Zhu et al., 2007).

Other modifying enzymes are less well characterised. Mono-ADP-ribosyltransferases (MARTs) and poly-ADP-ribose polymerases (PARPs) mediate ADP ribosylation of histones (Hassa et al., 2006). FPR4 isomerises a proline residue on the H3 subunit (Nelson et al., 2006) which in turn regulates methylation of the proline residue. Sumoylation is characterised by the conjugation of the SUMO protein to its histone substrate substrate by Ubc9 and is linked to transcriptional repression by antagonising acetylation and ubiquitylation (Johnson, 2004).

1.1.3 The transcription factory

Co-ordinated gene expression patterns require a combined effort of various transcription factors and chromatin modifiers to direct gene expression at various loci at certain developmental time point or in response to external stimuli. The previous sections described the transcriptional machinery at the molecular level but indeed such co-ordinated gene expression regulation also requires chromosomal organisation in a three-dimensional space of the nucleus. It has been suggested that active genes are repartitioned into nuclear territories for transcription to take place (Chambeyron and Bickmore, 2004; Williams et al., 2006). Looping and intra- or inter-chromosomal interactions between regulatory loci or active genes have also been documented to provide integrated expression (Spilianakis et al., 2005; Zhao et al., 2006). A more recent study illustrated that these kinds of

chromosomal re-organisation require induction by ligands and is facilitated by co-activators and components of the chromatin-remodelling complexes (Nunez et al., 2008).

1.2 Transcriptional regulatory networks

Studying gene regulation at a single gene locus fails to give a full picture of global regulatory patterns - genes across the genome interact with proteins, through time and space, within the cell to control their expression. Transcriptional regulatory networks are the programmes of multiple interactions within cells including transcription factor-DNA interaction and other factors that modulate these interactions biochemically to control the expression of genes. Such networks are crucial in dictating cellular behaviours in response to specific signals or at different stages of development.

1.2.1 Properties of a transcription network

Transcriptional regulatory networks, like other biological networks, consist of nodes which are connected by edges (Figure 1.3). Nodes include various transcription regulator proteins such as transcription factors (TF), co-factors and chromatin regulators and various DNA elements such as promoters and enhancers. Edges are the physical interactions between regulator proteins (protein-protein interactions) and between regulator proteins and DNA elements (protein-DNA interaction).

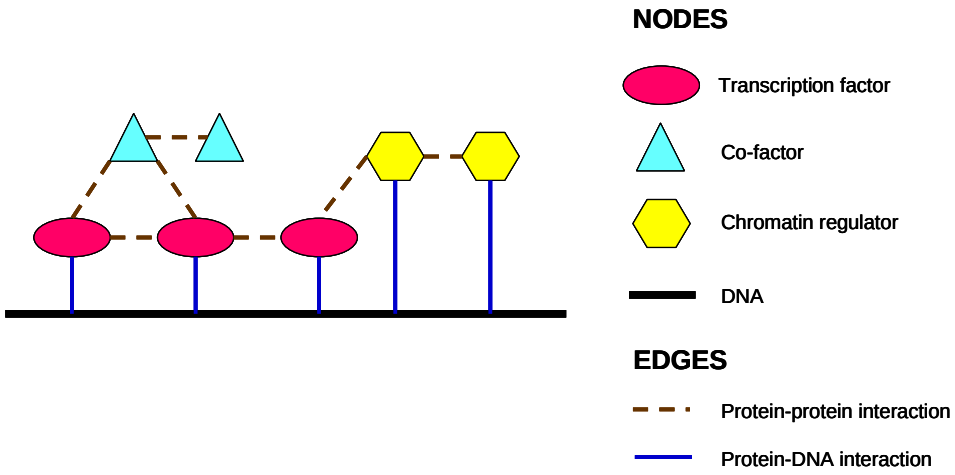

Figure 1.3. Nodes and edges of a transcription network. Schematic diagram shows the nodes and edges of a transcription network. Nodes include transcription factors (pink oval), co-factors (aqua triangle), chromatin regulator (yellow hexagon) and DNA (black line). Edges describe the relationships among the nodes which include protein-protein interactions (brown dotted line) and protein-DNA interaction (blue solid line).

How nodes and edges are related is key to understanding a transcriptional network. Network motifs are small networks with distinct properties, which in combination define the genetic control of the transcription programmes (Lee et al., 2002; Milo et al., 2002; Shen-Orr et al., 2002) (Figure 1.4). An autoregulation motif consists of a transcription factor binding to its own promoter to stimulate

expression. This ensures the stability of its own expression. In contrast, a multi-component loop motif involves more than one transcription factor binding to the promoters of one another to regulate expression. In a feed-forward loop motif, a transcription factor regulates the expression of another transcription factor while both of these transcription factors regulate the expression of a common target gene. A slight modification in the level of a master regulator can result in a significant increase or decrease in the target expression due to the presence of the second regulator which is under the control of the master regulator. A single input motif contains one transcription factor which co-regulates a number of target genes and this often ensures a co-ordinated expression pattern for a certain subset of genes. A multiple input motif involves a set of transcription factors binding to the promoters of the same set of target genes. This allows the expression of the targets to be co-ordinated in response to different signals which stimulates or inhibits expression of the regulator transcription factors. A dense overlapping region utilises a set of transcription factors that overlap to regulate a set of targets where each of these targets is regulated by a different transcription factor combination. A regulator chain motif consists of 3 or more regulator transcription factors in a series where the first transcription factor regulates the second transcription factor, which in turn regulates the third.

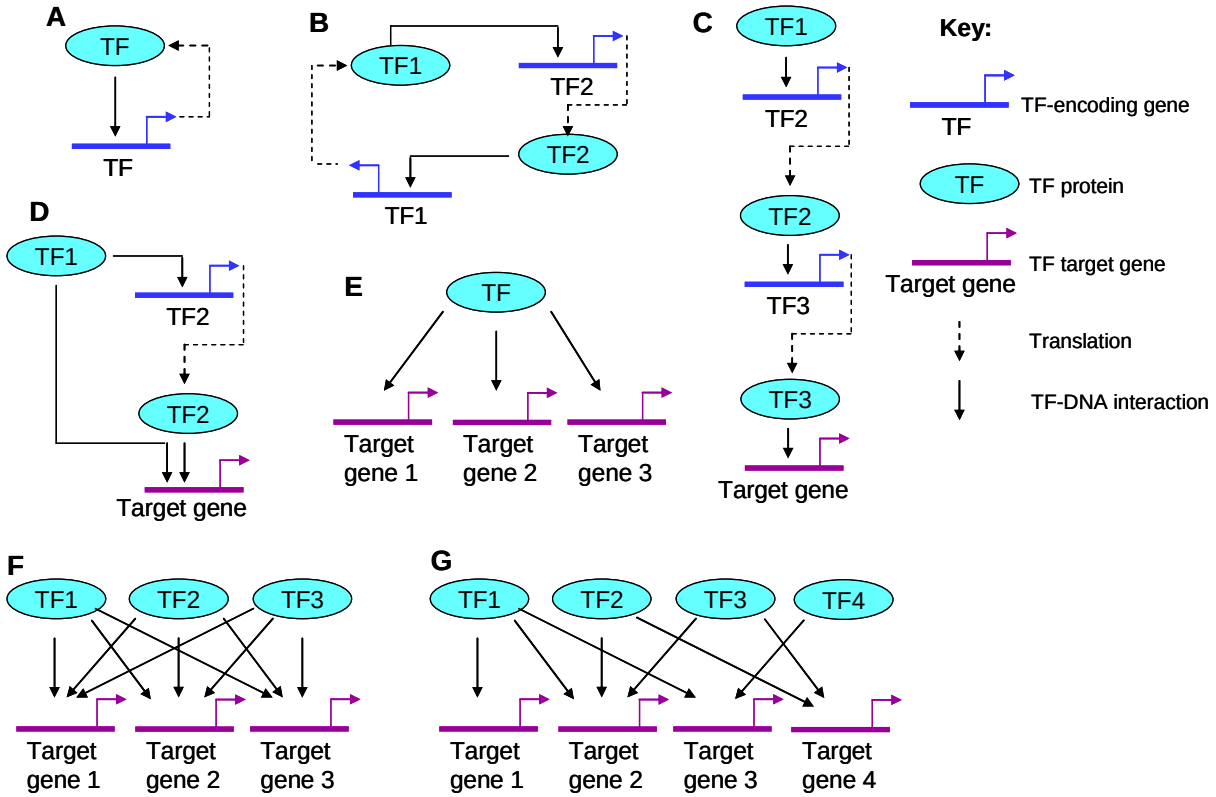

Figure 1.4. Network motifs of transcriptional regulatory networks. Combinations of these motifs regulate expression patterns in a transcriptional network. Black solid arrows indicate TF-DNA interactions and transcription factor regulation while black dotted arrows indicate translation of proteins. A: autoregulation motif. B: multi-component loop. C: regulator chain. D: feed forward loop. E: single input motif. F: multiple input motif. G: dense overlapping region. These motifs are explained in detail in the text.

Many studies have been performed to delineate transcriptional regulatory networks in various biological systems in eukaryotes. For example, in *Saccharomyces cerevisiae*, a comprehensive network of regulator-gene interactions of all known regulators has been determined (Lee et al., 2002). More specifically, yeast has been widely used in the study of transcriptional networks controlling the cell cycle (Lee et al., 2002; Oliva et al., 2005). In mouse, the Sonic hedgehog (Shh)-driven Gli-mediated transcriptional network defining neuronal development and specification has been characterised (Vokes et al., 2007).

1.2.2 Experimental and computational approaches for deciphering regulatory networks

To delineate transcriptional regulatory networks, or aspects thereof, a combination of experimental and computational approaches is required (Figure 1.5). The feasibility of such approaches have been demonstrated most effectively in model organisms such as *Saccharomyces cerevisiae* (Tavazoie et al., 1999) and *Caenorhabditis elegans* (Horner et al., 1998). Methods, such as expression profiling with microarrays (see section 1.3.2), which allow us to visualise the effects of perturbing a particular transcription factor in a biological system, facilitate the identification of direct and secondary target genes co-regulated by a transcription factor, and at the same time provide information about the mode of regulation (i.e. activation or repression). These data enable us to search for the common regulatory elements (e.g. transcription factor binding sites) in the co-regulated genes by computational methods. Further confirmation of the transcription factor binding sites at direct target genes can be achieved by chromatin immunoprepitation (ChIP), while ChIP combined with microarray (ChIP-chip) can be used to map the transcription factor binding sites on a genome-wide scale (Iyer et al., 2001; Ren et al., 2000). Direct and secondary targets of transcription factors can also be distinguished in this way. Chromatin structure affects the binding of transcription factors to regulatory sequences (Hassan et al., 2001; Jenuwein and Allis, 2001). Therefore, the study of chromatin structure and its biochemical modifications is crucial to understanding the complete picture of gene regulation (Lieb et al., 2001). Furthermore, chromosome correlation maps for the chromosomal locations of co-regulated genes can be generated which often show that genes co-localised to specific genomic regions are in open chromatin structures enabling active transcription of the region (Cohen et al., 2000). The combination of all or a subset of these approaches provides insights into the regulatory networks in biological systems (Shannon and Rao, 2002).

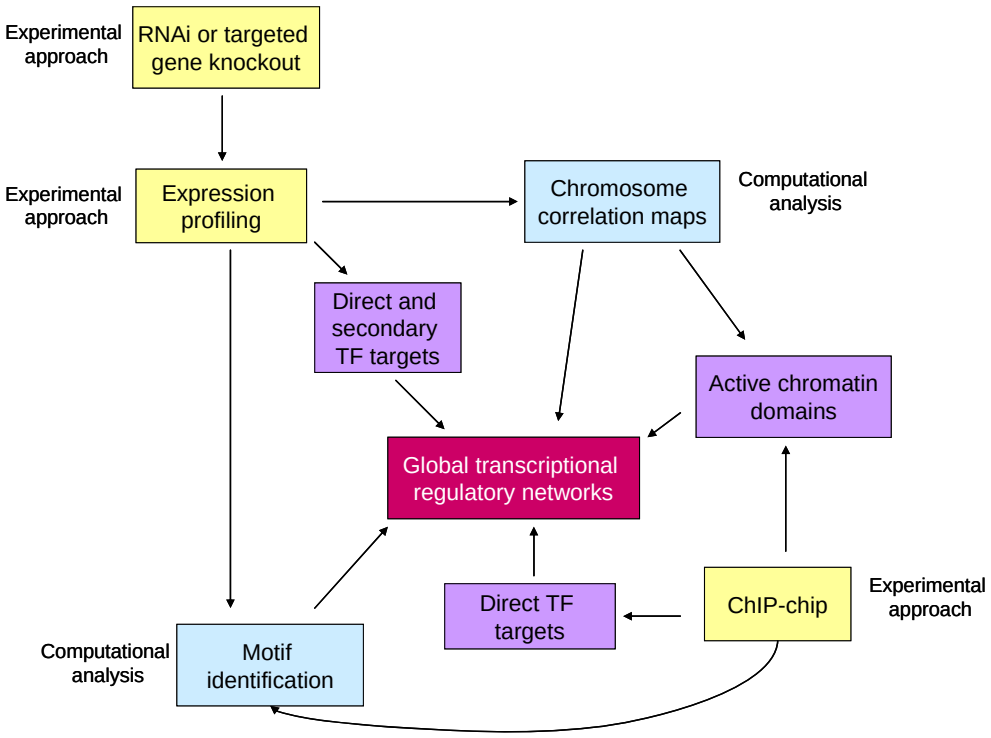

Figure 1.5. Experimental and computational approaches to delineate transcriptional regulatory networks. A combination of methods can be used to identify novel relationships between transcription factors and co-factors with their target genes. Yellow box: experimental approach; blue box: bioinformatics approach; purple box: experimental or informatic outcome. Expression profiling data obtained from gene knockdowns or knockouts enable us to identify direct and indirect targets, generate chromosome correlation maps and search for regulatory motifs. ChIP-chip generates genome-wide data on global TF-DNA binding and histone modifications. These data combined together provide insights into global transcriptional networks.

1.3 Experimental and computational approaches to understand transcriptional regulation

1.3.1 Gene perturbation by RNA interference

Perturbation of the activity of transcription factors has been widely used to study their function and to identify downstream target genes involved in transcriptional programmes. Traditionally, complete knockouts of the gene of interest provide the cleanest experimental paradigm to study. However, generating knockouts are time-consuming and can often result in lethality. With the discovery and advances in RNA interference, transient or stable gene knockdowns can be induced in the cell type of interest and are relatively rapid and inexpensive approaches for the delineation of downstream target genes of transcription factors.

1.3.1.1 Discovery and mechanism of RNA interference

Introduction of double-stranded RNA (dsRNA) was first found to silence genes with complementary sequences in *Caenorhabditis elegans* and has been termed RNA interference or

RNAi (Fire et al., 1991; Fire et al., 1998). Such silencing machinery by dsRNAs was first described as an anti-viral response to protect the organism from RNA viruses and the random integration of transposable elements (Waterhouse et al., 2001). The underlying molecular mechanism of RNAi involves two main steps. dsRNAs are processed into short interfering RNAs (siRNAs), which are about 22 nucleotides in length, by the RNase III enzyme Dicer. These mature siRNAs then associate with various proteins including the Argonaute protein family to form the RNA-induced silencing complex (RISC), where the siRNAs unwind. RISC then uses the unwound strand as the guide which identifies the substrates (Figure 1.6). Subsequent gene silencing occurs at various levels. At the post-transcriptional level, the identification of the target mRNA by the guide siRNA may trigger mRNA degradation by first cleaving the target mRNA. The mRNA cleavage requires siRNA and mRNA base-pairing together with the Argonaute protein which contains an RNaseH-like domain and all the critical active residues for endonucleolytic cleavage (Meister et al., 2004). The resultant cleaved mRNA fragments are directed to the general cellular mRNA degradation pathway which deadenylates the mRNA followed by 3' to 5' or 5' to 3' degradation. Post-transcriptional repression by RNAi can also be achieved by the inhibition of protein translation. In addition to post-transcriptional suppression of gene expression, RNAi is also implicated in silencing at the transcriptional level. siRNAs targeted to the promoter regions of genes can induce transcriptional silencing by DNA methylation in human cells (Kawasaki and Taira, 2004; Morris et al., 2004). However, the mechanism by which siRNAs enter the nucleus for DNA methylation remains unknown. siRNAs can also methylate histone H3 lysine 9 and 27 recruiting chromatin-remodelling complexes such as Mi2/NuRD and Sin3/HDAC resulting in the condensation of chromatin and transcriptional repression (Kawasaki and Taira, 2004; Weinberg et al., 2006).

In plants and *Caenorhabditis elegans*, the RNAi effect can be amplified through the mechanism of transitive RNAi (Sijen et al., 2001). siRNAs targeting the 3' end of a transcript results in the suppression of the mRNA and further production of siRNAs against the same region. siRNAs against sequences upstream of the original targeted region are also generated. Therefore, the RNAi effect is significantly enhanced even with the introduction of minute amount of exogenous dsRNA. Such amplification requires the plant RNA-directed RNA polymerase (RdRP) or the *C. elegans* homologue EGO1 which employ the target mRNA as a template (Schiebel et al., 1998; Smardon et al., 2000). However, this amplification system has not been demonstrated in mammalian systems, where no RdRP homologue has yet been identified.

RNAi also takes place in the endogenous gene silencing machinery using microRNAs (miRNAs). miRNAs are 21 to 23-nucleotide RNA duplexes which are transcribed by miRNA genes and have less than complete complementarities to their targets. Primary microRNAs (pri-miRNAs) are first

processed to form pre-miRNAs by the enzyme Drosha and pre-miRNAs then enter the RNAi pathway (Figure 1.6). The first miRNA, *lin-4*, was discovered in *C. elegans* and was found to control the timing of various stages of larval development by blocking translation of the protein LIN-14 (Lee et al., 1993; Olsen and Ambros, 1999; Wightman et al., 1993). Since then, many more miRNAs have been discovered in invertebrates and mammals and these have been shown to be critical during developmental timing, cell proliferation, differentiation and apoptosis, and signalling pathways. Genomic rearrangements resulting in altered expression of miRNA genes and/or changes in miRNA target sites have also implicated in cancer and other diseases (Kloosterman and Plasterk, 2006).

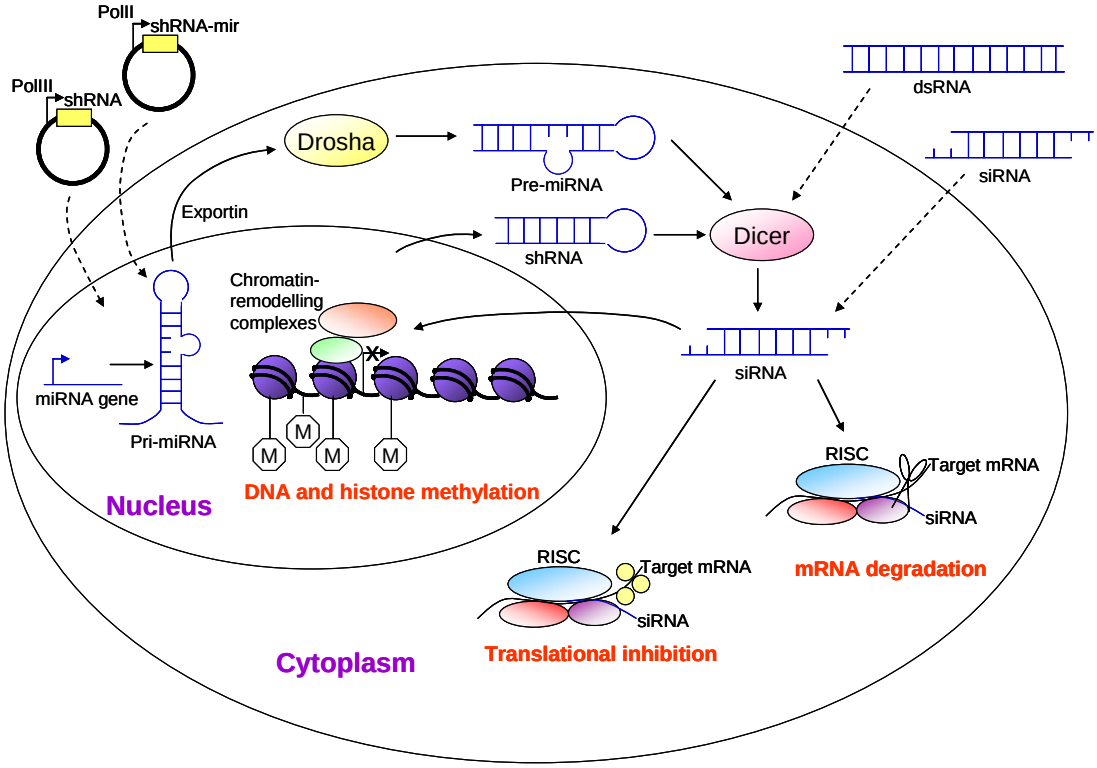

Figure 1.6. The mechanism of RNAi and various ways of triggering RNAi by exogenous sources. In addition to endogenous micro-RNA (miRNA), introduction of exogenous double-stranded RNA (dsRNA), short interfering RNA (siRNA) or plasmids containing short hairpin RNA (shRNA) or short hairpin RNA mir (shRNA-mir) genes all trigger RNA interference in the cell. Transcription of endogenous miRNA genes and exogenous shRNA-mirs inside the nucleus generates pri-miRNAs which are exported to the cytoplasm by exportin and further processed into pre-miRNA by Drosha. Pre-miRNAs, exogenous dsDNAs and shRNAs generated by the exogenous shRNA gene are then processed into siRNA by Dicer. Endogenously-produced siRNAs or exogenous siRNAs incorporate into the RNA-induced silencing complex (RISC) and mediate gene silencing by mRNA degradation, translational inhibition or DNA and histone methylation. Key: dotted arrows: introduction of RNAi triggers outside the cell; ovals: various proteins; dark blue circles inside nucleus: histones; M: methylation.

1.3.1.2 Inducing RNAi in various organisms

A. Invertebrate systems

Inducing RNAi in invertebrate systems such as *C. elegans* and *Drosophila* is relatively straightforward experimentally. Injecting small amounts of long dsRNAs of over 500 bp into the tail of *C. elegans* was first described to induce gene-specific silencing throughout the entire organism and the knockdown was persistent in the progeny (Fire et al., 1998). Similar methods of injection was also described in *Drosophila* (Kennerdell and Carthew, 1998). The spreading effect of RNAi across cell boundaries and inheritance through progeny is mainly due to the ability of *C. elegans* to amplify siRNAs. However, microinjections require expensive equipment and expertise. Simply feeding the worms with *Escherichia coli* expressing the dsRNAs was also found to induce repression in gene expression (Timmons and Fire, 1998). Improvements in the feeding method using a strain of *E. coli* deficient for RNaseIII and engineered to produce high quantities of specific dsRNAs when fed to *C. elegans* resulted in knockdown phenotypes comparable to complete gene knockouts (Timmons et al., 2001). Soaking the worms in solutions containing dsRNA was also demonstrated to be an alternative method of delivery (Tabara et al., 1998). Stable integration of inducible dsRNA-expressing constructs have been developed in *C. elegans* and *Drosophila* embryos and cultured cells where mutant lines can be maintained through multiple generations (Clemens et al., 2000; Kennerdell and Carthew, 2000; Tavernarakis et al., 2000).

B. Mammalian systems

The use of long dsRNA of over 30 bp in mammalian cells to silence genes was found to trigger the innate immune response (Williams, 1997). The enzyme dsRNA-dependent protein kinase (PKR) is activated on binding to long dsRNA, which results in the sequence-independent destruction of all RNAs and generalised repression of protein synthesis. This results in non-specific repression of gene expression within the cell. Therefore, other approaches have been used to induce gene-specific RNAi responses in mammalian cells (Figure 1.6). Each of these is described below and the advantages and disadvantages are summarised in Table 1.2.

(i) siRNAs for RNAi

Introduction of short interfering RNA (siRNA) (shorter than 30 base pairs) into mammalian cells, was found to induce the sequence-specific RNAi pathway (Caplen et al., 2001; Elbashir et al., 2001). These siRNAs are short duplexes of approximately 19 nucleotides in length with 2-nucleotide 3' overhangs on each strand. They bypass the Dicer processing step and enter the RNAi pathway by directly incorporating into the RISC complexes. siRNAs can be synthesised with or without chemical modification to increase their stability and specificity. Alternatively, they can be

generated by *in vitro* transcription of the target cDNA followed by cleavage by recombinant Dicer or bacterial RNase III (Myers et al., 2003; Yang et al., 2002). Dicer cleavage produces siRNA in its natural form for entry into RISC complex. It has been suggested that enzymatically prepared siRNAs can dramatically reduce the off-target effects (Kittler et al., 2007). The silencing effect of siRNA is dependent upon the transfection efficiency and the amount of siRNA used.

(iii) shRNAs for RNAi

Expressing short hairpin RNA (shRNA) from RNA polymerase III promoters in plasmid or viral-based vectors is an efficient way of silencing target genes. shRNAs are produced as single-stranded 50-70 nucleotides molecules which form stem-loop structures. The shRNA mimics the endogenous microRNA (miRNA) pathway to trigger the cleavage of shRNAs generating siRNA for the silencing of specific genes (Brummelkamp et al., 2002; Paddison et al., 2002). The shRNA-encoding DNA fragments can be made by chemically synthesising 50 to 70-nucleotide long oligonucleotides which can be annealed and cloned into a vector. Alternatively, they can be generated by PCR-based methods along with restriction enzyme digestion, which greatly enhance the efficiency of construct generation for RNAi screen (Gou et al., 2003; Sen et al., 2004).

(iii) shRNA-mirs for RNAi

shRNA-mirs are optimised sequences having miRNA-like properties. They are generated by miRNA precursors as the backbone for delivery of hairpin loops flanked by stem sequences found in miRNAs. shRNA-mirs have been demonstrated to successfully induce transient (Zeng et al., 2002) and stable gene knockdowns (Boden et al., 2004b; Dickins et al., 2005). The silencing effects produced by shRNA-mirs are significantly higher than for conventional shRNAs (Boden et al., 2004b; Silva et al., 2005). In some cases, even a single-copy integration can generate potent and stable knockdown (Dickins et al., 2005). This is particularly important for RNAi-based analyses as a reduction in concentration of siRNAs generated *in vivo* can lower the off-target effects (see section 1.3.1.3 B).

RNAi system	Advantages	Disadvantages
siRNA	• Efficient delivery methods available • Pre-validated siRNAs available	Transient silencing due to lack of cellular means to amplify and propagate siRNAs and dilution by actively dividing cells
shRNA	• Stable integration • Expression of shRNAs driven by inducible or constitutive promoter system (Gupta et al., 2004; Matsukura et al., 2003; van de Wetering et al., 2003)	Time and labour consuming in order to carry out the cloning of the constructs and the screening for stable transfectants.
shRNA-mir	• Transient and stable silencing • Stronger silencing effect	Comparatively less well-characterised technique than the other two methods

Table 1.2. Advantages and disadvantages of various RNAi systems in mammalian systems.

(iv) Delivery strategies

Various strategies have been developed to deliver siRNA or shRNA into mammalian cell types and they are summarised in Table 1.3.

Delivery strategy	Remarks	References
Electroporation of siRNA	Transient gene silencing	(MacKeigan et al., 2005)
Lipid-based transfection of siRNA	Transient gene silencing	(Aza-Blanc et al., 2003)
Influenza virosomes with encapsulated siRNA	Transient gene silencing	(de Jonge et al., 2006)
Electroporation of shRNA	Transient gene silencing	(Brummelkamp et al., 2003)
Transfection of shRNA-expressing plasmid vector	Generates stable RNAi by random integration and marker selection	(Brummelkamp et al., 2002)
Delivery of shRNA cassettes using retroviral systems	Used in cell types which are difficult to transfect (e.g. primary cells)	(Barton and Medzhitov, 2002)
Delivery of shRNA cassettes using lentiviral systems	Used in primary cells and non-dividing cells	(Moffat et al., 2006; Ngo et al., 2006; Stewart et al., 2003)
Delivery of shRNA cassettes using adeno-associated virus	Stable integration	(Boden et al., 2004a)
Delivery of shRNA cassettes using adenovirus	Stable integration	(Cao et al., 2005)

Table 1.3. Delivery strategies of RNAi in mammalian systems.

1.3.1.3 Non-specific effects of RNAi

Although PKR activation is not effectively triggered by siRNAs in mammalian systems, other non-specific effects are induced by RNAi triggers. These include the innate immune response (IFN response), off-targeting and saturation of the RNAi pathway as described below.

A. IFN response

dsRNAs which are longer than 30 nucleotides were found to trigger the PKR response in mammalian cells (Williams, 1997). However, it has also been demonstrated that transfection of siRNAs can activate PKR which results in the triggering of the interferon (IFN) pathway and induces a global upregulation of IFN-stimulated genes (ISGs) (Sledz et al., 2003). Some upregulated ISGs are dependent on siRNA concentration while others are not. siRNAs without 2- to 3-nucleotide 3' overhangs were shown to be recognised by the IFN system via the RNA helicase RIG-1 (Marques et al., 2006). These overhangs are the structural characteristics which distinguish synthetic siRNAs from endogenous Dicer-generated ones. siRNAs without 3' overhangs are more likely to be unwound, and this mediates IFN activation. The IFN response is also induced by H1 or U6 promoter-generated shRNAs *in vivo* (Bridge et al., 2003; Pebernard and Iggo, 2004). More detailed analyses of the U6 promoter vectors indicated that ISG induction is a consequence of the presence of an AA dinucleotide motif near the transcription start site of shRNAs.

B. Off-target effects

Off-target effects were first studied by examining the expression profiling of numerous siRNAs directed against the same target genes (Jackson et al., 2003). It was shown that a majority of gene expression patterns were siRNA-specific rather than target-specific. Off-target effects can be elicited by as few as 11 nucleotides of identity between the siRNA and its target. In other studies, different siRNAs against the MEN1 gene were characterised to induce variations in expression levels to different degree in p53 and p21, which are indicators of overall changes in cellular physiology (Scacheri et al., 2004). Off-target effects were further characterised by the ability of various siRNAs to induce changes in cell toxicity in a target-independent manner which generates toxic phenotypic changes (Fedorov et al., 2006). This toxic effect was found to relate to a UGGC motif in the siRNAs. Additional studies have been performed to understand the mechanism of off-target effects. It has been confirmed in various studies that the off-target transcripts have 3' UTR sequence partial complementary to the seed region of the siRNA which is similar to the endogenous miRNA pathway (Birmingham et al., 2006; Jackson et al., 2006).

C. Saturation of the RNAi pathway

Saturation of the RNAi pathway happens when there is an excessive dose of siRNA administered to cells or when shRNAs are highly expressed. Saturation can occur at different levels, depending on the siRNA/shRNA used. Depletion of Dicer and Ago2 was found to up-regulate expression of a large number of genes whose 3' UTRs show an enrichment of putative miRNA target sites (Schmitter et al., 2006). shRNAs or shRNA-mirs may also saturate Drosha or Exportin. Prolonged expression of shRNAs was found to be lethal in mice due to a saturation of Exportin 5 (Grimm et al., 2006).

1.3.1.4 Applications of RNAi

A. Study of gene functions and downstream pathways

RNAi technology has been used to knockdown the expression of specific genes which are of particular interest in certain pathways and diseases. In many cases, it has been used in conjunction with gene expression profiling to identify downstream target genes in signalling pathways (Jazag et al., 2005) and to understand disease mechanisms (Diakos et al., 2007). RNAi, together with expression analyses on microarray and computational or experimental promoter studies, have also been use to dissect transcriptional networks of key transcription factors involved in apoptosis (Elkon et al., 2005) and in embryonic stem cell self-renewal (Jiang et al., 2008).

B. Genetic screens

With the completion of the genome sequence of human and various model organisms, RNAi has been exploited as a tool to screen genes involved in specific pathways or disease. In contrast to the application above, a number of genes in the genome included in an RNAi library are knocked down simultaneously and their effects on specific pathway are studied to identify the genes which are important for the pathway. RNAi-based genetic screens have advantages over conventional knockouts as generating RNAi libraries is relatively cheap and easy. Genetic screens in *C. elegans* and *Drosophila* are summarised in Table 1.4.

Biological pathway studied	Organism	Genome coverage	Reference
Cell division	*C. elegans*	Chromosome III genes	(Gonczy et al., 2000)
DNA-damage responses	*C. elegans*	Whole genome	(van Haaften et al., 2006)
Embryonic developments	*C. elegans*	Whole genome	(Sonnichsen et al., 2005)
miRNA pathway	*C. elegans*	Whole genome	(Parry et al., 2007)
Hedgehog signalling pathway	*D. melanogaster*	Kinases and phosphatase; 43% predicted genes	(Lum et al., 2003)
Cell viability and growth	*D. melanogaster*	Whole genome	(Boutros et al., 2004)
Embryonic development	*D. melanogaster*	Whole genome	(Koizumi et al., 2007)
Chromatin-related transcriptional repression	*D. melanogaster*	Whole genome	(Stielow et al., 2008)

Table 1.4. RNAi genetic screen in *C. elegans* and *Drosophila*. This table summarises the biological pathways studied and the genome coverage of library in *C. elegans* and *Drosophila* RNAi genetic screens.

Genetic screens have also been widely used in mammalian systems (Table 1.5). Due to the size of mammalian genomes and their gene content, initial efforts of RNAi screening focused on libraries representing subsets of genes implicated in various processes and pathway. However, siRNA libraries representing the known human and mouse gene sets are now commercially available from several suppliers such as Ambion, Qiagen and Dharmacon. For stable integration of shRNAs, retroviral or lentiviral libraries targeting either subsets or all human genes have been generated and successfully employed in RNAi screens. Most conventional methods of RNAi screening involve the use of a single well/single gene approach which is relatively time-consuming. A small-scale pooled retroviral vector strategy involving the use of barcoded shRNAs and analyses on microarray was first described by Berns et al (2004) and has since been widely used. This greatly enhances the efficacy of global screening. Second generation plasmid-based shRNA-mir libraries covering all genes in the human and mouse genomes have also been described (Silva et al., 2005). Such libraries enable single-copy expression of the shRNAs which is important for pooled screening applications.

Biological pathway or disease studied	Form of RNAi trigger	Source of siRNA/ delivery of shRNA	Library genome coverage	Analysis strategy/ screen format	Reference
Negative regulation of phosphorylation	siRNA	Chemically synthesised	Phosphoinositide 3-kinase (PI3K) pathway human genes	Single well/ single gene	(Hsieh et al., 2004)
Cell division	siRNA	Endoribonuclease-prepared	> 15000 human genes	Single well/ single gene	(Kittler et al., 2004)
Huamn immunodeficiency virus (HIV) infection required host protein	siRNA	Chemically synthesised	All genes in the human genome (Dharmacon library)	Single well/ single gene	(Brass et al., 2008)
p53 pathway	shRNA	Retroviral delivery	~ 8000 human genes	Single well/ single gene and bar-coded siRNA + microarray screen	(Berns et al., 2004)
RAS activation pathway and tumorigenesis	shRNA	Retroviral delivery	~ 4000 human genes	Soft agar colony growth	(Kolfschoten et al., 2005)
Tumorigenesis	shRNA	Retroviral delivery	~ 7500 human genes	bar-coded siRNA + microarray screen	(Westbrook et al., 2005)
Diffused large B-cell lymphoma	shRNA	Retroviral delivery (Inducible expression)	~ 2500 human genes	bar-coded siRNA + microarray screen	(Ngo et al., 2006)
Cell division and proliferation	shRNA	Lentiviral delivery	~ 12000 human genes	Single well/ single gene	(Moffat et al., 2006)
Cell proliferation in mammary cells	shRNA-mir	Retroviral delivery	All genes in the human genome	bar-coded siRNA + microarray screen	(Silva et al., 2008)
Cell proliferation in cancer cells	shRNA-mir	Retroviral delivery	All genes in the human genome	bar-coded siRNA + microarray screen	(Schlabach et al., 2008)

Table 1.5. RNAi genetic screens in mammalian systems. This table summarises the diseases or biological pathways studied using RNAi screen in mammalian systems. The type and delivery of the RNAi trigger, library coverage and analysis methods are also described.

C. Disease therapy

In addition to using RNAi as an experimental tool, numerous studies have documented applying RNAi technology therapeutically in the treatment of various human diseases. A variety of RNAi triggers and delivery methods have been tested and summerised in Table 1.6. Ultimately, the choice of method of RNAi therapy depends on the disease and organ under treatment. For instance, siRNAs can be degraded by serum nucleases and can only provide a short-term suppression of gene

expression in a specific subset of tissues/organs, thus limiting their therapeutic benefits. Although shRNA delivery is more challenging, it initiates a more sustained therapeutic effect and provides treatment options to a broader range of diseases including viral infections and cancers. In particular, an *ex vivo* treatment protocol has been developed for HIV infection and is now under phase I clinical trial. Here cultured haematopoietic stem cells of HIV patients are incubated with lentiviral vectors carrying the anti-HIV shRNA. These HIV-resistant stem cells are then transplanted into the bone marrow of HIV-affected patients and allowed to proliferate and replace diseased cells.

Whatever method of delivery and RNAi trigger being used, one of the most important issues to be considered and overcome is the safety of RNAi. As mentioned before, RNAi induces a number of innate immune responses and silences non-specific targets which should be taken into account when designing the siRNA or shRNA trigger. Also, due to the high mutation rate in viruses, using a combination of multiple shRNAs against the viral genome, and also against host genes required for infection, are important issues to consider.

Disease	RNAi trigger	Delivery strategy	Route of administration	Reference
Respiratory syncytial virus (RSV) infection	siRNA	Complexing siRNA with nanoparticle	Intranasal	(Zhang et al., 2005a)
Apolipoprotein B (ApoB) related disease	siRNA	Encapsulating siRNA with stable nucleic acid particles (SNALP) Cholesterol-conjugated siRNA	Intravenous injection	(Zimmermann et al., 2006) (Wolfrum et al., 2007)
Human immunodeficiency virus (HIV) infection	siRNA	Complexing siRNA with antibody conjugate	Intravenous injection	(Song et al., 2005)
Haptitis B virus (HBV) infection	shRNA	Adeno-associated virus serotype 8 vector	Intrasplenic injection	(Chen et al., 2007)
Human immunodeficiency virus (HIV) infection	shRNA	Lentiviral vector	*Ex vivo* treatment and transplant	(Li et al., 2006)
Cancer (supression of Interleukin 8, a factor for tumor growth and metastasis)	shRNA	Adenoviral vector	Mouse model xenograft	(Yoo et al., 2008)

Table 1.6. Therapeutic intervention using RNAi. This table summarises the development of RNAi in the treatment of human diseases. The type of RNAi trigger, delivery strategies and route of administration are described.

1.3.2 Gene expression profiling

Measuring the expression of genes in various tissues, different stage during development or during perturbation experiments is essential for understanding complex transcriptional programmes. There are many different ways to profile the expression pattern of genes. These range from traditional low-throughput methods to genome-scale high-throughput methods as described below:

1.3.2.1 Conventional methods

A. Northern blotting

For northern blotting (Alwine et al., 1977), RNA samples are first separated by size via electrophoresis in an agarose or polyacrylamide gel under denaturing conditions. The RNA is then transferred to a membrane, crosslinked and hybridised with a radiolabelled or nonisotopically-labelled probe.

B. Nuclease protection assays

Nuclease protection assay (NPA) involves target-probe hybridisation in solution. A single-stranded labelled probe is incubated with an RNA sample so that DNA-RNA or RNA-RNA hybrids are formed. The mixture is then exposed to ribonucleases that specifically cleave only single-stranded RNA. The probe:target hybrids are precipitated and separated on a denaturing polyacrylamide gel and are either visualised by autoradiography or by secondary detection.

C. Differential display

Differential display (DD) is a technique involving PCR without relying on prior knowledge of gene sequences that can be used to isolate differentially expressed genes (Liang and Pardee, 1992). The mRNA samples are reverse transcribed into cDNA, amplified by PCR and labelled with radioisotopes or fluorescent dyes and separated by denaturing polyacrylamide gels. The cDNA populations from different samples can be visualised and compared, and differentially expressed genes can be identified and sequenced.

D. Quantitative real-time polymerase chain reaction (qRT-PCR)

Quantitative real-time PCR is a highly sensitive kinetics-based quantification technique where PCR products are measured in real time to monitor the concentration of nucleic acids. In qRT-PCR, PCR is performed as normal with a pair of oligonucleotide primers, however fluorescent dyes are used to measure the amount of PCR product. There are two different types of approaches that have been used which will be discussed in Chapter 3. qRT-PCR is commonly used for the validation of microarray data and for quantification where the starting material is limited. It has also been used in combination with other techniques in the study of transcriptional networks during macrophage activation (Nilsson et al., 2006).

Table 1.7 summarised the major advantages and disadvantages of conventional gene expression profiling techniques.

Conventional technique	Advantages	Disadvantages
Northern blotting	• Cheap • Simple procedures	• Not suitable for accurate quantitative analyses • Low-throughput • Low sensitivity • Target transcript sequence required
Nuclease protection assay	• Cheap • Simple procedures • Higher sensitivity • Mutli-probe analyses possible	• Not suitable for accurate quantitative analyses • Low-throughput • Target transcript sequence required
Differential display	• Target transcript sequence not required • Relatively high-throughput	• Not suitable for accurate quantitative analyses • Sequencing may be required
qRT-PCR	• Highly sensitive • Accurate quantitative analyses	• Low-throughput • Target transcript sequence required • Expensive

Table 1.7. Major advantages and disadvantages of conventional methods of gene expression profiling.

1.3.2.2 Sequencing-based methods

A. Serial analysis of gene expression (SAGE)

Serial analysis of gene expression (SAGE) is a sequencing-based technique used to measure gene expression (Velculescu et al., 1995) (http://www.ncbi.nlm.nih.gov/projects/SAGE/). Biotinylated double-stranded cDNAs are cleaved with a restriction enzyme which has a 4-bp recognition site. The 3'ends of the cDNAs are then collected with streptavidin beads. The cDNAs are separated into two pools, ligated with two different linkers (A and B) and each of them are cleaved by a type IIS restriction enzyme which cuts up to 20 bp downstream of the recognition sites. The cleaved fragments, each containing a gene-specific tag, are concatenated (by amplification with primers against the linkers A and B), cloned and sequenced. The number of each gene-specific tag is quantified and the tags are mapped to the annotated genome.

B. Cap analysis of gene expression (CAGE)

Cap analysis of gene expression (CAGE) is a very similar technique to SAGE (Kodzius et al., 2006; Shiraki et al., 2003). Instead of creating signature tag at the 3' end, CAGE clones the 5' ends of cDNA fragments. cDNAs are generated with random primers and isolated by a biotin cap trapper method, where the 5' cap of the mRNA is biotinylated and removed by streptavidin beads. Linkers are attached to the 5' ends of cDNAs to introduce a recognition site for the restriction enzyme. After amplification, the sequencing tags are concatenated for high-throughput sequencing.

C. Massively parallel signature sequencing (MPSS)

Massively parallel signature sequencing (MPSS) combines the technique of non-gel based sequencing and *in vitro* cloning of DNA fragments onto microbeads (Brenner et al., 2000a). Recent developments in sequencing technology such as the Illumina Solexa (Bennett, 2004) and the 454

sequencing platforms (Margulies et al., 2005) have also been developed with a similar principle. The initial steps of *in vitro* cloning are similar to other tagging approaches except that the plasmids used for cloning contains a tag sequence (Brenner et al., 2000b). PCR products having this tag sequence are generated and attached onto the microbeads carrying the anti-tag sequence by base-pairing. Each of these microbeads carries about 10^5 copies of the same cDNA fragments. The high concentration of DNA templates on the microbeads allows high-throughput sequencing to be monitored by detecting fluorescent signals from the beads. MPSS and related sequencing methods have been widely used for studying various biological pathways in different organisms (Table 1.8).

Biological pathway studied	Sequencing technology	Reference
Profile human ES cells markers	MPSS	(Brandenberger et al., 2004)
Profile fetal human neural precursor cells markers	MPSS	(Cai et al., 2006)
Identify cancer related genes	MPSS	(Chen et al., 2005)
Generate an atlas of gene expression in various cells and tisses in human	MPSS	(Jongeneel et al., 2005)
Profile microRNA expression patterns in mammalian embryonic development	MPSS	(Mineno et al., 2006)
Gene profiling and technology evaluation in *Drosophila melanogaster*	454 sequencing	(Torres et al., 2008)
Profile microRNA in human ES cells	Illumina/Solexa	(Morin et al., 2008)

Table 1.8. Applications of massively parallel signature sequencing technology in gene expression profiling.

D. Polony multiplex analysis of gene expression (PMAGE)

Polony multiplex analysis of gene expression (PMAGE) can be used to profile gene expression of rare transcripts and genes with low expression levels (<1 copy per cell) (Kim et al., 2007). Samples are subject to sequencing directly bypassing all the library amplification, concatenation and subcloning steps. These cDNAs samples are amplified with 1-micrometer polony beads carrying adapter primers in emulsion PCRs. Polony beads carrying DNA templates are cross-linked to aminosilylated glass with amino-ester bridges. Thus an *in vitro* library is generated for high-throughput sequencing.

1.3.2.3 Microarray-based methods

Microarray technology was first described in 1995 for quantitative expression analysis in *Arabidopsis* (Schena et al., 1995). Microarrays are libraries of DNA sequences from a genome which are arrayed at high density on a solid support. Since they were first described, the technology has advanced significantly and has been widely used in the expression studies in various organisms. To date, various microarray platforms are available as described below.

Spotted arrays (genomic clones, cDNAs, PCR products or oligonucleotides) were first developed for using array technology. Initially, double-stranded cDNAs were spotted onto glass microscope slides by a robotic device (Schena et al., 1995). The glass slides are usually coated with reactive

molecular groups such as poly-L-lysine or epoxy for DNA fragments to immobilise onto the surface. The major disadvantages of cDNA/PCR product arrays are that it is difficult to control non-specific hybridisation and hybridisation efficiency due to variations in the GC content of cDNAs. In contrast, spotted oligonucleotide arrays are usually 40 to 60-mers and are single-stranded. Thus, problems associated with cross-hybridisation and variations in hybridisation efficiency are theoretically significantly reduced for these arrays. Major drawbacks of spotted arrays lie in the discrepancy among different batches of arrays and the relatively low density of oligonucleotides that are immobilised onto the glass slides.

Oligonucleotides can also be directly synthesised at high density on the surface of the array by photolithography (Affymetrix) (Singh-Gasson et al., 1999), programmable optical mirrors (NimbleGen) (Lipshutz et al., 1999) and ink-jet devices (Agilent) (Hughes et al., 2001). The BeadArray technology (Illumina) has also been developed for synthesis of high density oligonucleotide arrays (Kuhn et al., 2004). This involves the assembly of silica beads carrying hundreds of thousands of copies of a specific oligonucleotide in microwells on fibre optic bundles or planar silica slides.

For comparing expression levels of genes in different RNA populations, either a two-colour or one-colour labeling approach can be used (Figure 1.7). For spotted arrays, target and reference samples are labelled with fluorescent dyes such as Cy3 and Cy5 and hybridised on the same array. The labelled samples will bind to the DNA sequences on the array in a competitive manner and the fluorescence intensities of the two channels are quantitated. For other types of array such as Affymetrix GeneChips and Illumina BeadArray, a one-colour approach is used where different samples labelled with the same fluorescent dye are hybridised onto separate arrays. The fluorescence intensity of a single channel is quantitated and then compared across separate arrays hybridised with either the target or the reference sample.

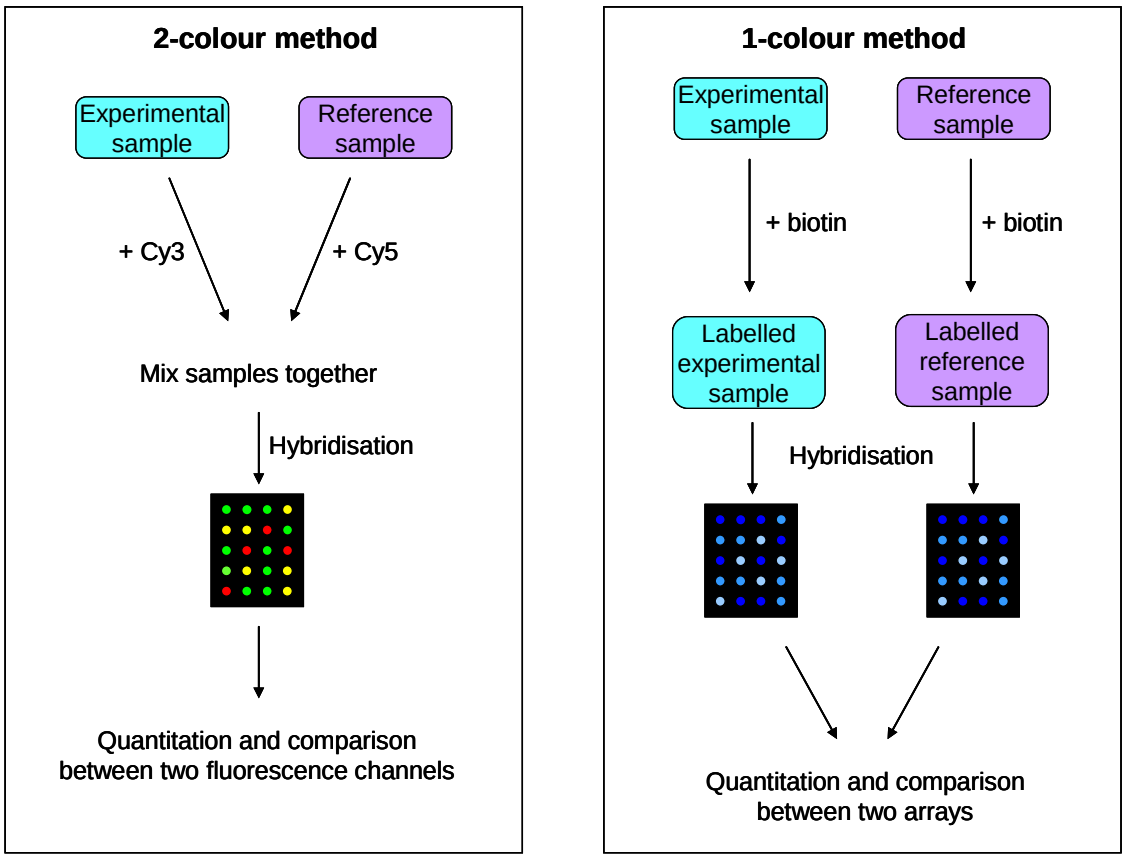

Figure 1.7. Principles of 1-colour and 2-colour microarray hybridisation. Left panel: 2-colour method; right panel: 1-colour method. In the 2-colour approach, experimental and reference samples are labelled, each with one of two different fluorescent dyes. The labelled samples are mixed together and hybridised to the same array. The two fluorescence channels are quantitated and compared. In the 1-colour approach, the experimental and reference samples are labelled with the same fluorescent dye/detection system and the labelled samples are hybridised to two separate arrays. The two arrays are quantitated and compared.

1.3.2.4 Applications of microarrays in gene expression profiling

A. Identification of pathway-specific genes

Typically, global gene expression profiles are monitored throughout a temporal program at different times within a pathway or at different stages of a developmental process. The DNA microarrays usually contain the whole genome or a complete set or subset of open reading frames (ORFs) of the organism. Differentially expressed genes are identified by comparing the expression profile at different time points. For example, microarrays have been used to study metabolic pathways in *S. cerevisiae* and to identify developmental-specific genes of metamorphosis in *Drosophila* (DeRisi et al., 1997; White et al., 1999).

B. Identification of downstream targets using genetic perturbation

Microarray expression profiling is commonly used to identify the effects on patterns of expression which occur when a biological system is perturbed for a gene of interest (for example, knockouts,

over-expression, knockdowns, ectopic expression, or introduction of mutations). Such approaches have been successfully used to identify downstream target genes of some key genes involved in cancer, for example, c-myc, p53 and ras (Coller et al., 2000; Milyavsky et al., 2005). However, this method of profiling cannot distinguish between primary and secondary target genes. Ultimately, determination of direct regulatory programmes controlled by a specific gene product (for example, a transcription factor) must be accompanied by other approaches such as ChIP (see section 1.3.3).

C. Profiling of human diseases and therapeutic responses

Gene expression profiling has also been used to study the molecular basis and identify the gene signatures of human cancer by comparing and classifying patient samples (Ferrando et al., 2002; Ge et al., 2006). It has also been used to study therapeutic effects of drugs and other treatments (Gyorffy et al., 2005; Marton et al., 1998; Shipp et al., 2002).

1.3.3 Characterisation of regulatory elements

Studying only mRNA expression patterns within a biological system cannot allow us to unequivocally identify direct target genes of transcription factors, because they fail to provide evidence about DNA-TF binding events. However, physical interaction between transcription factors and DNA can be determined both experimentally and computationally. This allows us to study where and how the transcription factor regulates the transcription of its target gene, which is important empirical evidence to support our understanding of transcriptional networks.

1.3.3.1 Conventional methods

A. DNase I hypersensitivity assays

As mentioned in section 1.1.2.5, chromatin structures are modified by the combinatorial action of chromatin-remodelling complexes and histone modification rendering the exposure of nucleosome-free DNA. Nucleosome-free DNA regions are often a characteristic of regulatory elements. These nucleosome-free DNA regions are extremely sensitive to the cleavage by DNase I - thus, they are regarded as DNase I hypersensitive sites (HSs). HSs have been shown to be associated with regulatory elements such as promoters, enhancers, silencers etc. (Gross and Garrard, 1988). Traditional and advances of HS assays are summarised in Table 1.9.

DNase I hypersensitivity assays	Desciption	References
Traditional assays	Genomic DNA was cleaved by DNase I followed by DNA purification, restriction enzyme digestion, southern blotting and hybridisation with a labelled probe	(Wu et al., 1979)
DNase I assays coupled with PCR and qPCR	Optimised PCR and quantitative PCR methods were used to increase the resolution, quantitation and sensitivity of the HSs mapping	(Follows et al., 2007; McArthur et al., 2001; Yoo et al., 1996)
Quantitative chromatin profiling (QCP)	Quantitation of tiled amplicons across a locus by quantitative real-time PCR	(Dorschner et al., 2004)
DNase I assays coupled with MPSS	Cloning of hypersensitive sites and analyses by massively parallel signature sequencing (MPSS)	(Crawford et al., 2004; Crawford et al., 2006b)
DNase I assays coupled with tiled microarray	Mapping of HSs by tiled microarray across a particular region of interest and potentially in a global manner in any sequenced genomes	(Crawford et al., 2006a; Follows et al., 2006)
In silico prediction of HSs	*In silico* prediction of HSs in the human genome using a supervised pattern recognition algorithm with high accuracy	(Noble et al., 2005)

Table 1.9. DNase I hypersensitivity assays and advances.

B. DNase I foot-printing

The DNase I foot-printing assay, also called the DNase I protection assay, is an *in vitro* assay used to identify protein-bound DNA elements (Galas and Schmitz, 1978). The procedure involves radioactively-labeling DNA fragments at one end. The DNA fragments are incubated with or without the protein of interest and then subjected to DNase I treatment followed by electrophoresis and autoradiography. DNA bound by proteins or transcription factors is more resistant to cleavage by DNase I than naked DNA and is absent on the autoradiograph (as gaps in the ladder of end-labelled fragments) and can be regarded as footprints for protein-bound regions.

C. Electrophoretic mobility shift assays

Electrophoretic Mobility Shift Assay (EMSA), also called gel shift assay, is another *in vitro* technique for studying protein-DNA interactions (Garner and Revzin, 1981). The gel shift assay is carried out by first incubating a radioactively-end-labelled or fluorescent-labelled (Onizuka et al., 2002) DNA fragment containing the putative protein binding site with or without a protein of interest. The reaction products are then analysed on a non-denaturing polyacrylamide gel followed by autoradiography. The protein-DNA complexes migrate more slowly than naked DNA and are retarded on the gel compared to the control sample.

Other *in vitro* assays include systematic evolution of ligands by exponential enrichment (SELEX) and cyclic amplification and selection of targets (CASTing), both of which screen pools of nucleic acid ligands with the protein of interest (Tuerk and Gold, 1990; Wright et al., 1991).

D. Reporter gene assays

The identification of putative regulatory elements alone fails to provide information on their activity within the cell. Often the functionality of these elements is tested by reporter gene assays (Weber et al., 1984) which can also be adapted for genome-wide screens. The putative regulatory elements of interest (or random genomic fragments for large scale screening), are cloned into a plasmid containing a reporter gene encoding chloramphenicol acetyltransferase (CAT), β-galatosidase, green fluorescent protein (GFP) or luciferase. The plasmid construct is then transfected stably or transiently into cultured cells by electroporation or lipofection and the activity of the reporter is quantified. The reporter construct is made according to the different type of regulatory elements to be tested (Figure 1.8).

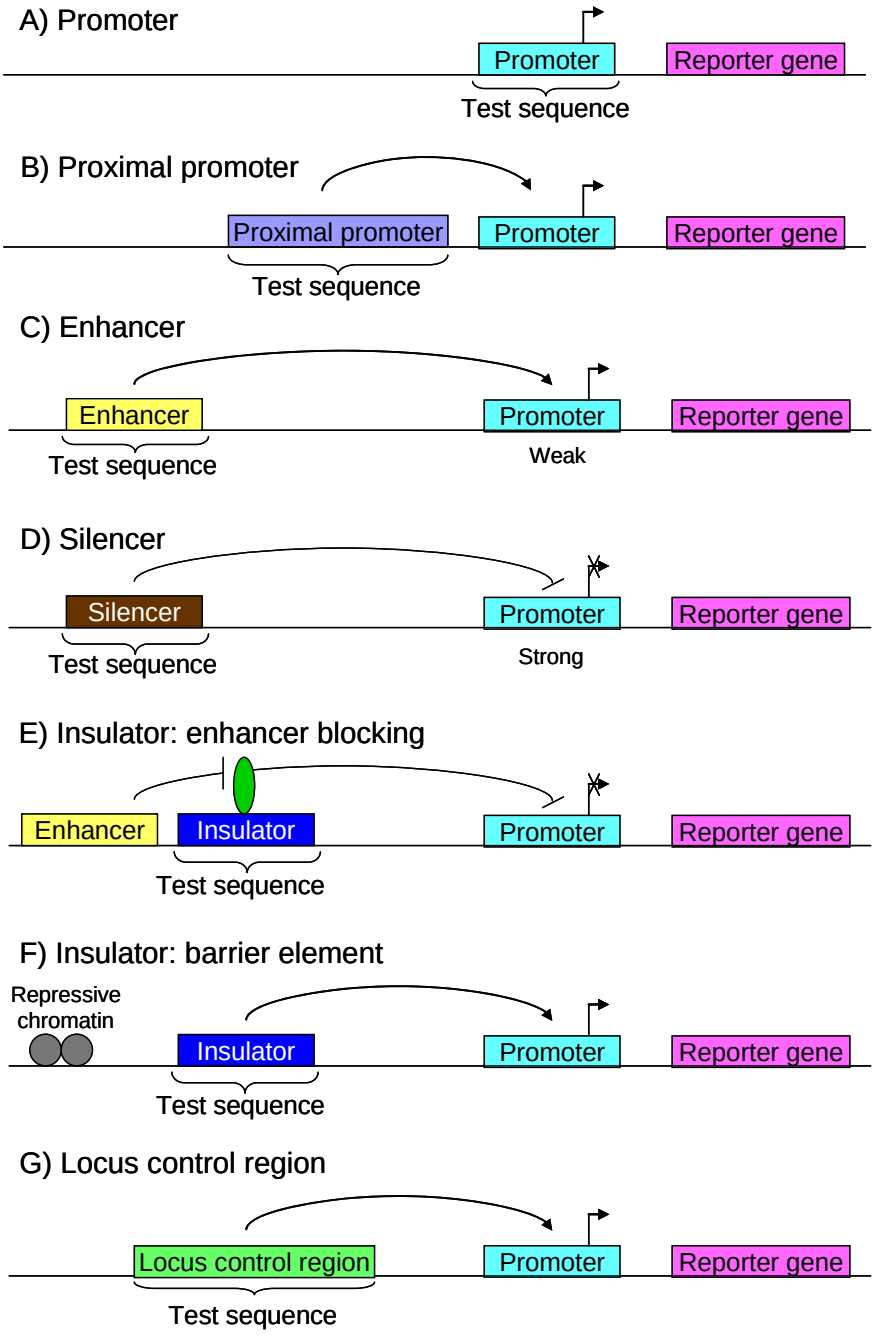

Figure 1.8. Functional reporter gene assays for the identification of regulatory elements. A) A genomic element representing a putative promoter is cloned immediately upstream of a reporter gene lacking an endogenous promoter. B)-D) Sequences representing putative proximal promoters, enhancers and silencers are cloned upstream of a reporter gene directed by an appropriate strength promoter. E) Insulators with an enhancer blocking activity interfere with enhancer-promoter communication and repress gene expression. F) Insulators having a barrier activity avoid the spread of repressive chromatin. G) Locus control regions confer correct gene expression patterns.

1.3.3.2 ChIP-based methods

Chromatin immunoprecipitation (ChIP) is a well developed and powerful technique to study *in vivo* interaction between protein and DNA. This is an approach where protein and DNA can be cross-linked in the native chromatin structure *in vivo* and which overcomes the obstacles presented with the traditional methods which use *in vitro* based assays (Figure 1.9). Cells are grown under the desired experimental condition and fixed with cross-linking agents whilst intact, effectively resulting in covalently interactions between proteins and DNA. The cross-linked chromatins are sonicated to shear the DNA fragments to approximately 200-1000 bp. The protein-DNA complexes are immunoprecipitated with an antibody against the protein of interest. The crosslinks are then reversed and the DNA bound to the protein is purified. The ChIP DNAs can be quantified by Southern blot, PCR or quantitative PCR to identify specifically enriched DNA fragments (Das et al., 2004). An alternative approach named ChIP-on-beads which combines a conventional PCR with tagged primers and captures the products onto microbeads followed by analyses by flow cytometry was developed for larger scale analyses (Szekvolgyi et al., 2006). However, all these methods require prior knowledge of the putative sequence that the protein may bind and are relatively low-throughput. Some of the issues and limitations associated with ChIP are discussed in Chapter 5.

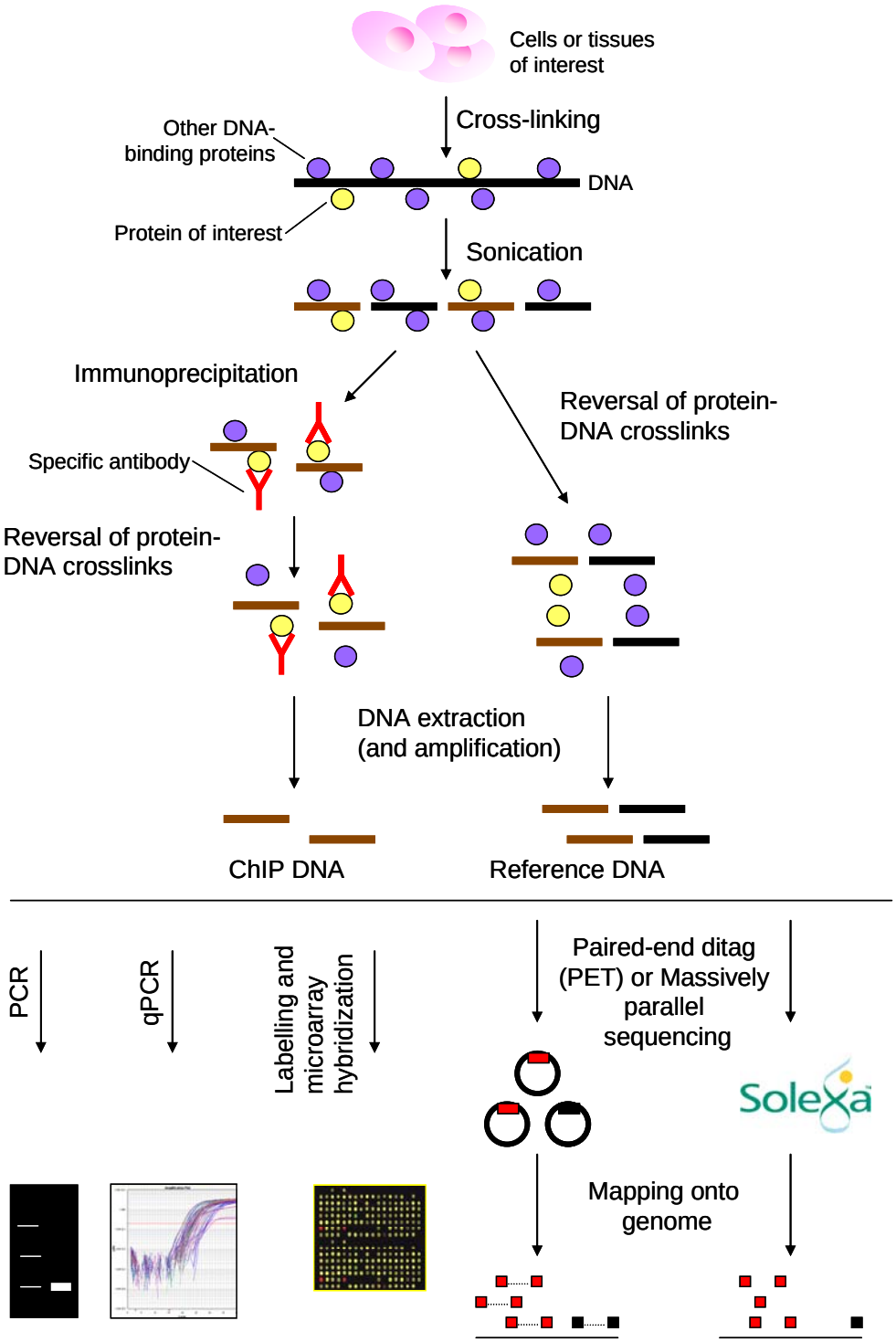

Figure 1.9. A schematic diagram of the chromatin immunoprecipition (ChIP) assay and subsequent analyses. DNA-protein complexes in the cells or tissues of interest are cross-linked, sonicated, and immunoprecipitated followed by reversal of crosslinks and DNA purification (with or without amplification). Purified DNAs are analysed by PCR, qPCR, microarrays, paired-end ditag or massively parallel sequencing.

1.3.3.3 High-throughput ChIP applications

A. ChIP-on-chip

To map protein binding sites on a genome-wide scale, ChIP coupled with microarrays (ChIP-on-chip or ChIP-chip) is an extremely powerful technique which is widely used (Figure 1.9). The DNA

from ChIP is usually amplified or several IPs of material are pooled to provide sufficient DNA for labeling with fluorescent dyes such as Cy3 or Cy5. The labelled DNA is hybridised onto the microarray while DNA that is not immunoprecipitated (or a mock immunoprecipitation) is used as a reference for comparison.

ChIP-on-chip was first successfully demonstrated in yeast, where the complexity of the genome allows one to study all genomic sequences in relatively simple array-based experiments. Table 1.10 summarised some examples of ChIP-on-chip studies performed in yeast.

ChIP-on-chip study in yeast	References
Mapping of binding sites of the transcription factor Rap1	(Lieb et al., 2001)
Mapping of binding sites of the transcription factors SBF and MBF	(Iyer et al., 2001)
Mapping of binding sites of the transcription factors Gal4 and Ste12	(Ren et al., 2000)
Annotation of all transcription factor and DNA binding protein sites across the yeast genome	(Lee et al., 2002)
Study of recruitment of TATA-binding proteins to promoters	(Kim and Iyer, 2004)
Study of recruitment of RNA polymerase II to promoters and open reading frames	(Pokholok et al., 2002)
Study of DNA replication	(Wyrick et al., 2001)
Study of DNA recombination	(Gerton et al., 2000)
Study of chromatin structures and histone modifications	(Bernstein et al., 2002; Kurdistani et al., 2004; Nagy et al., 2002; Pokholok et al., 2005; Robert et al., 2004; Robyr et al., 2002)

Table 1.10. Application of ChIP-on-chip studies in yeast (*S. cerevisiae*).

Because of the complexity of mammalian genomes, ChIP-on-chip studies have traditionally involved the use of arrays which contain features representing only a sub-set of the genome. Human promoter arrays were first used in the mapping of E2F in cell cycle progression and proliferation (Ren et al., 2002). Since then, many similar studies have been performed in human and mouse (Table 1.11). CpG island arrays have also been used in identifying *in vivo* targets of the E2F family, pRb, c-Myc (Mao et al., 2003; Oberley et al., 2003; Weinmann et al., 2002; Wells et al., 2003). CpG island and promoter arrays are particularly useful for studying the global regulation of a particular transcription factor across the whole genome. However, the disadvantage of these arrays is that they only represent a subset of sequences in a genome and, thus, not all transcription factor binding sites can be detected. In otherwords, transcription factors binding to regulatory elements outside promoters and CpG islands such as enhancers and repressors cannot be examined with this biased approach.

To circumvent the limitations of promoter and CpG island arrays in mammals, arrays whose features spanned entire gene loci, chromosomal regions, or whole mammalian genomes have been used in ChIP-on-chip approaches. Tiling path arrays across the human β-globin locus were first

used to map GATA1 binding sites (Horak et al., 2002). Other tiling arrays representing non-repetitive regions in the human genomes were also explored for the study of transcription factor binding and histone modifications (Table 1.11). Using arrays which tiled the ENCyclopedia of DNA Elements (ENCODE) Project (2004a) pilot regions (1% of the human genome) both transcription factors and histone modifications have been characterised (2007; Koch et al., 2007).

ChIP-on-chip study	Organism	Type of array	References
Mapping of E2F in cell cycle progression and proliferation	Human	Promoter array	(Ren et al., 2002)
Mapping of c-Myc and Max in human cancer cells	Human	Promoter array	(Li et al., 2003)
Mapping of HNFs in the human liver and pancreas	Human	Promoter array	(Odom et al., 2004)
Mapping of myogenic regulatory factors (MRFs) and myocytes enhancer factor 2 (MEF2) in muscle development	Human	Promoter array	(Blais et al., 2005)
Mapping of OCT4, SOX2 and NANOG in human embryonic stem pluripotency and self-renewal	Human	Promoter array	(Boyer et al., 2005)
Mapping of the polycomb repressive complexes in the study of embryo developments and pluripotency	Mouse	Promoter array	(Boyer et al., 2006)
Identifying *in vivo* targets of c-Myc	Human	CpG island array	(Mao et al., 2003)
Identifying *in vivo* targets of E2Fs in human cancer cells	Human	CpG island array	(Oberley et al., 2003; Weinmann et al., 2002)
Study of pRb in cell cycles	Human	CpG island array	(Wells et al., 2003)
Mapping of GATA1 binding site across the human β-globin locus	Human	Tiling array	(Horak et al., 2002)
Mapping of the binding sites of CREB (Cyclic AMP-responsive element-binding protein) across the non-repetitive regions of chromosome 22	Human	Tiling array	(Euskirchen et al., 2004)
Mapping of the binding sites of the NF-kappaB family across the non-repetitive regions of chromosome 22	Human	Tiling array	(Martone et al., 2003)
Mapping of binding sites of Sp1, c-Myc and p53 on chromosome 21 and 22	Human	Tiling array	(Cawley et al., 2004)
Study of histone modifications in human chromosome 21 and 22 and comparison with mouse loci	Human and mouse	Tiling array	(Bernstein et al., 2005)
Characterisation of histone modification in 1% of human genome	Human	Tiling array (ENCODE)	(Koch et al., 2007)
Mapping of all promoters in the human genome	Human	Tiling array (All non-repetitive regions)	(Kim et al., 2005b; Kim et al., 2005c)
Mapping of transcriptional initiation sites	Human	Tiling array (All non-repetitive regions)	(Guenther et al., 2007)
Mapping of polycomb repressive complex in human embryonic stem cells and developmental regulators at transcriptional repressive nucleosome regions	Human	Tiling array (Whole genome)	(Lee et al., 2006)

Table 1.11. Application of ChIP-on-chip studies in mammals.

B. Sequencing-based analyses

An alternative approach to analyse the ChIP DNAs is to directly sequence them and map the sequencing reads onto the genome. This is an unbiased approach which is not limited by what is presented on an array. DNA fragments are usually cloned and sequenced (Weinmann et al., 2001). However, the large number of non-specifically immunoprecipitated fragments makes cloning unpractical. The ChIP-display technique concentrates the target DNA sequences and scatters the non-specific ones (Barski and Frenkel, 2004). The ChIP DNAs are digested with a particular restriction enzyme and analysed by gel electrophoresis. DNA elements enriched in the ChIP reaction will show bands of the same size reproducibly on the gel. These concentrated DNA fragment can then be cloned and sequenced. However, ChIP-display is not suitable for mapping histone modifications and transcription factors with a lot of binding sites and cannot produce precise quantification of the enrichments at a particular location. Another technology called STAGE (Sequence Tag Analysis of Genomic Enrichment) was developed which is based on high-throughput sequencing of concatemerised tags derived from target DNA enriched in ChIP (Kim et al., 2005a). Despite its high-throughput nature, this mono-tagging technology leaves ambiguity in the mapping of short sequences onto the genome. A similar technique ChIP-PET (for paired-end ditag) combines the chromatin immunoprecipitation strategy with the paired-end ditag strategy of high-throughput sequencing. The ChIP DNA fragments are cloned to create a library which is further digested and concatemerised to create the PET library having 36-bp signatures of 18 bp of the 5' and 3' ends of the original fragment for sequencing. With the advances in sequencing technology such as the Solexa platform (Bennett, 2004), the time-consuming cloning steps in the above methods can now be circumvented. ChIP-sequencing (ChIP-seq) combines traditional chromatin-immunoprecipitaion with ultra high-throughput Solexa or 454 sequencing platforms for identifying and quantifying enriched DNA elements. Table 1.12 summarised the applications of ChIP coupled with various sequencing techniques in the study of transcriptional regulation.

ChIP coupled with sequencing study	Organism	Sequencing technique	References
Identification of E2F-regulated promoters	Human	ChIP coupled with cloning and traditional sequencing	(Weinmann et al., 2001)
Identification of target genes of RUNX2	Mouse	ChIP-display	(Barski and Frenkel, 2004)
Mapping of target sites of TATA-binding proteins in yeast and E2F4 in human cells	Yeast and human	STAGE	(Kim et al., 2005a)
Mapping of p53 genomic binding sites	Human	ChIP-PET	(Wei et al., 2006)
Mapping of c-Myc genomic binding sites	Human	ChIP-PET	(Zeller et al., 2006)
Delineation of the transcription regulatory networks of Nanog and Oct4 in stem cell pluripotency	Mouse	ChIP-PET	(Loh et al., 2006)
Mapping of *in vivo* binding sites of REST	Human	ChIP-seq (Solexa)	(Johnson et al., 2007)
Generation of high resolution genome-wide map of histone methylation	Human	ChIP-seq (Solexa)	(Barski et al., 2007)
Study of association of histone modification with nucleosomes	Human	ChIP-seq (Solexa)	(Schones et al., 2008)

Table 1.12. Application of ChIP coupled with sequencing techniques.

1.3.3.4 Alternative ChIP approaches

A. DamID assays

An alternative method which circumvents the need for performing ChIP experiments is DamID. This involves the labeling of DNA near the protein binding site (van Steensel and Henikoff, 2000). In this method, the transcription factor of interest is fused with the *Escherichia coli* DNA-adenine methyltransferase (Dam) protein and is expressed in a cell culture system. The Dam protein methylates the adenine base in GATC sites 1.5 to 2 kb around the binding site of the transcription factor-Dam fusion protein. DNAs from this experimental sample and from a control sample, where only the Dam protein is expressed, are extracted, digested with a restriction enzyme (*Dpn* I), labelled and hybridised onto microarrays. This method has been used in *Drosophila* (Orian et al., 2003) and in mammals (Vogel et al., 2006; Vogel et al., 2007).

B. DIP-ChIP

A modification of the ChIP-chip protocol, called DIP-chip, has been developed for the immunoprecipitation of DNA with the protein of interest *in vitro* followed by microarray analyses (Liu et al., 2005). Purified and tagged proteins of interest are mixed with genomic DNA *in vitro* and the protein-bound DNA is isolated by affinity purification, amplified and hybridised onto a genomic array. The advantage of this method is that no specific antibodies are needed as the fusion partner of the protein makes purification much easier.

1.3.4 Computational approaches to study gene regulation

With the completion of genome sequence in many species, computational tools play a significant role in the study of gene regulation and transcription networks, in combination with experimental approaches. Softwares and databases for promoter and transcription factor binding sites (TFBS) predictions and comparative sequence analyses are now widely available.

1.3.4.1 Promoter prediction

Predicting the location of promoters is particularly useful for targeting regions of interest to study with respect to regulatory interactions. However, it is rather challenging considering the core promoter may be distant from the exons and the combination of core elements may differ from promoter to promoter. The most successful programs are based on the analyses of training data sets from known promoter sets as a means of identifying functionally defined sequences conserved across promoters. These programs then scan for these conserved signatures in genomic sequence. These include PromoterInspector (Scherf et al., 2000), FirstEF (Davuluri et al., 2001) and Eponine (Down and Hubbard, 2002). Nevertheless, these programs have limited sensitivity and specificity for genome-scale analyses as they are heavily dependent on the data sets of known promoters (which may have biased representations). Promoters associated with CpG islands are generally well-predicted compared to those which are not (Bajic et al., 2004).

1.3.4.2 Transcription factor binding site prediction

Transcription factor binding sites (TFBSs) are conserved sequences with a certain degree of degeneracy which transcription factors recognise and bind. The binding sites of the most well-characterised transcription factors are compiled in online databases such as TRANSFAC, TRRD and COMPEL (Heinemeyer et al., 1998). Programs such as MATCH (Kel et al., 2003) or online tools such as TESS and TFSEARCH (Akiyama, 1998; Schug, 1997) make use of the TRANSFAC database to identify TFBSs in input genomic sequences. However, one of the major drawbacks of these methods is that there can be a large number of false positive or true negatives owing to the quality of data used initially to populate the databases. Tools such as JASPAR have been developed recently which use more sophisticated statistically-based models of TFBSs (Sandelin et al., 2004).

To overcome the potential problems mentioned above, more intuitive motif discovery approaches identify sets of common sequence motifs in the upstream regions of a set of genes which are likely to be co-regulated. This allows researchers to identify known as well as novel motifs that might be associated with a transcription factor. The algorithms available include AlignACE (Roth et al., 1998), MEME (Bailey and Elkan, 1995), MDScan (Liu et al., 2002) and NestedMICA (Down and Hubbard, 2005). Such method has been used to identify sequence motifs or clusters of motifs in the

promoter regions of co-expressed genes inferred from gene expression data in yeast (Segal et al., 2003).

1.3.4.3 Comparative sequence analyses

Comparative sequence analyses have long been used as a tool to identify evolutionally conserved and functionally important DNA sequences. Traditionally, it has been applied to the coding regions of genomes to predict novel genes, and more recently, for the identification of *cis*-regulatory elements. Many algorithms and softwares have been developed to aid these kinds of analyses. These include, but are not limited to, BLAST (Altschul et al., 1990), FootPrinter (Blanchette and Tompa, 2003), PhastCons (Siepel et al., 2005), LAGAN (Brudno et al., 2003) and VISTA (Visel et al., 2007). One of the early applications of comparative sequence analyses in regulatory element prediction identified a new enhancer in the SCL locus (Gottgens et al., 2000). Large-scale genome comparative analyses have also been perfomed recently to identify enhancers (Pennacchio et al., 2006; Woolfe et al., 2005). In particular, Pennacchio et al. (2006) identified a subset of enhancers which are highly active in neuronal development and functionally validated 45% of them using *in vivo* enhancer trap assays.

However, comparative sequence analyses have limitations. First, not all the conserved regions contain functional regulatory motifs (Balhoff and Wray, 2005). Secondly, transcription factor binding sites may not be conserved among species (Dermitzakis and Clark, 2002). One reason to explain this is that TFBSs have some degree of degeneracy. Therefore, perfect sequence conservation at the binding site may not necessarily be required for function. Recently, a ChIP-on-chip study of four tissue-specific transcription factors in mouse and human hepatocytes revealed that many occupied binding sites for these transcription factors are not conserved between the two species (Odom et al., 2007). It was shown in this study that, in many instances, a transcription factor can bind to a particular TFBS in human, but it binds to a completely different site in the mouse, irrespective of whether sequences are conserved between the two species. This suggests that sequence conservation alone cannot predict transcription factor occupancy.

1.4 Haematopoiesis

Haematopoiesis is an accessible mammalian system to study the processes associated with the regulation of gene expression and the relationships between genes and their protein products in transcriptional networks. The study of human haematopoiesis formed the basis of the biological system used in this thesis.

1.4.1 Embryonic origin and lineages of haematopoiesis

Haematopoiesis is the process of formation of mature blood cells from haematopoietic stem cells (HSCs). Pluripotent HSCs differentiate to form various blood progenitor cells which further give rise to mature and terminally-differentiated blood cells in specific lineages. In mammals, haematopoiesis occurs in two consecutive phases: primitive (or embryonic) haematopoiesis in early embryonic development and definitive haematopoiesis in late embryonic development and adults. Various tissues have been demonstrated to serve as the reservoirs of haematopoietic cells and/or sites of haematopoietic differentiation during different time of the developmental and differentiation process. These include yolk sac, para-aortic-spanchnopleura (PAS), aorta-gonad-mesonephros (AGM), liver, spleen and thymus.

The initial phase of blood development, primitive haematopoiesis, first takes place in the yolk sac around embryonic day 7 (E7) in mice or during the second to third week in human gestation. Here the undifferentiated mesodermal cells form extraembryonic blood islands where endothelial cells, precursors for the formation of blood vessels, differentiate at the edges of the mesoderm while primitive erythrocytes form in the interior regions. Thus, both endothelial and haematopoietic lineages are derived from the same origin. There is evidence supporting the existence of a bi-potential common precursor of endothelial and haematopoietic cells: the haemangioblast (Choi et al., 1998). Primitive haematopoiesis results in the production of mainly large, nucleated erythroblasts, as well as some megakaryocytes and macrophages. It is a robust yet transient process to generate large amount of blood cells for growth and development of the young embryo. Primitive haematopoiesis only occurs at early stages of embryonic development until around day 13 (E13), after which time the yolk sac begins to degenerate (Figure 1.10).

Unlike primitive haematopoiesis which is mainly erythropoietic, definitive haematopoiesis gives rise to all haematopoietic lineages (Figure 1.11). Definitive haematopoiesis occurs both in the extra-embryonic yolk sac and the intraembryonic, mesoderm-derived para-aorta-splanchnopleura (PAS) which later contributes to the aorta-gonad-mesonephros (AGM). Therefore; within the yolk sac, definitive progenitor cells are produced by a population of mesodermal cells having a fetal-adult fate rather than purely a primitive fate. This supports the idea that there is a temporal overlap between primitive and definitive haematopoiesis and that they share a common precursor (Kennedy et al., 1997). These definitive progenitors do not mature in the yolk sac, but instead they migrate to other tissues for maturation.

Definitive haematopoiesis is mainly derived from haematopoietic stem cells (HSCs). HSCs are defined as a cell population which can contribute to the long-term repopulation of the haematopoietic system of irradiated adult mice. HSCs are required for haematopoietic development

during the entire life of an organism. HSCs are characterised by their ability to self-renew and the expression of markers such as CD34 and c-kit. There has been some controversy regarding the origin of HSCs. The yolk sac has long been regarded as the site of HSCs generation as removal of yolk sac was shown to abolish haematopoiesis in the embryo (Moore and Metcalf, 1970). More recent studies have also isolated HSCs in yolk sacs prior to day 9 (E9) and confirmed their long-term multilineage activity (Yoder et al., 1997). It has been proposed that HSCs produced in the yolk sac migrate to the AGM which serves as a reservoir of HSCs. However, there are also findings opposing the yolk sac as the unique origin of HSCs. The AGM was shown to generate and expand the population of HSCs from day 10 (E10) (Medvinsky and Dzierzak, 1996).

Differentiation of HSCs does not occur in the AGM. Instead, HSCs circulate to other intraembryonic tissues such as the fetal liver for terminal differentiation and maturation of haematopoietic cells (Godin et al., 1999). Here enucleated erythrocytes producing adult globins as well as myeloid cells become mature and appear in the circulation around E12. At the same time, the fetal thymus is the site for T-lymphoid development. The fetal spleen becomes the main site of haematopoiesis during late embryogenesis until around the time of birth, when the bone marrow becomes the major site of haematopoiesis throughout the life of the animal (Godin and Cumano, 2002; Kumaravelu et al., 2002).

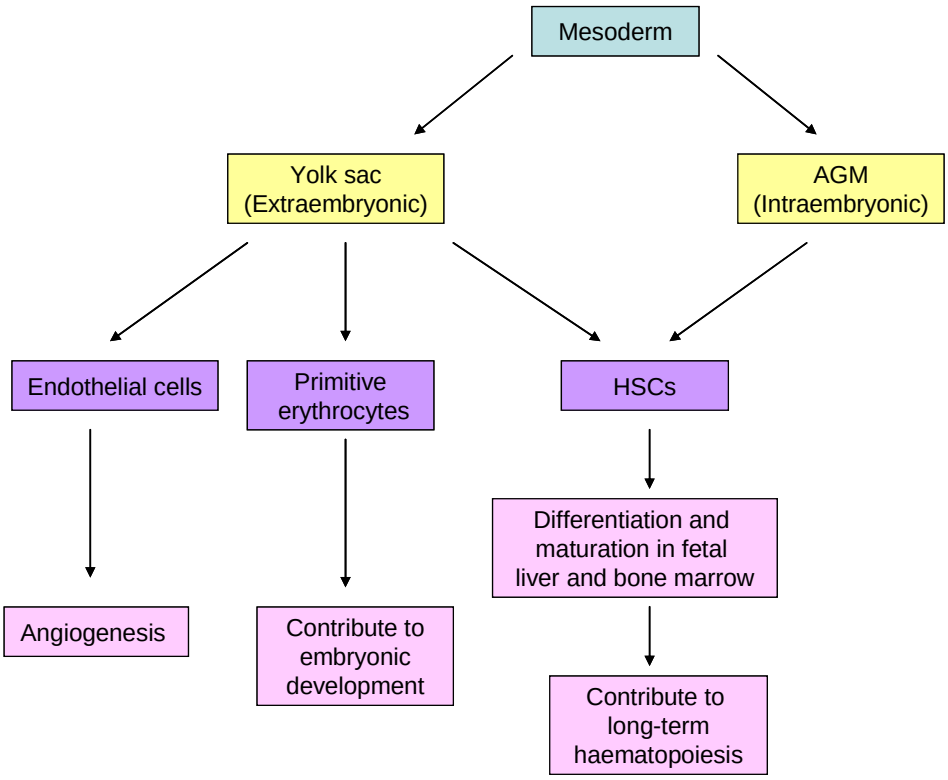

Figure 1.10. A flow diagram of the development of the endothelial lineage and primitive and definitive haematopoeisis from their embryonic origins. The extraembryonic yolk sac from the mesoderm gives rise to endothelial cells, primitive erythrocytes and haematopoietic stem cells (HSCs). Later in embryonic development, the

intraembryonic aorta-gonad-mesonephros (AGM) also gives rise to HSCs. Endothelials cells are implicated in vasculogenesis and angiogenesis. Primitive erythrocytes are critical for supporting embryonic development. HSCs migrate to the fetal liver or bone marrow for differentiation and maturation of various blood lineages and contribute to long-term haematopoiesis.

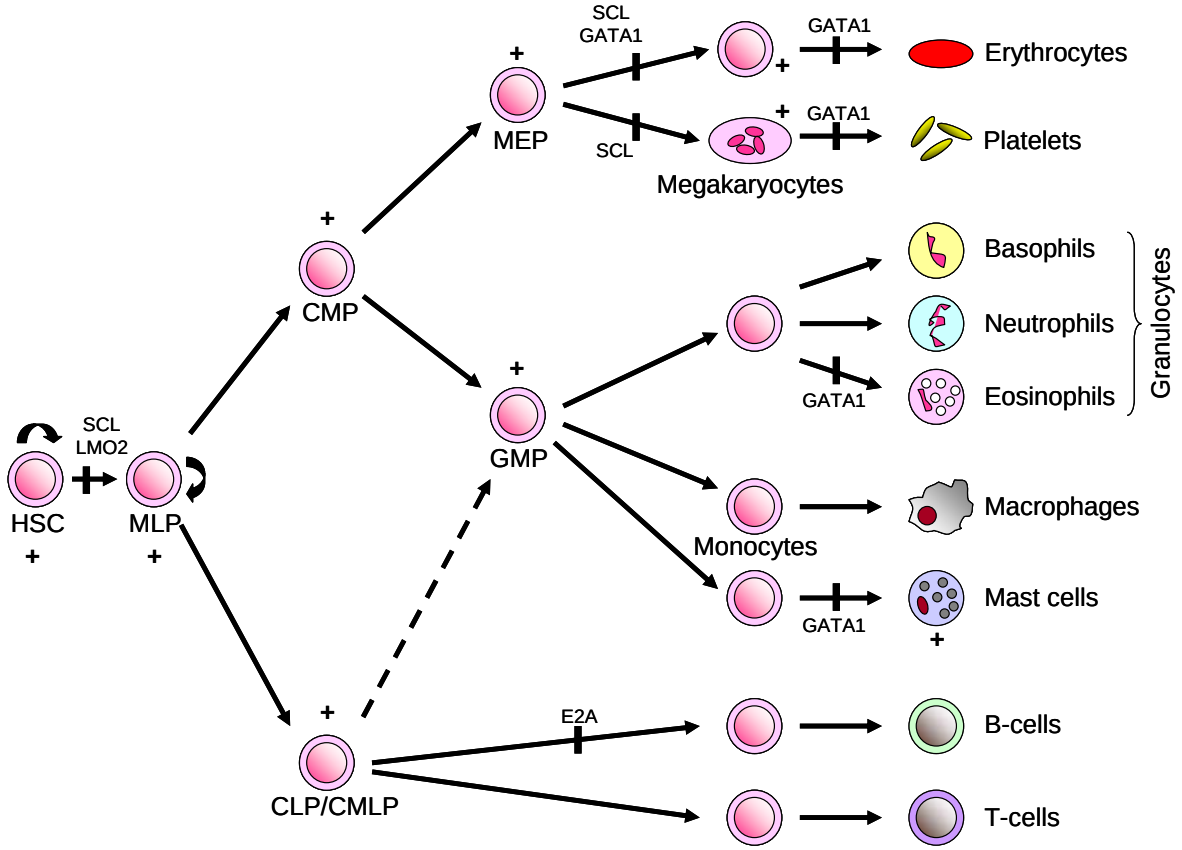

Figure 1.11. A schematic digram of haematopoietic lineage pathways from pluripotent haematopoietic stem cells to mature blood cells. The haematopoietic stem cell (HSC) is the highest in the hierarchy and gives rise to multi-lineage progenitors (MLP) which differentiate to form common myeloid progenitors (CMP) and common lymphoid progenitors (CLP) or common myeloid lymphoid progenitors (CMLP). CLPs give rise to B-cells or T-cells while CMPs give rise to megakaryocyte-erythrocyte progenitors (MEP) or granulocyte-monocyte progenitors (GMP). Alternatively, CMLPs can give rise to GMPs, B-cells and T-cells. GMPs further differentiate to form granulocytes, macrophages and mast cells while MEPs give rise to erythrocytes and platelets. The + signs in the diagram show the major expression pattern common to both SCL and GATA1. Important roles of various transcription factors (discussed in the text) in the haematopoietic lineages are indicated. The bold bars represent developmental blocks when the corresponding protein is removed. Figure modified from Ferreira et al. (2005).

1.4.2 Regulation of haematopoiesis

Haematopoietic commitment and differentiation is regulated by a tightly controlled transcriptional regulatory programme. At different stages of development, different combinations of transcription factors are expressed to further regulate expression of downstream haematopoietic specific genes in the cascade. Thus, a complex transcription network is involved to govern the molecular mechanism leading to differentiation of specific blood lineages. Often transcription factors expressed together

at certain stages form multi-protein complexes which work co-operatively for downstream regulation to take place. The master regulator of haematopoiesis, SCL or TAL1, together with the some of its interacting partners, is discussed below.

1.4.2.1 SCL

A. The SCL gene

SCL, also named TAL1, is a transcription factor which is thought to be a master regulator of haematopoietic development (Begley et al., 1989b). It was first identified in T-cell acute lymphoblastic leukemia (T-ALL) which is resulted from the translocation of human chromosome 1 at p32-33 with the T-cell receptor (TCR) delta locus on chromosome 14q11 (Begley et al., 1989a). The translocation results in the expression of a fusion transcript of TCR-delta and an unknown gene which was thereafter named the stem cell leukaemia (SCL) gene due to its involvement in T-ALL.

The human SCL gene is composed of 8 exons spanning 16 kb on chromosome 1p32-33 (Aplan et al., 1990a). The first five exons are non-coding and there are two promoters (the erythroid-specific promoter 1a and the myeloid-specific promoter 1b) located in the 5' non-coding regions. Alternative splicing events occur at this region generating different mRNA species (Aplan et al., 1990a). An additional promoter, located within exon 4 is cryptically active in T-ALL (Bernard et al., 1992). The murine SCL gene locus is structurally very similar and consists of 7 exons distributed across a 20 kb region of mouse chromosome 4 (Begley et al., 1991).

B. Expression patterns of SCL

Since SCL was identified as a leukaemic translocation fusion protein, it has been shown that SCL expression is critical to various stages of haematopoietic development. Many studies have been performed to examine the expression patterns of SCL. It has been found to be expressed in both the haematopoietic and non-haematopoietic compartments, as discussed below.

- **Haematopoietic expression**

SCL was first found to be expressed in fetal liver, regenerative bone marrow, early myeloid cell lines and leukaemic T-cell lines by northern blot analyses (Begley et al., 1989b). It was later established that SCL is also expressed in human and murine erythroid cell lines, mast cell lines and megakaryocytic cell lines (Green et al., 1992; Green et al., 1991). SCL expression was detected in normal human erythroid, mast, and megakaryocytic cell populations by in situ hybridisation and by RT-PCR (Mouthon et al., 1993). In an *in vitro* differentiation study, expression of SCL was detected and increased during the differentiation of embryonic stem cells to embryoid bodies, the *in vitro* counterpart of haematopoietic progenitors (Elefanty et al., 1997). SCL was found to be expressed in the aorta-associated $CD34^+$ high proliferative potential haematopoietic cells which are

proposed to be HSCs present later in fetal liver and bone marrow (Labastie et al., 1998). In addition to early haematopoietic and erythroid cells, the SCL transcript was also found in pre-B cells (Green et al., 1992). Using a combination of fluorescence-activated cell-sorting (FACS) and RT-PCR for quantitative analysis of expression, SCL expression was confirmed in all haematopoietic cells having an erythroid potential and present, but down-regulated, in common lymphoid progenitors (CLP) and granulocyte/monocyte progenitors (GMP) (Zhang et al., 2005b). No SCL expression was detected in immature and mature cells of the non-myeloid lineage (Figure 1.11).

- **Non-haematopoietic expression**

The overlapping origin of endothelial and haematopoietic lineages suggests a similar expression of key regulators. The SCL transcript was first demonstrated to be expressed in endothelial cells in the spleen (Hwang et al., 1993). The SCL protein was subsequently detected in endothelial progenitors in blood island and in endothelial cells in a variety of tissues including spleen, thymus, placenta and kidney (Kallianpur et al., 1994; Pulford et al., 1995). SCL mRNA has been detected in the endothelial cell clusters of the ventral endothelium of the aorta (Labastie et al., 1998).

SCL is also expressed in the nervous system. It was shown to be expressed in the human neuroepithelial cell lines (Begley et al., 1989b) and subsequently in the murine post-mitotic neurons in the metencephalon and roof of the mesencephalon (Green et al., 1992). Using a knock-in mouse, SCL was shown to be widely expressed in the thalamus, midbrain and hindbrain in the adult and the developing embryonic central nervous system (van Eekelen et al., 2003). SCL expression has also been described in vascular and smooth muscle cells in the aorta and bladder and uterine smooth muscle cells, and in the developing skeleton (Kallianpur et al., 1994; Pulford et al., 1995).

C. Co-ordinated expression pattern of SCL together with GATA1 and LYL1

The expression profiles of SCL and the transcription factors GATA1 and LYL1 are shown to be highly similar in a number of studies. Both SCL and GATA1 were co-ordinately expressed in early haematopoietic and erythroid lineages and their expression undergo biphasic modulations during erythroid and myeloid differentiation in mouse (Green et al., 1992). An early transient decrease followed by an increase of both SCL and GATA1 expression was demonstrated for induced erythroid differentiation. An early transient increase, an initial recovery, followed by a prolonged inhibition was observed during myeloid differentiation (Green et al., 1993). GATA1 and SCL were also found to be co-expressed in erythroid, megakaryocytic and mast cell lineage and down-regulated in terminal erythroid and megakaryocytic maturation (Mouthon et al., 1993). Their expression was shown to be restricted to commited progenitor cells ($CD34^+/CD38^+$) but not the most primitive cells ($CD34^+/CD38^-$). In addition, GATA1 and SCL were found to be expressed in

the extraembryonic mesoderm (precursor of yolk sac), although SCL expression was detected earlier than GATA1 (Silver and Palis, 1997).

The bHLH protein LYL1, which is structurally similar to SCL, also forms a translocation fusion protein during T-ALL (Mellentin et al., 1989) and has overlapping expression pattern with SCL in mouse (Visvader et al., 1991). Similar to SCL, LYL1 is expressed in the erythroid and myeloid lineages and in ascular tissues in mice (Visvader et al., 1991). However, unlike SCL, LYL1 is not expressed in the nervous system (Giroux et al., 2007). Its expression is initiated slightly later than SCL during haematopoietic specification, beginning during haemangioblast differentiation (Chan et al., 2007).

D. Functions of SCL

(i) Haematopoietic development and lineage specification.

SCL is one of the earliest acting regulators of haematopoietic development. Ablation of SCL resulted in embryonic death in mice at E9.5 due to the lack of blood cells (Robb et al., 1995; Shivdasani et al., 1995). Further investigation of $SCL^{-/-}$ embryonic stem (ES) cells in a mouse chimera showed that they are unable to contribute to any haematopoietic lineages, which revealed that SCL is required for both primitive and definitive haematopoiesis (Porcher et al., 1996; Robb et al., 1996) although endothelial cells were still observed in the $SCL^{-/-}$ knockout mice (Robb et al., 1996). Rescue experiments in $SCL^{-/-}$ ES cells revealed that SCL is required for primitive and definitive haematopoiesis at the mesodermal stage (Endoh et al., 2002).

Conditional knockout studies in mice, which circumvent the early lethality observed in $SCL^{-/-}$ mice, demonstrated that SCL is crucial for erythroid and megakaryocytic development in adult mice (Hall et al., 2003). Ablation of SCL completely disrupts erythropoiesis and megakaryopoiesis while the myeloid lineage remains unaffected. Primitive progenitors were also shown to lose their ability to generate erythroid and megakaryocytic cells. However, despite being an important gene for erythropoiesis, SCL is not essential for the generation of mature red blood cells in adults suggesting a possible alternative factor governing this process (Hall et al., 2005).

(ii) Endothelial development.

In addition to being a regulator in haematopoietic development, SCL has been shown to play a crucial role in endothelial development and angiogenesis. In a study of a transgenic knock-in disruption of the SCL locus and an separate study using transgenic rescue of $SCL^{-/-}$ embryos, SCL has been shown to be required for the remodelling of capillary networks to form complex branching vitelline vessels in yolk sacs (Elefanty et al., 1999; Visvader et al., 1998).

SCL has been shown to be required for the generation of blast colonies from blast colony forming cells (BL-CFCs), an *in vitro* equivalent of the haemangioblast (Chung et al., 2002; Robertson et al., 2000). However, contradicting studies showed that $SCL^{-/-}$ cells initiate colony growth but cannot generate endothelial and haematopoietic progeny (D'Souza et al., 2005). The ability to give rise to blast colonies can, however, be rescued by ectopic expression of SCL. This suggests that SCL is essential for commitment of haematopoietic and endothelial lineages from haemangioblast but not for its development.

(iii) HSC self-renewal and repopulating activity.

The role of SCL in the development of HSCs has also been defined. Conditional knockout of SCL in mice demonstrated that SCL is dispensable for the long-term repopulating activity and differentiation into myeloid and lymphoid lineage of HSCs, but is required for the genesis of HSCs (Mikkola et al., 2003). Consistent with this finding, Curtis et al. also suggested that SCL is not required for self-renewal of HSCs but is important for their short-term repopulating capacity (Curtis et al., 2004). A contradicting study on enforced expression in long-term SCID (severe combined immunodeficient) mouse-repopulating cells (LT-SRCs), demonstrated that the expression level of SCL plays a pivotal role in the self-renewal and engraftment of HSCs and this regulation requires the DNA-binding domain of SCL (Reynaud et al., 2005).

(iv) T-cell leukaemia.

Chromosomal rearrangement of SCL is the most common cause of T-ALL and results in the activation of SCL expression in T cells, where it is normally down-regulated. The majority of translocations involves the TCR delta locus which results in the disruption of the promoter and 5' regulatory regions of SCL whilst the full-length coding sequence is unaffected (Begley et al., 1989a; Bernard et al., 1991). Translocation breakpoints were also identified downstream of the SCL coding regions resulting in the formation of an amino truncated protein under the cryptic promoter located in exon 4 (Bernard et al., 1992). An additional rearrangement involved a 5' interstitial deletion in the SCL locus which removes the 5' regulatory elements of SCL and the coding sequence of the SIL gene located immediately upstream of SCL. This results in the expression of SCL under the control of the SIL promoter (Aplan et al., 1990b; Bernard et al., 1991).

There are other known molecular mechanisms, related to SCL binding partners (see section F below), by which chromosomal rearrangement induces tumour formation. LMO1 and LMO2 are also targets of chromosomal rearrangement in T-ALL and are found to be co-expressed with SCL in T-ALL (Wadman et al., 1994). Thus, abberant expression of SCL and LMO2 in T-ALL may induce the expression of genes which are normally silent in T cells including RALDH-2 and TALLA-1

although the relevance of these genes to leukaemogenesis remains obscure (Ono et al., 1997, 1998). A different mechanism involves the sequestering of SCL-interacting partners E2A/TCF3 and HEB by the heterodimerisation between SCL and its partners. This disrupts the homodimerisation or heterodimerisation of E2A/TCF3 and HEB resulting in impaired regulation by these proteins (O'Neil et al., 2004).

E. Transcriptional regulation of SCL

To ensure appropriate expression during haematopoietic differentiation, expression of SCL is tightly regulated as described below (Figure 1.12).

(i) Promoters.

Three promoters have been identified to control transcription of SCL and are conserved in mouse and human: promoter 1a, 1b and an additional promoter in exon 4 (active only in leukaemic T-cells and T-ALL) (Bernard et al., 1992). Promoter 1a is active in erythroid and megakaryocytic lineages and mast cells while promoter 1b is silent in erythroid cells but active in primitive myeloid progenitors and mast cells. GATA1 cooperates with SP1 and SP3 to regulate the promoter 1a of SCL in erythroid cells and mast cells (Bockamp et al., 1998; Lecointe et al., 1994). Promoter 1b is active in primitive myeloid cells but functions in a GATA1-independent manner (Bockamp et al., 1997). Transcription factors PU.1, Elf-1, SP1, and SP3 were found to bind to promoter 1b and transactivate promoter 1b in mast cells (Bockamp et al., 1998).

(ii) Enhancers.

DNase I hypersensitivity assays and reporter assays have been used to identify and characterise putative regulatory elements at the SCL locus in human (Leroy-Viard et al., 1994) and mouse (Fordham et al., 1999; Gottgens et al., 1997). It has been demonstrated that the activity of these regulatory elements have overlapping but distinct features in various haematopoietic cell types. The putative enhancers at the human and mouse locus and their usage in different cell types and tissues are summarised in Figure 1.12.

The stem cell enhancer (+17 /+18 in mouse; +20/+21 in human) was demonstrated to be active in erythroid and mast cells but silent in primitive myeloid cells (Fordham et al., 1999; Gottgens et al., 1997). It was shown to activate both promoters 1a and 1b (Fordham et al., 1999). Further characterisation of this enhancer in transgenic mice demonstrated it targets expression in extra-embryonic mesoderm and both endothelial cells and haematopoietic progenitor cells in the yolk sacs, AGM, fetal liver and bone marrow (Sanchez et al., 1999). Exogenous SCL expression driven by the stem cell enhancer was shown to rescue early haematopoietic development in $SCL^{-/-}$ embryos which further strengthens its involvement in SCL regulation during stem cell development

(Sanchez et al., 2001). Analyses were done to further refine the core region for the enhancer activity and identified a 641 bp region containing the +19 site (Gottgens et al., 2002). This +19 core enhancer contains conserved Myb, Ets and GATA sites and these sites were shown to be bound by GATA2, Fli-1 and Elf-1. The activity of the core +19 enhancer is similar to the +18/+19 enhancer except that it is not sufficient to drive expression in definitive erythroid cells, suggesting addition elements are required for full function (Silberstein et al., 2005). Despite its proven enhancer activity, it was later shown that this stem cell enhancer is dispensable for SCL transcription and haematopoietic cell formation in a mouse knockout study (Gottgens et al., 2004). This suggests that additional regulatory elements are necessary for SCL expression. A -3.8 enhancer was identified subsequently which targets expression in haematopoietic progenitors and endothelium and is bound by Fli-1 and Elf-1 (Gottgens et al., 2004).

A systemic mapping of histone acetylation at the SCL locus identified peaks at the known SCL enhancers and promoters and one additional site 40 kb downstream of exon 1a in mouse (called the +40 region) and at the corresponding conserved +50/+51 region in human (Delabesse et al., 2005). The +40 region was shown to have enhancer activity *in vitro* and target expression in primitive but not definitive erythroid cells *in vivo*. Further analyses of this +40 enhancer indicated that it also targets expression in midbrain but not endothelial cells, and at the same time identified two indispensable GATA/E-box motifs which are bound by SCL and GATA1 in mouse erythroid cells *in vivo* (Ogilvy et al., 2007). The putative +50/+51 enhancer in human was shown to have highly conserved GATA/E-box motifs at +51 and that GATA1, SCL, and LDB1 are bound to this region in a human erythroid cell line (Pawan Dhami, PhD thesis). Transient reporter assays also demonstrated its enhancer activity. Thus, the murine +40 and the human +51 enhancers may function in the auto-regulation of SCL expression in erythroid cells.

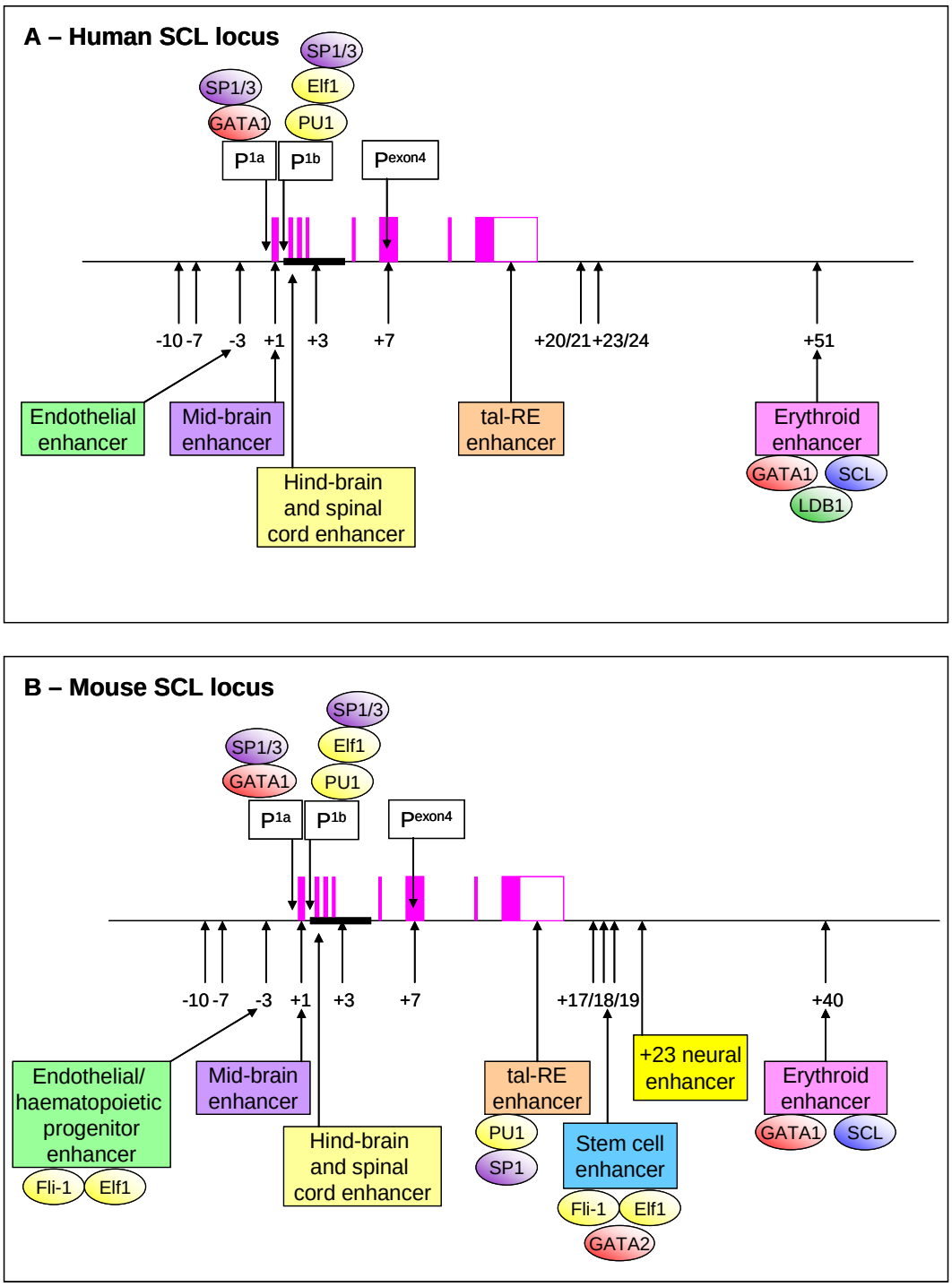

Figure 1.12. Schematic diagrams of regulatory elements at the human and mouse SCL loci. A: human SCL locus; B: mouse SCL locus. The pink boxes show the exons of the SCL gene. +/- numbers refer to the distance in kilobases of each DNase hypersensitive sites (HSs) from promoter 1a. Coloured boxes show the known enhancers while white boxes show the known promoters. Ovals denote proteins which are known to bind to the enhancers/promoters. Detailed description of each component is provided in the text (images shown are not to scale).

F. The SCL protein, interacting partners and downstream targets

The SCL gene encodes a class B basic-helix-loop-helix (bHLH) protein with two isoforms - a 42 kDa full length protein and a 22 kDa amino truncated form (Elwood et al., 1994; Goldfarb et al.,

1992). The bHLH region remains present in both isoforms and is required for nuclear localisation, DNA binding and protein-protein interactions (Hsu et al., 1991; Hsu et al., 1994b).

SCL protein heterodimerises with class A bHLH proteins such as E2A/TCF3, HEB and E2-2 which is a requirement for DNA binding (Hsu et al., 1991; Hsu et al., 1994a). In addition, SCL forms a transactivating protein complex in erythroid cells with other transcription factors including the haematopoietic specific proteins GATA1, LMO2 and ubiquitously expressed proteins E2A/TCF3 (transcript variants E12 and E47) and LDB1 (Wadman et al., 1997) (Figure 1.13). This complex binds to a bipartite DNA motif consisting of an E-box (CANNTG) ~9 bp upstream of a GATA site. The SCL-E2A heterodimer binds to the E-box motif while GATA1 binds to the GATA site. LMO2 and LDB1 do not bind DNA directly - instead, they act as bridging proteins between the SCL-E2A heterodimer and GATA1. More complete descriptions of these binding partners are found in sections 1.4.2.2 – 1.4.2.5.

In addition to these members of the SCL complex, a novel component, ETO2, was found to be recruited to the complex by interacting with E2A (Goardon et al., 2006). ETO2 is a repressor protein and was shown to negatively regulate expression of one target gene GPA (see below). It was demonstrated that changes in the amount of ETO2 protein in this complex governs the expression of erythroid specific genes and is a key determinant in terminal erythroid differentiation.

Since the discovery of SCL and its multiprotein erythroid complex, only a few target genes have thus far been identified in the erythroid lineage (see below). Regulation by SCL in T-cell acute lymphoblastic leukaemia (T-ALL) was studied by comparing the expression of genes in the human genome between SCL-expressing and non-expressing human T-ALL samples using expression arrays and ChIP-on-chip analysis with promoter arrays (Palomero et al., 2006). The results demonstrated that SCL functions as both a repressor and an activator in T-ALL.

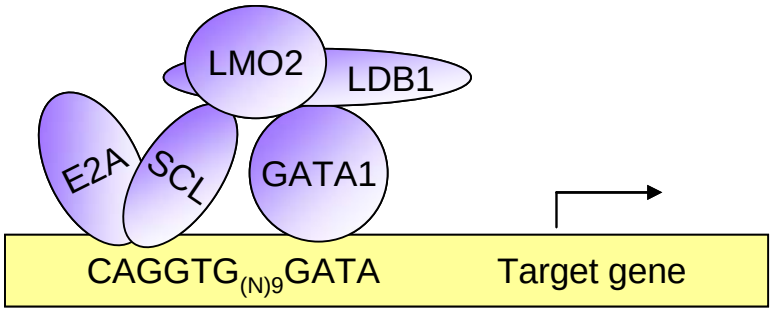

Figure 1.13. The SCL erythroid complex. The ovals or circles indicate the proteins involved in this complex. GAGGTG is the E-box motif; GATA is the GATA site; (N)9 indicates the two motifs are separated by 9 nucleotides. The arrow indicates transcription of target genes. A detailed description of the complex and its target genes are provided in the text.

(i) c-kit.

c-kit was one of the first SCL target genes identified. The c-kit gene encodes a tyrosin receptor kinase which is required for normal haematopoiesis. Expression of c-kit was shown to correlate with SCL expression suggesting a regulatory role of SCL in $CD34^+$ haematopoietic cells (Krosl et al., 1998). Chromatin immunoprecipitation studies revealed that members of the SCL erythroid complex, together with a novel member, specificity protein 1 (Sp1, a zinc-finger protein) occupies the c-kit promoter and the combinatorial interaction of all the members of this complex is essential for the synergistic transactivation of c-kit (Lecuyer et al., 2002). GATA2, another member of the GATA family, was also found in the complex and was shown to convey greater transcriptional activation on the c-kit promoter than GATA1. In a separate study, the pentamer protein complex consisting of SCL, E12, LMO2, LDB1 and the retinoblastoma protein (pRb) was shown to inhibit c-kit expression during erythropoiesis (Vitelli et al., 2000).

(ii) Glycophorin A (GPA).

The erythroid cell-specific glycophorin A gene (GPA) was identified as a target of the SCL erythroid complex (including Sp1) in primary hematopoietic cells (Lahlil et al., 2004). The complex was shown to occupy the GPA promoter *in vivo* and to activate GPA expression with GATA1, rather than GATA2, conveying a greater degree of transcriptional activation.

(iii) α- and β-globin genes.

The SCL erythroid complex was also found to occupy the human β-globin locus control region (LCR) during erythroid differentiation (Song et al., 2007). The long range interaction between the β-globin LCR and the active β-globin promoter requires LDB1 for the formation of the loop structure. The mouse and the human α-globin loci were found to be co-occupied by SCL, E2A, GATA1/2, LMO2 and LDB1 in DNase I hypersensitivity assays and ChIP-on-chip (Anguita et al., 2004; De Gobbi et al., 2007). However, no functional analysis of the α-globin clusters has been performed to investigate the role of the complex in globin regulation.

(iv) Protein 4.2 (P4.2).

The gene for protein 4.2 (P4.2), an important component of the erythrocyte cell membrane skeleton, is also a target gene of the SCL erythroid complex in mouse. SCL, E47, GATA1, LMO2 and LDB1 were demonstrated to activate P4.2 expression via two GATA E-box elements in the P4.2 promoter in erythroid cells (Xu et al., 2003). Maximal transcription requires both GATA and E-box sites and all five members of the complex. The SWI/SNF protein Brg1 was also found to associate with the complex and down-regulate P4.2 expression by recruiting chromatin-remodelling complexes and histone modification enzymes (Xu et al., 2006).

(v) GATA1.

The *cis*-acting regulatory element (HS-1) upstream of the promoter of GATA1 in mouse contains a composite E-box GATA site and was shown to be bound by the SCL eyrthroid complex (Vyas et al., 1999b). Mutations in the GATA1 site, but not the E-box site, significantly abolish the activation activity of the element.

(vi) FLK-1.

The tyrosine receptor kinase FLK-1 is important for the generation of common precursors for both the endothelial and haematopoietic lineages. The FLK intronic enhancer contains two E-box motifs, one indispensable GATA site and two ETS binding sites. These sites have been demonstrated to be bound by SCL, GATA1 and ETS proteins respectively. Mutations on these sites abolished the enhancer activity. Combinatorial action of these transcription factors regulates FLK-1 expression in both haematopoietic and vascular development (Kappel et al., 2000).

(vii) RUNX1/AML1.

The transcription factor RUNX1/AML1 is an important regulator of haematopoiesis and has recently been shown to be regulated by a multiprotein complex containing SCL in mouse. SCL, together with LMO2, LDB1, GATA2 and ETS were found to bind to the putative +23 enhancer of RUNX1 located 23 kb downstream of the transcription start site of RUNX1 *in vivo* in a myeloid progenitor cell line. This +23 enhancer contributes to expression of RUNX1 in early haematopoiesis (Nottingham et al., 2007). Direct binding of SCL, LMO2 and GATA2 was confirmed by chromatin immunoprecipitation in another study and gene expression profiling also revealed that RUNX1, together with RUNX3, are downstream targets of SCL in early haematopoietic development (Landry et al., 2008).

1.4.2.2 GATA1

A. <u>The GATA1 gene and its expression</u>

GATA1, also named NF-E1, NF-1, Ery-1 and GF-1, was first identified as a protein bound to the 3' enhancer of the β-globin gene (Wall et al., 1988). It was later mapped to human chromosome X at Xp21-22 (Zon et al., 1990). The GATA1 gene locus contains six exons where the first exon is non-coding. GATA1 belongs to the GATA family of genes including GATA1 to 6 where GATA1, 2 and 3 are important in haematopoietic development.

GATA1 is widely expressed in various lineages of haematopoietic development. In many respects, its expression patterns mirror those of SCL (see section 1.4.2.1). It is expressed in primitive and definitive erythroid cells (Tsai et al., 1989; Zon et al., 1990), megakaryocytes (Martin et al., 1990), eosinophils (Zon et al., 1993) and mast cells (Martin et al., 1990). It is also expressed in testis

Sertoli cells in mice (Yomogida et al., 1994). At earlier stages of haematopoietic development, GATA1 is expressed in HSCs, common myeloid progenitors (CMP), megakaryocyte/erythrocyte lineage restricted progenitors (MEP) and haemangioblast (Akashi et al., 2000; Kuhl et al., 2005; Yokomizo et al., 2007). During erythroid differentiation from HSCs, GATA1 was found to be expressed at low levels initially while its expression gradually increases as erythroid differentiation progresses.

B. Functions of GATA1

- **Erythropoiesis**

Several lines of evidence have demonstrated a crucial participation of GATA1 in erythroid development. Deletion of GATA1 in mouse ES cells resulted in contribution to all haematopoietic tissues except mature red blood cells in chimeric mice (Pevny et al., 1991). *In vitro* colony assays further suggested that the GATA1 null cells failed to mature beyond proerythroblasts, a cell type found at an early stage of terminal differentiation (Pevny et al., 1995). Similarly, GATA1 null chimeric mice died between E10.5 to E11.5 of anaemia and displayed embryonic erythroid cells arrested at the proerythroblast stage (Fujiwara et al., 1996). This further established the importance of GATA1 in both primitive and definitive erythropoiesis. *In vitro* differentiation of GATA1 null ES cells confirmed that the proerythroblast arrest and death by apoptosis and thus suggesting GATA1 supports the viability of red blood cell precursors by suppressing apoptosis (Weiss et al., 1994; Weiss and Orkin, 1995). Inducible rescue of GATA1 null erythroblasts demonstrated that GATA1 promotes terminal erythroid maturation and G1 cell cycle arrest by suppressing the expression of c-MYC, a proto-oncogene which regulate cell proliferation and differentiation (Rylski et al., 2003). The interaction between GATA1 and its co-factor FOG1 has been shown to be required for terminal erythroid maturation (Rylski et al., 2003).

- **Megakaryopoiesis**

GATA1 also plays a critical role in megakaryocytic development. GATA1-deficient mice were shown to have reduced platelet counts as well as expansion of immature megakaryocytes (Shivdasani et al., 1997). These megakaryocytes have abnormal morphology, are unable to mature and exhibit a marked hyperproliferation *in vivo* and *in vitro* (Vyas et al., 1999a). At the molecular level, GATA1 activates transcription of megakaryocyte specific genes including NF-E2, GP1bα and platelet factor 4.

- **Eosinophils and mast cells development**

It was first shown that GATA1 could convert chicken myeloblasts, mouse common lymphoid progenitors and human myeloid progenitors to eosinophils (Hirasawa et al., 2002; Iwasaki et al.,

2003; Kulessa et al., 1995). Disruption of GATA sites in the GATA1 promoter resulted in selective loss of the eosinophil lineage (Yu et al., 2002a). Mast cell development was also shown to be disrupted in GATA1^{low} mice where the first enhancer and distal promoter of GATA1 are deleted (Migliaccio et al., 2003); the mast cells produced were defective in terminal maturation and had increased apoptosis. At the molecular level, GATA1 has been shown to activate expression of eosinophil specific genes such as MBP (Yamaguchi et al., 1998; Yamaguchi et al., 1999).

- **GATA1 and leukaemia**

One of the more well-studied disorders associated with mutation in GATA1 is the transient myeloproliferative disorder (TMD) which occurs in about 10% of children with Down syndrome. In 20% of the TMD cases, patients develop Down syndrome-related acute megakaryocytic leukaemia (DS-AMKL) later in life. In most cases, the mutations in GATA1 introduce a premature stop codon or a splice site in the N-terminal activation domain which results in the translation of a GATA1s isoform lacking the N-terminal activation domain (Wechsler et al., 2002). GATA1s has diminished transactivation potential in *in vitro* assays and causes a reduction in differentiation of megakaryoctic precursor cells.

A number of missense mutations in the N-finger of GATA1 have also been found in patients with X-linked thrombocytopenia and anaemia (Table 1.13). In most of these cases, the ability of GATA1 to interact with FOG1 or to bind DNA is affected. The severity of the disease depends on the particular type of mutation.

Mutation	FOG1-binding	DNA-binding	Phenotype	References
V205M	Strongly reduced	Not affected	Macrothrombocytopenia; Dyserythropoietic anaemia	(Nichols et al., 2000)
D218G	Reduced	Not affected	Macrothrombocytopenia; Dyserythropoiesis without anaemia	(Freson et al., 2001)
D218Y	Strongly reduced	Not affected	Macrothrombocytopenia; anaemia	(Freson et al., 2002)
G208S	Reduced	Not affected	Macrothrombocytopenia	(Mehaffey et al., 2001)
R216Q	Not affected	Reduced binding to complex and palindromic sites	Macrothrombocytopenia; β-thalassemia	(Balduini et al., 2004; Yu et al., 2002b)

Table 1.13. Mutations of GATA1 in X-linked thrombocytopenia and anaemia.

C. Transcriptional regulation of GATA1

- **Cis-regulatory elements**

Together with its upstream region, the first untranslated exon of GATA1 contains regulatory elements for GATA1 expression which are conserved across vertebrates (Figure 1.14). The erythroid-specific promoter region located upstream of the erythroid first exon (IE) contains a CACCC box and a double GATA site necessary and sufficient to drive expression in erythroid cells

(Zon et al., 1990). The CACCC box is essential for initiation of GATA1 gene expression as mutations or deletions therein completely disrupted promoter activity in zebrafish (Meng et al., 1999). The GATA sites, in contrast, are not essential for lineage-specific expression (Nicolis et al., 1991; Yu et al., 2002a). In mouse, there is an additional promoter upstream of the first exon testis-specific exon (IT) which is used in Sertoli cells (Onodera et al., 1997b).

Upstream regions of the GATA1 IE promoter there is an enhancer element for GATA1 expression in primitive and definitive erythropoiesis, as well as in megakaryocytes and eosinophils (Onodera et al., 1997a). This region is denoted as the enhancer G1HE. The core regions of G1HE which contains a GATA site or a GATA site plus a CACCC box are required for expression in erythroid cells and megakaryocytes respectively (Nishimura et al., 2000; Vyas et al., 1999b). An additional intronic enhancer intron-SP located in the first intron which contains GATA and AP1 repeats is required for efficient expression in definitive erythroid cells (Onodera et al., 1997a). The testis-specific enhancer in mouse G1TAR is required for activation of the IT promoter (Wakabayashi et al., 2003).

- **Trans-acting proteins**

The presence of a number of GATA sites in the regulatory region of GATA1 suggests that GATA factors may bind to these regions. Indeed, GATA1 has been shown to bind to the G1HE, the double GATA site in promoter IE and the intronic enhancer intron-SP *in vivo* by chromatin-immunprecipitation assays (Valverde-Garduno et al., 2004). Overexpression of GATA1 upregulates a transgenic GATA1 reporter gene in zebrafish and the self-association of GATA1 is required for this regulation (Kobayashi et al., 2001; Nishikawa et al., 2003; Shimizu et al., 2007). Furthermore, suppression of GATA2 gene expression down-regulates GATA1 expression in blast cells but has no effect in differentiated cells (Lugus et al., 2007; Tsai and Orkin, 1997). All these findings suggest that GATA1 is under the control of GATA2 at an early stage of development, while GATA1 is involved in autoregulation later in development.

Other transcription factors also play important roles in GATA1 transcriptional regulation. PU.1 antagonises GATA1 expression by hindering the binding of GATA1 to the GATA1 locus and thus inhibits autoregulation (Zhang et al., 2000). It is also suggested that PU.1 inhibits GATA1 expression by creating a repressive chromatin structure (Stopka et al., 2005). The SCL erythroid complex containing SCL, LMO2, LDB1, E2A and GATA1 (Section 1.4.2.1 F) is recruited to the G1HE in erythroid cells *in vivo* (Valverde-Garduno et al., 2004). CP2 has been shown to bind to the upstream region of erythroid specific first exon at two CP2-binding sites adjacent to the double GATA site bound by GATA1. Mutation in these CP2 sites impair promoter activity in erythroid

cells (Bose et al., 2006). This suggests a functional cooperation of the two factors in controlling expression of GATA1.

The Sp1/Krüppel-like factor (KLF) family binds to the CACCC box. Expression of KLF2 and KLF6 has been shown to correlate with expression of GATA1 in mice suggesting a regulatory role played by these factors (Basu et al., 2005; Matsumoto et al., 2006).

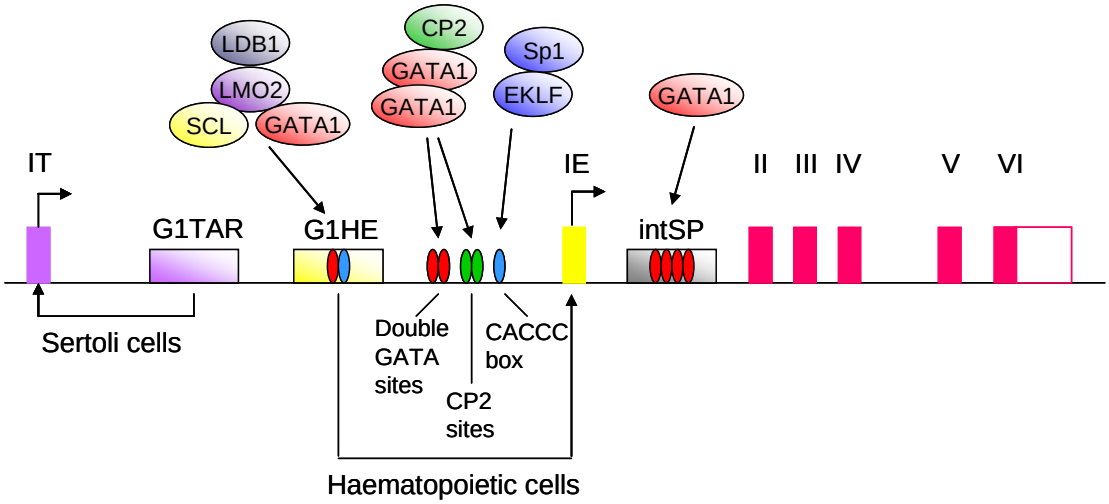

Figure 1.14. A schematic diagram of the regulatory regions of GATA1. The pink boxes show the exons of the GATA1 gene. The GATA1 testis activation region (G1TAR) activates transcription from the testis-specific first exon (IT). Both G1TAR and IT are only found in mouse. The GATA1 haematopoietic enhancer (G1HE) activates transcription from the haematopoietic-specific first exon (IE). The intSP is an erythroid-specific enhancer found in the first intron. Small ovals indicate specific motifs in the regulatory elements: red: GATA sites; blue: CACCC box; green: CP2 sites. Large ovals denote proteins which are known to bind to the enhancers/promoters. Detailed description of each component is provided in the text (image shown is not to scale).

D. The GATA1 protein and interacting partners

GATA1 encodes a protein which belongs to the GATA family of transcription factors (GATAs 1 through 6) and contains three functional domains: the N- and C-terminal zinc finger motifs and the N-terminal activation domain. The C-terminal zinc finger binds to the DNA consensus sequence (A/T)GATA(A/G) whereas the N-terminal zinc finger functions by binding to DNA and recruiting co-factors and contributes to the stability and specificity of DNA-binding. The N-terminal activation domain confers transcriptional activation to target genes (Martin and Orkin, 1990).

GATA1 has been shown to physically interact with a variety of nuclear proteins, as well as to self-dimerise. Such interactions are essential for the function of GATA1 as a transcriptional regulator and are pivotal in haematopoietic development. Table 1.14 summarises the co-factors or transcription factors which interact with GATA1.

Co-factor	Protein functions	Mechanism of action	Reference
GATA1	Zinc-finger protein required for erythroid and megakaryocytic development	Homodimerisation requires N-finger-C-finger contacts and induce transcriptional activation.	(Mackay et al., 1998)
Friend of GATA1 (FOG1)	Zinc-finger protein required for erythroid and megakaryocytic development	Interacts with N-finger of GATA1; co-operates with GATA1 in erythroid and megakaryocytic differentiation; both synergises with or represses GATA1-mediated activation.	(Fox et al., 1999; Tsang et al., 1997)
LMO2	LIM-domain protein implicated in T-ALL	Interacts with GATA1 as a requirement for the formation of the SCL erythroid complex together with LDB1 and E2A; whole complex trans-activates a number of erythroid-specific genes,	(Wadman et al., 1997) Section 1.4.2.1E
p300/CBP	Histone acetyltransferase	Interacts with GATA1 *in vivo* and *in vitro*; stimulates GATA1 activity by acetylating GATA1.	(Boyes et al., 1998)
Erythroid Krüppel-like factor (EKLF)	Erythroid-specific Zinc finger protein of the Krüppel-like factor family	Interacts with C-finger of GATA1; functions as a co-regulator with GATA1.	(Gregory et al., 1996; Merika and Orkin, 1995)
Sp1	Krüppel-like factor family protein required for early embryonic development	Interacts with C-finger of GATA1; activates GATA1 transcriptional activity.	(Gregory et al., 1996; Merika and Orkin, 1995)
PU.1	Ets protein required for lymphoid and granulocytic development	Interacts with GATA1 via the DNA-binding domain; antagonises GATA1 transcriptional activation by preventing its binding to DNA.	(Rekhtman et al., 1999; Zhang et al., 2000)
CP2	Transcription factor which stimulates α-globin expression	Interact with C-finger of GATA; functions as a co-regulator with GATA1 in regulation of erythroid genes	(Bose et al., 2006)
c-myb	Oncogene required for haematoppoiesis	Interacts with GATA1 and inhibits its DNA-binding activity.	(Matsumura et al., 2000)
Fli-1	Ets protein important for megakaryocyte differentiation	Interacts with zinc fingers of GATA1 via its own Ets domain; functions as a co-activator for genes involved in terminal megakaryocytic differentiation.	(Eisbacher et al., 2003)

Table 1.14. Interacting partners of GATA1.

E. Downstream targets of GATA1.

A large set of genes, especially those related to haematopoiesis, have been characterised as target genes of GATA1 and some of the more well-characterised target genes are discussed below. Genes whose regulation is mediated through the SCL erythroid complex containing GATA1 are not discussed here (refer to section 1.4.2.1 F).

(i) α- and β-globin genes.

GATA1 was first identified by its binding to an enhancer at the β-globin locus (Wall et al., 1988). ChIP-on-chip analyses of the β-globin locus demonstrated that GATA1 binds to a region of the HS2 core element and an additional region upstream of γ-globin gene (Horak et al., 2002). Subsequently,

GATA1 was shown to bind to the α-globin locus (Evans et al., 1988). It has been shown that GATA2 drives α-globin expression in multipotent progenitors while GATA1 replaces GATA2 in committed erythroid progenitors where it is bound to the α-globin promoter (Anguita et al., 2004).

(ii) EKLF.

In addition to being an interacting partner of GATA1, EKLF was also identified as a target gene of GATA1. Forced expression of GATA1 in non-erythroid cells induced activation of the EKLF promoter while one of the GATA sites in the promoter of EKLF was found to be indispensable for promoter function (Crossley et al., 1994). GATA1 was shown to bind to two GATA sites in a GATA-E-box-GATA motif in the promoter of EKLF which is essential for EKLF expression (Anderson et al., 1998, 2000). Functional interaction of CP2 and GATA1 may contribute to the regulation at the EKLF promoter (Bose et al., 2006).

(iii) GATA2.

GATA1 has been shown to repress expression of GATA2 in erythroid differentiation (Weiss et al., 1994). GATA1 binds to a region upstream of promoter 1G of GATA2 which is normally bound by GATA2 itself (Grass et al., 2003). GATA2, when bound to its own promoter, recruits CBP leading to histone acetylation and transcriptional activation. Displacement of GATA2 by GATA1 disrupts this autoregulation and thus represses GATA2 expression.

(iv) Epo and EpoR.

Erythropoietin (Epo) is a major growth factor for erythroid cells which binds to the Epo receptor (EpoR), a cell surface marker, resulting in proliferation and differentiation of erythroid progenitors. GATA1 was found to bind and transactivate the EpoR promoter (Zon et al., 1991). Conversely, GATA1 acts as a repressor for Epo expression and binds to a GATA site in the Epo promoter (Imagawa et al., 2002; Imagawa et al., 1997).

(v) NF-E2.

Abrogation of GATA1 expression was shown to significantly reduce expression of NF-E2 in megakaryocytes (Vyas et al., 1999a). GATA1-mediated activation acts in concert with human FOG2 (Holmes et al., 1999). Further analyses of the NF-E2 promoter in mouse demonstrated that GATA1 bind to the proximal promoter 1B located in the first intron (Moroni et al., 2000).

(vi) GFI-1B.

Gfi-1B is an erythroid-specific transcription factor which plays an essential role in erythropoiesis. ChIP assays demonstrated that GATA1 binds to the promoter region of GFI-1B (Huang et al., 2004). Ectopic expression of GATA1 in non-erythroid cells activates the GFI-1B promoter. This

direct activation is also dependent on NF-Y which also binds to GFI-1B promoter. GFI-1B itself suppresses the GATA1-mediated activation by protein-protein interaction (Huang et al., 2005).

(vii) FOG1.

FOG1 is a co-factor and binding partner of GATA 1 (see Table 1.14). Global expression analysis revealed that FOG1 expression is rapidly induced by GATA1 expression and ChIP studies confirmed the binding of GATA1 to GATA motifs in the *cis*-regulatory elements of FOG1 (Welch et al., 2004). This suggests a regulatory hierarchy where GATA1 first induces expression of its co-factor for a co-operative activation of the β-globin gene (Welch et al., 2004).

(viii) c-MYC.

The proto-oncogene c-MYC is a transcription factor which binds to E-box motifs and recruits histone acetyltransferases. GATA1 has been shown to repress c-MYC expression and binds to its promoter in mouse erythroid cells (Rylski et al., 2003).

1.4.2.3 E2A/TCF3

A. <u>The E2A/TCF3 gene and gene products</u>

The E2A gene, also named TCF3, was first identified as two highly similar cDNA clones whose dimerised products bind specifically to the human immunoglobulin kappa chain enhancer (Murre et al., 1989). The gene was mapped on chromosome 19p13.3 and contains 19 exons. E12 and E47 are two splicing variants produced by alternative splicing of exons 17 and 18.

E12 and E47 are the founding members of basic helix-loop-helix (bHLH) family of transcription factors. They belong to class A of the bHLH proteins including HEB and E2-2 which bind to DNA elements with the consensus E-box sequence CANNTG. Both E12 and E47 are virtually identical except that the C-terminal bHLH domains are slightly different (due to the alternative splicing of exons 17 and 18). There are two activation domains, AD1 and AD2, located at the N-terminus and in the central region of the protein, which mediate transcriptional activation by recruiting histone acetyltransferases (Massari et al., 1999; Qiu et al., 1998). E12 and E47 form homodimers or heterodimers with class B bHLH protein such as MyoD where the protein interaction is mediated by the bHLH domain. E47 homodimers, and heterodimers between MyoD and E47 or E12 can bind DNA; whereas E12 homodimers fail to bind DNA due to the presence of an inhibitory domain in the basic region of E12 (Sun and Baltimore, 1991).

Like other class A bHLH proteins, E2A proteins are ubiquitously expressed in a variety of cell types and tissues. However, expression of E2A has been shown to be up-regulated during B-cell lineage commitment (Zhuang et al., 2004).

B. Functions of E2A

- **Regulation of tissue-specific differentiation**

Despite its ubiquitous expression pattern, E2A can still function as a co-regulator in tissue-specific differentiation. This is mediated by the formation of heterodimers between E2A and class B tissue-specific bHLH proteins. One of the more well-studied examples is the regulation of myogenesis. E2A dimerises with MyoD and regulates expression of several downstream muscle-specific regulators which, in turn, control muscle differentiation (Lassar et al., 1991). Another example is the dimerisation of E2A and SCL which leads to the formation of multiprotein complexes and direct transcriptional activation or repression of erythroid-specific genes (Section 1.4.2.1 F).

- **Transcriptional activation of B-cell specific genes**

E2A is a key transcription factor regulating transcription of B-cell specific genes. Early B-cell factor (EBF), an important regulator of B-cell commitment and lineage-specific gene expression, is one of the more well-characterised targets of E2A. Ectopic expression of E12 induced expression of EBF and the promoter of EBF functionally interacts with E47 (Kee and Murre, 1998; Smith et al., 2002). However, E2A itself is not sufficient to drive EBF expression, as PU.1 has been shown to work independently or in a cooperative manner with E2A to direct EBF expression (Medina et al., 2004).

E2A and EBF are involved in the regulation of an overlapping set of B-lineage specific genes including genes crucial for gene rearrangement and BCR expression (Mansson et al., 2004). However, some of these genes may be secondary targets of E2A mediated by EBF activation, although a subset have been shown to have direct association between E2A and their regulatory region in ChIP (Greenbaum and Zhuang, 2002).

- **Regulation of lymphoid development**

Homozygous E2A mutant mice or knockout mice contained no mature B cells while all other haematopoietic lineages were intact (Bain et al., 1994; Zhuang et al., 1994). Detailed examination of the defect in B-cell differentiation revealed that B-cell development was blocked at the stage before IgH DJ rearrangement and before the Pro-B cell formation. More recent studies in E2A knockout haematopoietic progenitor cells show a characteristic pro-B cell signature indicative that these cells are pluripotent (Ikawa et al., 2004); they expressed genes specific to other lineage but not the B-cell lineage, and they could contribute to all blood lineage except B-cells. Taken together, these data indicate that E2A is required for B-lineage restriction and commitment.

E2A, together with other class A bHLH proteins like HEB, are also involved in T-cell development and lineage commitment which requires the formation of heterodimers between E2A and HEB

(Barndt et al., 2000). In contrast to the B- and T- lineages, suppression of E2A function has been implicated in natural killer (NK)-cell development. This is mediated by dimerisation of E2A with the Id protein, thus inhibiting E2A from binding to DNA (Heemskerk et al., 1997).

- **Translocation and leukaemia**

Chromosomal rearrangements involving the E2A gene result in acute lymphoblastic leukemia (ALL). Translocation between chromosome 1 and 19 resulted in the formation of a fusion E2A-PBX1 protein while translocation between chromosome 17 and 19 leads to expression of E2A-HLF (Inaba et al., 1992; Nourse et al., 1990). Such translocations disrupt the normal gene regulatory networks of the proteins involved and the fusion proteins may also cause abnormal transcriptional upregulation of its target genes.

1.4.2.4 LMO2

A. <u>The LMO2 gene and gene product</u>

LMO2 belongs to the LIM-only family of genes which was first discovered by virtue of its translocation product in T-ALL (Boehm et al., 1991). It is located on chromosome 11p13 and contains three exons. It encodes a protein comprising two tandem LIM domains which are zinc-binding finger-like domains structurally similar to the DNA-binding GATA fingers. However, unlike the GATA fingers, LIM domains have not been shown to bind DNA but are restricted to protein-protein interaction. LMO2 is ubiquitously expressed in blood progenitor cells and endothelial cells (Delassus et al., 1999; Yamada et al., 2000).

B. <u>Functions of LMO2</u>

- **Regulation of haematopoietic and endothelial development**

LMO2 plays a critical role in erythropoietic and endothelial development. LMO2 null mice display defects in primitive erythropoiesis and lethality around E10.5 (Warren et al., 1994). Additional studies of homozygous mutant $LMO2^{-/-}$ mouse ES cells showed that all haematopoietic lineages are disrupted in these $LMO2^{-/-}$ cells while ectopic expression of LMO2 rescues this phenotype (Yamada et al., 1998). Thus, LMO2 is necessary for all stages of definitive haematopoiesis and it functions at least at the level of pluripotent stem cell. These studies, when considered together, show that LMO2 has a very similar function as SCL in early haematopoietic development. This suggests that protein-protein interaction between LMO2 and SCL and possibly other transcription factors governs haematopoiesis (Wadman et al., 1997). A further LMO2 null study has also demonstrated that LMO2 is not required for the generation of the primary capillary network (vasculogenesis) but it is

crucial for remodelling of this capillary network into the mature vascular system (angiogenesis) (Yamada et al., 2000).

- **Translocation and leukaemia**

As mentioned above, LMO2 is involved in chromosome translocation between chromosome 11 and 14 or 11 and 7 which causes T-ALL. The translocation breakpoint is upstream of the LMO2 promoter and thus enforced expression of the full-length LMO2 protein is observed in T-cells. Ectopic expression of LMO2 in T-cells from transgenic mice resulted in a marked accumulation of immature T-cells indicating that LMO2 induces cell proliferation and blocks T-cell differentiation (Neale et al., 1995). Enforced expression of both SCL and LMO2 further enhanced this effect leading to the hypothesis that interaction between these two proteins can alter T cell development and potentiate tumorigenesis (Larson et al., 1996).

There are two models describing the mechanism of tumorigenesis by LMO2 translocations. In the first model, an LMO2 complex was described which is similar to its analog in erythroid cells (Grutz et al., 1998). This complex involves two E2A/SCL heterodimers which bind to two E-box motifs separated by one helix turn. LMO2 and LDB2 proteins are bridging protein for this multimer complex. This complex may regulate a specific subset of target genes involved in tumor development. The second model suggests that the abnormal expression of LMO2 may sequester members of protein complexes or dimmers, thus disrupting their normal functions.

1.4.2.5 LDB1

A. The LDB1 gene and gene product

The LIM-domain binding protein 1 (LDB1), also named CLIM2 or NL1, was first discovered due to its ability to interact with LIM domain proteins LIM-homeo-domain (LIM-HD) and LIM-only (LMO) (Agulnick et al., 1996). It was mapped on chromosome 10q24-25 and contains 11 exons. LDB1 is ubiquitously expressed in various tissues. The LDB1 protein contains three domains: a conserved nuclear localisation sequence (NLS), an N-terminal dimerisation domain (DD) and a C-terminal LIM interaction domain (LID). The DD mediates homodimerisation of LDB1, while LID mediates interaction with LIM-HD or LMO proteins. However, no DNA-binding or enzymatic activity has been observed in LDB1. Thus, it is likely that LDB1 functions exclusively through protein-protein interaction.

B. Functions of LDB1

- **Developmental regulator**

Deletion of LDB1 in mice induced severe defects in anterior-posterior patterning, truncation of head structures, posterior axis duplication and a lack of heart embryonic cells (Mukhopadhyay et al., 2003). This suggests that LDB1 plays important roles in diverse developmental process. The exact mechanism underlying the regulation of developmental processes regulated by LDB1 is not fully understood. However, the precise stoichiometry of LDB1 and its interacting partners LIM-HD or LMO proteins may be a critical criterion in determining the downstream pathways regulating various biological processes. One piece of evidence is that overexpression of LDB1 inhibits erythroid differentiation (Visvader et al., 1997).

- **Transcriptional modulator**

Although LDB1 does not bind DNA directly, its interaction with LIM domain proteins contributes significantly to its role as a transcriptional regulator. In additional to the target genes in the SCL erythroid complex containing LDB1 (described in section 1.4.2.1 F), LDB1 has also been found to repress the synergistic activation of the insulin enhancer by LMX1 and E47 (Jurata and Gill, 1997).

1.4.2.6 Transcriptional regulatory networks in haematopoietic development

Haematopoietic differentiation and lineage-specific commitment is a complex process regulated by multiple transcription factors or developmental critical genes. The 5 genes described above (SCL, GATA1, E2A, LMO2 and LDB1) belong to only a small subset of these regulators. These regulators, of which many are transcription factors and co-factors, both physically interact and/or are transcriptionally regulated by one another. This results in the generation of a global regulatory network (Section 1.2).

Swiers et al. 2006 first attempted to build a network for erythropoiesis based on data in mouse (Swiers et al., 2006). The authors identified the network motifs first described in *E. coli* and yeast which play essential roles at different stages of haematopoietic development. For instance, multi-input motifs such as SCL and Hex which are both regulated by the cooperation of GATA2, Fli-1 and Elf-1 are important for co-ordinating gene expression in specific lineage. Feed-forward motifs such as GATA1 → FOG1 → β-globin control the temporal expression of lineage-specific genes and prevent immature activation of certain genes. Autoregulation, such as in the case of GATA1, generates a forward momentum and stabilises expression of GATA1 in specific cell types. The authors concluded that, in summary, network motifs function together to provide a complex regulation of haematopoiesis.

1.5 Aims of this thesis

Although a lot is known about the functional roles of SCL and its binding partners during haematopoietic development, the molecular means by which these functions are regulated are not well understood. For example, only a handful of target genes have been identified in human or mouse for the SCL erythroid complex. The central aim of this thesis was to further characterise the regulatory network governed by this complex in human erythroid cells using a combination of approaches. Specifically, the aims of this work were:

1. To develop perturbation assays for SCL, GATA1, E2A, LMO2 and LDB1 by siRNA-mediated knockdown in the human erythroid cell line K562.
2. To study the downstream effects of each of the knockdowns and identify putative primary or secondary target genes by genome-wide expression analyses using Affymetrix GeneChips.
3. To identify and confirm direct target genes of the 5 transcription factors using ChIP coupled with a human transcription factor promoter array (ChIP-chip).
4. To generate a transcription network of the SCL erythroid complex using an integration of various experimental approaches described in aims 1-3.

(2004a). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636-640.
(2004b). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.
(2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.
Agulnick, A.D., Taira, M., Breen, J.J., Tanaka, T., Dawid, I.B., and Westphal, H. (1996). Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. Nature *384*, 270-272.
Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193-197.
Akiyama, Y. (1998). TFSEARCH: Searching Transcription Factor Binding Sites. http://wwwrwcporjp/papia/.
Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation And Methylation Of Histones And Their Possible Role In The Regulation Of Rna Synthesis. Proc Natl Acad Sci U S A *51*, 786-794.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.
Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci U S A *74*, 5350-5354.
Anderson, K.P., Crable, S.C., and Lingrel, J.B. (1998). Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. J Biol Chem *273*, 14347-14354.
Anderson, K.P., Crable, S.C., and Lingrel, J.B. (2000). The GATA-E box-GATA motif in the EKLF promoter is required for in vivo expression. Blood *95*, 1652-1655.
Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. Embo J *23*, 2841-2852.
Antes, T.J., Namciu, S.J., Fournier, R.E., and Levy-Wilson, B. (2001). The 5' boundary of the human apolipoprotein B chromatin domain in intestinal cells. Biochemistry *40*, 6731-6742.
Aplan, P.D., Begley, C.G., Bertness, V., Nussmeier, M., Ezquerra, A., Coligan, J., and Kirsch, I.R. (1990a). The SCL gene is formed from a transcriptionally complex locus. Mol Cell Biol *10*, 6426-6435.
Aplan, P.D., Lombardi, D.P., Ginsberg, A.M., Cossman, J., Bertness, V.L., and Kirsch, I.R. (1990b). Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. Science *250*, 1426-1429.

Aza-Blanc, P., Cooper, C.L., Wagner, K., Batalov, S., Deveraux, Q.L., and Cooke, M.P. (2003). Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. Mol Cell *12*, 627-637.

Baeuerle, P.A., and Baltimore, D. (1988). I kappa B: a specific inhibitor of the NF-kappa B transcription factor. Science *242*, 540-546.

Bailey, T.L., and Elkan, C. (1995). The value of prior knowledge in discovering motifs with MEME. Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB *3*, 21-29.

Bain, G., Maandag, E.C., Izon, D.J., Amsen, D., Kruisbeek, A.M., Weintraub, B.C., Krop, I., Schlissel, M.S., Feeney, A.J., van Roon, M.*, et al.* (1994). E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell *79*, 885-892.

Bajic, V.B., Tan, S.L., Suzuki, Y., and Sugano, S. (2004). Promoter prediction analysis on the whole human genome. Nature biotechnology *22*, 1467-1473.

Balduini, C.L., Pecci, A., Loffredo, G., Izzo, P., Noris, P., Grosso, M., Bergamaschi, G., Rosti, V., Magrini, U., Ceresa, I.F.*, et al.* (2004). Effects of the R216Q mutation of GATA-1 on erythropoiesis and megakaryocytopoiesis. Thrombosis and haemostasis *91*, 129-140.

Balhoff, J.P., and Wray, G.A. (2005). Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. Proceedings of the National Academy of Sciences of the United States of America *102*, 8591-8596.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell *33*, 729-740.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299-308.

Barndt, R.J., Dai, M., and Zhuang, Y. (2000). Functions of E2A-HEB heterodimers in T-cell development revealed by a dominant negative mutation of HEB. Mol Cell Biol *20*, 6677-6685.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Barski, A., and Frenkel, B. (2004). ChIP Display: novel method for identification of genomic targets of transcription factors. Nucleic Acids Res *32*, e104.

Barton, G.M., and Medzhitov, R. (2002). Retroviral delivery of small interfering RNA into primary cells. Proceedings of the National Academy of Sciences of the United States of America *99*, 14943-14945.

Basu, P., Morris, P.E., Haar, J.L., Wani, M.A., Lingrel, J.B., Gaensler, K.M., and Lloyd, J.A. (2005). KLF2 is essential for primitive erythropoiesis and regulates the human and murine embryonic beta-like globin genes in vivo. Blood *106*, 2566-2571.

Begley, C.G., Aplan, P.D., Davey, M.P., Nakahara, K., Tchorz, K., Kurtzberg, J., Hershfield, M.S., Haynes, B.F., Cohen, D.I., Waldmann, T.A., *et al.* (1989a). Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proc Natl Acad Sci U S A *86*, 2031-2035.

Begley, C.G., Aplan, P.D., Denning, S.M., Haynes, B.F., Waldmann, T.A., and Kirsch, I.R. (1989b). 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.

Begley, C.G., Visvader, J., Green, A.R., Aplan, P.D., Metcalf, D., Kirsch, I.R., and Gough, N.M. (1991). Molecular cloning and chromosomal localization of the murine homolog of the human helix-loop-helix gene SCL. Proc Natl Acad Sci U S A *88*, 869-873.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Bennett, S. (2004). Solexa Ltd. Pharmacogenomics *5*, 433-438.

Bernard, O., Azogui, O., Lecointe, N., Mugneret, F., Berger, R., Larsen, C.J., and Mathieu-Mahul, D. (1992). A third tal-1 promoter is specifically used in human T cell leukemias. J Exp Med *176*, 919-925.

Bernard, O., Lecointe, N., Jonveaux, P., Souyri, M., Mauchauffe, M., Berger, R., Larsen, C.J., and Mathieu-Mahul, D. (1991). Two site-specific deletions and t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5' part of the tal-1 gene. Oncogene *6*, 1477-1488.

Berns, K., Hijmans, E.M., Mullenders, J., Brummelkamp, T.R., Velds, A., Heimerikx, M., Kerkhoven, R.M., Madiredjo, M., Nijkamp, W., Weigelt, B., *et al.* (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature *428*, 431-437.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A *99*, 8695-8700.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., *et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., *et al.* (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nature methods *3*, 199-204.

Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., and Dynlacht, B.D. (2005). An initial blueprint for myogenic differentiation. Genes Dev *19*, 553-569.

Blanchette, M., and Tompa, M. (2003). FootPrinter: A program designed for phylogenetic footprinting. Nucleic Acids Res *31*, 3840-3842.

Bockamp, E.O., Fordham, J.L., Gottgens, B., Murrell, A.M., Sanchez, M.J., and Green, A.R. (1998). Transcriptional regulation of the stem cell leukemia gene by PU.1 and Elf-1. J Biol Chem *273*, 29032-29042.

Bockamp, E.O., McLaughlin, F., Gottgens, B., Murrell, A.M., Elefanty, A.G., and Green, A.R. (1997). Distinct mechanisms direct SCL/tal-1 expression in erythroid cells and CD34 positive primitive myeloid cells. J Biol Chem *272*, 8781-8790.

Bode, J., Benham, C., Knopp, A., and Mielke, C. (2000). Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). Critical reviews in eukaryotic gene expression *10*, 73-90.

Bode, J., Goetze, S., Heng, H., Krawetz, S.A., and Benham, C. (2003). From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. Chromosome Res *11*, 435-445.

Boden, D., Pusch, O., Lee, F., Tucker, L., and Ramratnam, B. (2004a). Efficient gene transfer of HIV-1-specific short hairpin RNA into human lymphocytic cells using recombinant adeno-associated virus vectors. Mol Ther *9*, 396-402.

Boden, D., Pusch, O., Silbermann, R., Lee, F., Tucker, L., and Ramratnam, B. (2004b). Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. Nucleic acids research *32*, 1154-1158.

Boehm, T., Foroni, L., Kaneko, Y., Perutz, M.F., and Rabbitts, T.H. (1991). The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. Proceedings of the National Academy of Sciences of the United States of America *88*, 4367-4371.

Bose, F., Fugazza, C., Casalgrandi, M., Capelli, A., Cunningham, J.M., Zhao, Q., Jane, S.M., Ottolenghi, S., and Ronchi, A. (2006). Functional interaction of CP2 with GATA-1 in the regulation of erythroid promoters. Mol Cell Biol *26*, 3942-3954.

Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Paro, R., and Perrimon, N. (2004). Genome-wide RNAi analysis of growth and viability in Drosophila cells. Science (New York, NY *303*, 832-835.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Boyes, J., Byfield, P., Nakatani, Y., and Ogryzko, V. (1998). Regulation of activity of the transcription factor GATA-1 by acetylation. Nature *396*, 594-598.

Brandenberger, R., Khrebtukova, I., Thies, R.S., Miura, T., Jingli, C., Puri, R., Vasicek, T., Lebkowski, J., and Rao, M. (2004). MPSS profiling of human embryonic stem cells. BMC developmental biology *4*, 10.

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science (New York, NY *319*, 921-926.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., *et al.* (2000a). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol *18*, 630-634.

Brenner, S., Williams, S.R., Vermaas, E.H., Storck, T., Moon, K., McCollum, C., Mao, J.I., Luo, S., Kirchner, J.J., Eletr, S., *et al.* (2000b). In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. Proc Natl Acad Sci U S A *97*, 1665-1670.

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729-736.

Bridge, A.J., Pebernard, S., Ducraux, A., Nicoulaz, A.L., and Iggo, R. (2003). Induction of an interferon response by RNAi vectors in mammalian cells. Nature genetics *34*, 263-264.

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., and Batzoglou, S. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res *13*, 721-731.

Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science (New York, NY *296*, 550-553.

Brummelkamp, T.R., Nijman, S.M., Dirac, A.M., and Bernards, R. (2003). Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-kappaB. Nature *424*, 797-801.

Cai, J., Shin, S., Wright, L., Liu, Y., Zhou, D., Xue, H., Khrebtukova, I., Mattson, M.P., Svendsen, C.N., and Rao, M.S. (2006). Massively parallel signature sequencing profiling of fetal human neural precursor cells. Stem cells and development *15*, 232-244.

Cao, H.B., Wang, A., Martin, B., Koehler, D.R., Zeitlin, P.L., Tanawell, A.K., and Hu, J. (2005). Down-regulation of IL-8 expression in human airway epithelial cells through helper-dependent adenoviral-mediated RNA interference. Cell research *15*, 111-119.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. Proceedings of the National Academy of Sciences of the United States of America *98*, 9742-9747.
Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., *et al.* (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499-509.
Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev *18*, 1119-1130.
Chan, W.Y., Follows, G.A., Lacaud, G., Pimanda, J.E., Landry, J.R., Kinston, S., Knezevic, K., Piltz, S., Donaldson, I.J., Gambardella, L., *et al.* (2007). The paralogous hematopoietic regulators Lyl1 and Scl are coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl-/- phenotype. Blood *109*, 1908-1916.
Chen, C.C., Ko, T.M., Ma, H.I., Wu, H.L., Xiao, X., Li, J., Chang, C.M., Wu, P.Y., Chen, C.H., Han, J.M., *et al.* (2007). Long-term inhibition of hepatitis B virus in transgenic mice by double-stranded adeno-associated virus 8-delivered short hairpin RNA. Gene therapy *14*, 11-19.
Chen, J.D., and Evans, R.M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature *377*, 454-457.
Chen, L., and Widom, J. (2005). Mechanism of transcriptional silencing in yeast. Cell *120*, 37-48.
Chen, Y.T., Scanlan, M.J., Venditti, C.A., Chua, R., Theiler, G., Stevenson, B.J., Iseli, C., Gure, A.O., Vasicek, T., Strausberg, R.L., *et al.* (2005). Identification of cancer/testis-antigen genes by massively parallel signature sequencing. Proc Natl Acad Sci U S A *102*, 7940-7945.
Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J.C., and Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. Development *125*, 725-732.
Chung, Y.S., Zhang, W.J., Arentson, E., Kingsley, P.D., Palis, J., and Choi, K. (2002). Lineage analysis of the hemangioblast as defined by FLK1 and SCL expression. Development *129*, 5511-5520.
Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A., and Dixon, J.E. (2000). Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proceedings of the National Academy of Sciences of the United States of America *97*, 6499-6503.
Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet *26*, 183-186.

Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., and Golub, T.R. (2000). Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. Proc Natl Acad Sci U S A *97*, 3260-3265.
Courey, A.J., and Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell *55*, 887-898.
Crossley, M., Tsang, A.P., Bieker, J.J., and Orkin, S.H. (1994). Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. J Biol Chem *269*, 15440-15444.
Curtis, D.J., Hall, M.A., Van Stekelenburg, L.J., Robb, L., Jane, S.M., and Begley, C.G. (2004). SCL is required for normal function of short-term repopulating hematopoietic stem cells. Blood *103*, 3342-3348.
Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J., *et al.* (2004). Histone deimination antagonizes arginine methylation. Cell *118*, 545-553.
D'Souza, S.L., Elefanty, A.G., and Keller, G. (2005). SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development. Blood *105*, 3862-3870.
Dahmus, M.E. (1995). Phosphorylation of the C-terminal domain of RNA polymerase II. Biochimica et biophysica acta *1261*, 171-182.
Daniel, J.A., Torok, M.S., Sun, Z.W., Schieltz, D., Allis, C.D., Yates, J.R., 3rd, and Grant, P.A. (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. J Biol Chem *279*, 1867-1871.
Das, P.M., Ramachandran, K., vanWert, J., and Singal, R. (2004). Chromatin immunoprecipitation assay. BioTechniques *37*, 961-969.
Davuluri, R.V., Grosse, I., and Zhang, M.Q. (2001). Computational identification of promoters and first exons in the human genome. Nat Genet *29*, 412-417.
De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. Blood *110*, 4503-4510.
de Jonge, J., Holtrop, M., Wilschut, J., and Huckriede, A. (2006). Reconstituted influenza virus envelopes as an efficient carrier system for cellular delivery of small-interfering RNAs. Gene therapy *13*, 400-411.
Delabesse, E., Ogilvy, S., Chapman, M.A., Piltz, S.G., Gottgens, B., and Green, A.R. (2005). Transcriptional regulation of the SCL locus: identification of an enhancer that targets the primitive erythroid lineage in vivo. Mol Cell Biol *25*, 5215-5225.

Delassus, S., Titley, I., and Enver, T. (1999). Functional and molecular analysis of hematopoietic progenitors derived from the aorta-gonad-mesonephros region of the mouse embryo. Blood *94*, 1495-1503.
DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science (New York, NY *278*, 680-686.
Dermitzakis, E.T., and Clark, A.G. (2002). Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Molecular biology and evolution *19*, 1114-1121.
Diakos, C., Krapf, G., Gerner, C., Inthal, A., Lemberger, C., Ban, J., Dohnal, A.M., and Panzer-Gruemayer, E.R. (2007). RNAi-mediated silencing of TEL/AML1 reveals a heat-shock protein- and survivin-dependent mechanism for survival. Blood *109*, 2607-2610.
Dickins, R.A., Hemann, M.T., Zilfou, J.T., Simpson, D.R., Ibarra, I., Hannon, G.J., and Lowe, S.W. (2005). Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. Nature genetics *37*, 1289-1295.
Down, T.A., and Hubbard, T.J. (2002). Computational detection and location of transcription start sites in mammalian genomic DNA. Genome Res *12*, 458-461.
Down, T.A., and Hubbard, T.J. (2005). NestedMICA: sensitive inference of over-represented motifs in nucleic acid sequence. Nucleic Acids Res *33*, 1445-1453.
Eisbacher, M., Holmes, M.L., Newton, A., Hogg, P.J., Khachigian, L.M., Crossley, M., and Chong, B.H. (2003). Protein-protein interaction between Fli-1 and GATA-1 mediates synergistic expression of megakaryocyte-specific genes through cooperative DNA binding. Mol Cell Biol *23*, 3427-3441.
Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.
Elefanty, A.G., Begley, C.G., Hartley, L., Papaevangeliou, B., and Robb, L. (1999). SCL expression in the mouse embryo detected with a targeted lacZ reporter gene demonstrates its localization to hematopoietic, vascular, and neural tissues. Blood *94*, 3754-3763.
Elefanty, A.G., Robb, L., Birner, R., and Begley, C.G. (1997). Hematopoietic-specific genes are not induced during in vitro differentiation of scl-null embryonic stem cells. Blood *90*, 1435-1447.
Elkon, R., Rashi-Elkeles, S., Lerenthal, Y., Linhart, C., Tenne, T., Amariglio, N., Rechavi, G., Shamir, R., and Shiloh, Y. (2005). Dissection of a DNA-damage-induced transcriptional network using a combination of microarrays, RNA interference and computational promoter analysis. Genome biology *6*, R43.
Elwood, N.J., Green, A.R., Melder, A., Begley, C.G., and Nicola, N. (1994). The SCL protein displays cell-specific heterogeneity in size. Leukemia *8*, 106-114.

Endoh, M., Ogawa, M., Orkin, S., and Nishikawa, S. (2002). SCL/tal-1-dependent process determines a competence to select the definitive hematopoietic lineage prior to endothelial differentiation. Embo J *21*, 6700-6708.
Euskirchen, G., Royce, T.E., Bertone, P., Martone, R., Rinn, J.L., Nelson, F.K., Sayward, F., Luscombe, N.M., Miller, P., Gerstein, M., *et al.* (2004). CREB binds to multiple loci on human chromosome 22. Mol Cell Biol *24*, 3804-3814.
Evans, T., Reitman, M., and Felsenfeld, G. (1988). An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc Natl Acad Sci U S A *85*, 5976-5980.
Fedorov, Y., Anderson, E.M., Birmingham, A., Reynolds, A., Karpilow, J., Robinson, K., Leake, D., Marshall, W.S., and Khvorova, A. (2006). Off-target effects by siRNA can induce toxic phenotype. RNA (New York, NY *12*, 1188-1196.
Ferrando, A.A., Neuberg, D.S., Staunton, J., Loh, M.L., Huard, C., Raimondi, S.C., Behm, F.G., Pui, C.H., Downing, J.R., Gilliland, D.G., *et al.* (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer cell *1*, 75-87.
Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. Mol Cell Biol *25*, 1215-1227.
Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development *113*, 503-514.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.
Fordham, J.L., Gottgens, B., McLaughlin, F., and Green, A.R. (1999). Chromatin structure and transcriptional regulation of the stem cell leukaemia (SCL) gene in mast cells. Leukemia *13*, 750-759.
Fox, A.H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. Embo J *18*, 2812-2822.
Freeman, L.A., and Garrard, W.T. (1992). DNA supercoiling in chromatin structure and gene expression. Critical reviews in eukaryotic gene expression *2*, 165-209.
Freson, K., Devriendt, K., Matthijs, G., Van Hoof, A., De Vos, R., Thys, C., Minner, K., Hoylaerts, M.F., Vermylen, J., and Van Geet, C. (2001). Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. Blood *98*, 85-92.

Freson, K., Matthijs, G., Thys, C., Marien, P., Hoylaerts, M.F., Vermylen, J., and Van Geet, C. (2002). Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation. Human molecular genetics *11*, 147-152.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C., and Orkin, S.H. (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.

Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res *5*, 3157-3170.

Garner, M.M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res *9*, 3047-3060.

Ge, Y., Dombkowski, A.A., LaFiura, K.M., Tatman, D., Yedidi, R.S., Stout, M.L., Buck, S.A., Massey, G., Becton, D.L., Weinstein, H.J.*, et al.* (2006). Differential gene expression, GATA1 target genes, and the chemotherapy sensitivity of Down syndrome megakaryocytic leukemia. Blood *107*, 1570-1581.

Gershenzon, N.I., and Ioshikhes, I.P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics (Oxford, England) *21*, 1295-1300.

Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O., and Petes, T.D. (2000). Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *97*, 11383-11390.

Giroux, S., Kaushik, A.L., Capron, C., Jalil, A., Kelaidi, C., Sablitzky, F., Dumenil, D., Albagli, O., and Godin, I. (2007). lyl-1 and tal-1/scl, two genes encoding closely related bHLH transcription factors, display highly overlapping expression patterns during cardiovascular and hematopoietic ontogeny. Gene Expr Patterns *7*, 215-226.

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

Godin, I., and Cumano, A. (2002). The hare and the tortoise: an embryonic haematopoietic race. Nat Rev Immunol *2*, 593-604.

Godin, I., Garcia-Porrero, J.A., Dieterlen-Lievre, F., and Cumano, A. (1999). Stem cell emergence and hemopoietic activity are incompatible in mouse intraembryonic sites. The Journal of experimental medicine *190*, 43-52.

Goldfarb, A.N., Goueli, S., Mickelson, D., and Greenberg, J.M. (1992). T-cell acute lymphoblastic leukemia--the associated gene SCL/tal codes for a 42-Kd nuclear phosphoprotein. Blood *80*, 2858-2866.
Gonczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S.J., Copley, R.R., Duperon, J., Oegema, J., Brehm, M., Cassin, E.*, et al.* (2000). Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. Nature *408*, 331-336.
Gonzalez, G.A., and Montminy, M.R. (1989). Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell *59*, 675-680.
Gottgens, B., Barton, L.M., Gilbert, J.G., Bench, A.J., Sanchez, M.J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M.*, et al.* (2000). Analysis of vertebrate SCL loci identifies conserved enhancers. Nature biotechnology *18*, 181-186.
Gottgens, B., Broccardo, C., Sanchez, M.J., Deveaux, S., Murphy, G., Gothert, J.R., Kotsopoulou, E., Kinston, S., Delaney, L., Piltz, S.*, 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.
Gottgens, B., McLaughlin, F., Bockamp, E.O., Fordham, J.L., Begley, C.G., Kosmopoulos, K., Elefanty, A.G., and Green, A.R. (1997). Transcription of the SCL gene in erythroid and CD34 positive primitive myeloid cells is controlled by a complex network of lineage-restricted chromatin-dependent and chromatin-independent regulatory elements. Oncogene *15*, 2419-2428.
Gottgens, B., Nastos, A., Kinston, S., Piltz, S., Delabesse, E.C., Stanley, M., Sanchez, M.J., Ciau-Uitz, A., Patient, R., and Green, A.R. (2002). Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. Embo J *21*, 3039-3050.
Gou, D., Jin, N., and Liu, L. (2003). Gene silencing in mammalian cells by PCR-based short hairpin RNA. FEBS letters *548*, 113-118.
Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., and Workman, J.L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. J Biol Chem *274*, 5895-5900.
Grass, J.A., Boyer, M.E., Pal, S., Wu, J., Weiss, M.J., and Bresnick, E.H. (2003). GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. Proc Natl Acad Sci U S A *100*, 8811-8816.
Green, A.R., Lints, T., Visvader, J., Harvey, R., and Begley, C.G. (1992). SCL is coexpressed with GATA-1 in hemopoietic cells but is also expressed in developing brain. Oncogene *7*, 653-660.
Green, A.R., Rockman, S., DeLuca, E., and Begley, C.G. (1993). Induced myeloid differentiation of K562 cells with downregulation of erythroid and megakaryocytic transcription factors: a novel experimental model for hemopoietic lineage restriction. Exp Hematol *21*, 525-531.

Green, A.R., Salvaris, E., and Begley, C.G. (1991). Erythroid expression of the 'helix-loop-helix' gene, SCL. Oncogene *6*, 475-479.
Greenbaum, S., and Zhuang, Y. (2002). Identification of E2A target genes in B lymphocyte development by using a gene tagging-based chromatin immunoprecipitation system. Proceedings of the National Academy of Sciences of the United States of America *99*, 15030-15035.
Gregory, R.C., Taxman, D.J., Seshasayee, D., Kensinger, M.H., Bieker, J.J., and Wojchowski, D.M. (1996). Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. Blood *87*, 1793-1801.
Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F., and Kay, M.A. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature *441*, 537-541.
Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu Rev Biochem *57*, 159-197.
Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell *51*, 975-985.
Grutz, G.G., Bucher, K., Lavenir, I., Larson, T., Larson, R., and Rabbitts, T.H. (1998). The oncogenic T cell LIM-protein Lmo2 forms part of a DNA-binding complex specifically in immature T cells. The EMBO journal *17*, 4594-4605.
Gu, W., and Roeder, R.G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell *90*, 595-606.
Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.
Gupta, S., Schoer, R.A., Egan, J.E., Hannon, G.J., and Mittal, V. (2004). Inducible, reversible, and stable RNA interference in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 1927-1932.
Gyorffy, B., Surowiak, P., Kiesslich, O., Denkert, C., Schafer, R., Dietel, M., and Lage, H. (2005). Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. Int J Cancer.
Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol *11*, 394-403.
Hall, M.A., Curtis, D.J., Metcalf, D., Elefanty, A.G., Sourris, K., Robb, L., Gothert, J.R., Jane, S.M., and Begley, C.G. (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.

Hall, M.A., Slater, N.J., Begley, C.G., Salmon, J.M., Van Stekelenburg, L.J., McCormack, M.P., Jane, S.M., and Curtis, D.J. (2005). Functional but abnormal adult erythropoiesis in the absence of the stem cell leukemia gene. Mol Cell Biol *25*, 6355-6362.

Harris, M.B., Mostecki, J., and Rothman, P.B. (2005). Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. J Biol Chem *280*, 13114-13121.

Hassa, P.O., Haenni, S.S., Elser, M., and Hottiger, M.O. (2006). Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev *70*, 789-829.

Hassan, A.H., Neely, K.E., and Workman, J.L. (2001). Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. Cell *104*, 817-827.

Heemskerk, M.H., Blom, B., Nolan, G., Stegmann, A.P., Bakker, A.Q., Weijer, K., Res, P.C., and Spits, H. (1997). Inhibition of T cell and promotion of natural killer cell development by the dominant negative helix loop helix factor Id3. The Journal of experimental medicine *186*, 1597-1602.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., *et al.* (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res *26*, 362-367.

Heinzel, T., Lavinsky, R.M., Mullen, T.M., Soderstrom, M., Laherty, C.D., Torchia, J., Yang, W.M., Brard, G., Ngo, S.D., Davie, J.R., *et al.* (1997). A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. Nature *387*, 43-48.

Hirasawa, R., Shimizu, R., Takahashi, S., Osawa, M., Takayanagi, S., Kato, Y., Onodera, M., Minegishi, N., Yamamoto, M., Fukao, K., *et al.* (2002). Essential and instructive roles of GATA factors in eosinophil development. J Exp Med *195*, 1379-1386.

Holmes, M., Turner, J., Fox, A., Chisholm, O., Crossley, M., and Chong, B. (1999). hFOG-2, a novel zinc finger protein, binds the co-repressor mCtBP2 and modulates GATA-mediated activation. J Biol Chem *274*, 23491-23498.

Hope, I.A., Mahadevan, S., and Struhl, K. (1988). Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. Nature *333*, 635-640.

Hoppe, K.L., and Francone, O.L. (1998). Binding and functional effects of transcription factors Sp1 and Sp3 on the proximal human lecithin:cholesterol acyltransferase promoter. Journal of lipid research *39*, 969-977.

Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. (2002). GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci U S A *99*, 2924-2929.

Horner, M.A., Quintin, S., Domeier, M.E., Kimble, J., Labouesse, M., and Mango, S.E. (1998). pha-4, an HNF-3 homolog, specifies pharyngeal organ identity in Caenorhabditis elegans. Genes Dev *12*, 1947-1952.

Hsieh, A.C., Bo, R., Manola, J., Vazquez, F., Bare, O., Khvorova, A., Scaringe, S., and Sellers, W.R. (2004). A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. Nucleic acids research *32*, 893-901.

Hsu, H.L., Cheng, J.T., Chen, Q., and Baer, R. (1991). Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. Mol Cell Biol *11*, 3037-3042.

Hsu, H.L., Wadman, I., and Baer, R. (1994a). Formation of in vivo complexes between the TAL1 and E2A polypeptides of leukemic T cells. Proc Natl Acad Sci U S A *91*, 3181-3185.

Hsu, H.L., Wadman, I., Tsan, J.T., and Baer, R. (1994b). Positive and negative transcriptional control by the TAL1 helix-loop-helix protein. Proc Natl Acad Sci U S A *91*, 5947-5951.

Huang, D.Y., Kuo, Y.Y., and Chang, Z.F. (2005). GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. Nucleic Acids Res *33*, 5331-5342.

Huang, D.Y., Kuo, Y.Y., Lai, J.S., Suzuki, Y., Sugano, S., and Chang, Z.F. (2004). GATA-1 and NF-Y cooperate to mediate erythroid-specific transcription of Gfi-1B gene. Nucleic Acids Res *32*, 3935-3946.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., *et al.* (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol *19*, 342-347.

Hwang, L.Y., Siegelman, M., Davis, L., Oppenheimer-Marks, N., and Baer, R. (1993). Expression of the TAL1 proto-oncogene in cultured endothelial cells and blood vessels of the spleen. Oncogene *8*, 3043-3046.

Ikawa, T., Kawamoto, H., Wright, L.Y., and Murre, C. (2004). Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. Immunity *20*, 349-360.

Imagawa, S., Suzuki, N., Ohmine, K., Obara, N., Mukai, H.Y., Ozawa, K., Yamamoto, M., and Nagasawa, T. (2002). GATA suppresses erythropoietin gene expression through GATA site in mouse erythropoietin gene promoter. International journal of hematology *75*, 376-381.

Imagawa, S., Yamamoto, M., and Miura, Y. (1997). Negative regulation of the erythropoietin gene expression by the GATA transcription factors. Blood *89*, 1430-1439.

Inaba, T., Roberts, W.M., Shapiro, L.H., Jolly, K.W., Raimondi, S.C., Smith, S.D., and Look, A.T. (1992). Fusion of the leucine zipper gene HLF to the E2A gene in human acute B-lineage leukemia. Science *257*, 531-534.

Iwasaki, H., Mizuno, S., Wells, R.A., Cantor, A.B., Watanabe, S., and Akashi, K. (2003). GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. Immunity *19*, 451-462.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533-538.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol *21*, 635-637.

Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L., and Linsley, P.S. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA (New York, NY *12*, 1179-1187.

Jazag, A., Ijichi, H., Kanai, F., Imamura, T., Guleng, B., Ohta, M., Imamura, J., Tanaka, Y., Tateishi, K., Ikenoue, T.*, et al.* (2005). Smad4 silencing in pancreatic cancer cell lines using stable RNA interference and gene expression profiles induced by transforming growth factor-beta. Oncogene *24*, 662-671.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nature cell biology *10*, 353-360.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.

Johnson, E.S. (2004). Protein modification by SUMO. Annu Rev Biochem *73*, 355-382.

Jongeneel, C.V., Delorenzi, M., Iseli, C., Zhou, D., Haudenschild, C.D., Khrebtukova, I., Kuznetsov, D., Stevenson, B.J., Strausberg, R.L., Simpson, A.J.*, et al.* (2005). An atlas of human gene expression from massively parallel signature sequencing (MPSS). Genome Res *15*, 1007-1014.

Jurata, L.W., and Gill, G.N. (1997). Functional analysis of the nuclear LIM domain interactor NLI. Molecular and cellular biology *17*, 5688-5698.

Kallianpur, A.R., Jordan, J.E., and Brandt, S.J. (1994). The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis. Blood *83*, 1200-1208.

Kappel, A., Schlaeger, T.M., Flamme, I., Orkin, S.H., Risau, W., and Breier, G. (2000). Role of SCL/Tal-1, GATA, and ets transcription factor binding sites for the regulation of flk-1 expression during murine vascular development. Blood *96*, 3078-3085.

Kawasaki, H., and Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature *431*, 211-217.
Kee, B.L., and Murre, C. (1998). Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12. The Journal of experimental medicine *188*, 699-713.
Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., and Wingender, E. (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res *31*, 3576-3579.
Kennedy, M., Firpo, M., Choi, K., Wall, C., Robertson, S., Kabrun, N., and Keller, G. (1997). A common precursor for primitive erythropoiesis and definitive haematopoiesis. Nature *386*, 488-493.
Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell *95*, 1017-1026.
Kennerdell, J.R., and Carthew, R.W. (2000). Heritable gene silencing in Drosophila using double-stranded RNA. Nat Biotechnol *18*, 896-898.
Kim, J., Bhinge, A.A., Morgan, X.C., and Iyer, V.R. (2005a). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. Nature methods *2*, 47-53.
Kim, J., and Iyer, V.R. (2004). Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles. Mol Cell Biol *24*, 8104-8112.
Kim, J.B., Porreca, G.J., Song, L., Greenway, S.C., Gorham, J.M., Church, G.M., Seidman, C.E., and Seidman, J.G. (2007). Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. Science (New York, NY *316*, 1481-1484.
Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M.*, et al.* (2005b). Direct isolation and identification of promoters in the human genome. Genome Res *15*, 830-839.
Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005c). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.
Kittler, R., Pelletier, L., Heninger, A.K., Slabicki, M., Theis, M., Miroslaw, L., Poser, I., Lawo, S., Grabner, H., Kozak, K.*, et al.* (2007). Genome-scale RNAi profiling of cell division in human tissue culture cells. Nature cell biology *9*, 1401-1412.
Kittler, R., Putz, G., Pelletier, L., Poser, I., Heninger, A.K., Drechsel, D., Fischer, S., Konstantinova, I., Habermann, B., Grabner, H.*, et al.* (2004). An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. Nature *432*, 1036-1040.
Kloosterman, W.P., and Plasterk, R.H. (2006). The diverse functions of microRNAs in animal development and disease. Developmental cell *11*, 441-450.

Kobayashi, M., Nishikawa, K., and Yamamoto, M. (2001). Hematopoietic regulatory domain of gata1 gene is positively regulated by GATA1 protein in zebrafish embryos. Development (Cambridge, England) *128*, 2341-2350.
Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., *et al.* (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res *17*, 691-707.
Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., *et al.* (2006). CAGE: cap analysis of gene expression. Nature methods *3*, 211-222.
Koizumi, K., Higashida, H., Yoo, S., Islam, M.S., Ivanov, A.I., Guo, V., Pozzi, P., Yu, S.H., Rovescalli, A.C., Tang, D., *et al.* (2007). RNA interference screen to identify genes required for Drosophila embryonic nervous system development. Proceedings of the National Academy of Sciences of the United States of America *104*, 5626-5631.
Kolfschoten, I.G., van Leeuwen, B., Berns, K., Mullenders, J., Beijersbergen, R.L., Bernards, R., Voorhoeve, P.M., and Agami, R. (2005). A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. Cell *121*, 849-858.
Krosl, G., He, G., Lefrancois, M., Charron, F., Romeo, P.H., Jolicoeur, P., Kirsch, I.R., Nemer, M., and Hoang, T. (1998). Transcription factor SCL is required for c-kit expression and c-Kit function in hemopoietic cells. J Exp Med *188*, 439-450.
Kuhl, C., Atzberger, A., Iborra, F., Nieswandt, B., Porcher, C., and Vyas, P. (2005). GATA1-mediated megakaryocyte differentiation and growth control can be uncoupled and mapped to different domains in GATA1. Mol Cell Biol *25*, 8592-8606.
Kuhn, K., Baker, S.C., Chudin, E., Lieu, M.H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T.K., and Chee, M.S. (2004). A novel, high-performance random array platform for quantitative gene expression profiling. Genome Res *14*, 2347-2356.
Kulessa, H., Frampton, J., and Graf, T. (1995). GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Genes Dev *9*, 1250-1262.
Kumaravelu, P., Hook, L., Morrison, A.M., Ure, J., Zhao, S., Zuyev, S., Ansell, J., and Medvinsky, A. (2002). Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. Development *129*, 4891-4899.
Kurdistani, S.K., Tavazoie, S., and Grunstein, M. (2004). Mapping global histone acetylation patterns to gene expression. Cell *117*, 721-733.
Labastie, M.C., Cortes, F., Romeo, P.H., Dulac, C., and Peault, B. (1998). Molecular identity of hematopoietic precursor cells emerging in the human embryo. Blood *92*, 3624-3635.

Lahlil, R., Lecuyer, E., Herblot, S., and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. Mol Cell Biol *24*, 1439-1452.
Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. Science *280*, 547-553.
Landry, J.R., Kinston, S., Knezevic, K., de Bruijn, M.F., Wilson, N., Nottingham, W.T., Peitz, M., Edenhofer, F., Pimanda, J.E., Ottersbach, K.*, et al.* (2008). Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. Blood *111*, 3005-3014.
Larson, R.C., Lavenir, I., Larson, T.A., Baer, R., Warren, A.J., Wadman, I., Nottage, K., and Rabbitts, T.H. (1996). Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. The EMBO journal *15*, 1021-1027.
Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell *66*, 305-315.
Lecointe, N., Bernard, O., Naert, K., Joulin, V., Larsen, C.J., Romeo, P.H., and Mathieu-Mahul, D. (1994). GATA-and SP1-binding sites are required for the full activity of the tissue-specific promoter of the tal-1 gene. Oncogene *9*, 2623-2632.
Lecuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C.G., Porcher, C., Orkin, S.H., and Hoang, T. (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood *100*, 2430-2440.
Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005a). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol Cell Biol *25*, 9674-9686.
Lee, D.Y., Teyssier, C., Strahl, B.D., and Stallcup, M.R. (2005b). Role of protein methylation in regulation of transcription. Endocrine reviews *26*, 147-170.
Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.
Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.
Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I.*, et al.* (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799-804.
Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annual review of genetics *34*, 77-137.

Leroy-Viard, K., Vinit, M.A., Lecointe, N., Mathieu-Mahul, D., and Romeo, P.H. (1994). Distinct DNase-I hypersensitive sites are associated with TAL-1 transcription in erythroid and T-cell lines. Blood *84*, 3819-3827.

Letting, D.L., Chen, Y.Y., Rakowski, C., Reedy, S., and Blobel, G.A. (2004). Context-dependent regulation of GATA-1 by friend of GATA-1. Proc Natl Acad Sci U S A *101*, 476-481.

Leuther, K.K., Salmeron, J.M., and Johnston, S.A. (1993). Genetic evidence that an activation domain of GAL4 does not require acidity and may form a beta sheet. Cell *72*, 575-585.

Li, M., Li, H., and Rossi, J.J. (2006). RNAi in combination with a ribozyme and TAR decoy for treatment of HIV infection in hematopoietic cell gene therapy. Annals of the New York Academy of Sciences *1082*, 172-179.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood *100*, 3077-3086.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A *100*, 8164-8169.

Liang, P., and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science (New York, NY *257*, 967-971.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet *28*, 327-334.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev *18*, 1606-1617.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. Nat Genet *21*, 20-24.

Liu, X., Noll, D.M., Lieb, J.D., and Clarke, N.D. (2005). DIP-chip: rapid and accurate determination of DNA-binding specificity. Genome Res *15*, 421-427.

Liu, X.S., Brutlag, D.L., and Liu, J.S. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nature biotechnology *20*, 835-839.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

Lorch, Y., Zhang, M., and Kornberg, R.D. (1999). Histone octamer transfer by a chromatin-remodeling complex. Cell *96*, 389-392.

Louie, M.C., Yang, H.Q., Ma, A.H., Xu, W., Zou, J.X., Kung, H.J., and Chen, H.W. (2003). Androgen-induced recruitment of RNA polymerase II to a nuclear receptor-p160 coactivator complex. Proc Natl Acad Sci U S A *100*, 2226-2230.
Lugus, J.J., Chung, Y.S., Mills, J.C., Kim, S.I., Grass, J., Kyba, M., Doherty, J.M., Bresnick, E.H., and Choi, K. (2007). GATA2 functions at multiple steps in hemangioblast development and differentiation. Development (Cambridge, England) *134*, 393-405.
Lum, L., Yao, S., Mozer, B., Rovescalli, A., Von Kessler, D., Nirenberg, M., and Beachy, P.A. (2003). Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells. Science (New York, NY *299*, 2039-2045.
Mackay, J.P., Kowalski, K., Fox, A.H., Czolij, R., King, G.F., and Crossley, M. (1998). Involvement of the N-finger in the self-association of GATA-1. J Biol Chem *273*, 30560-30567.
MacKeigan, J.P., Murphy, L.O., and Blenis, J. (2005). Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. Nature cell biology *7*, 591-600.
Mahadevan, L.C., Willis, A.C., and Barratt, M.J. (1991). Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. Cell *65*, 775-783.
Mansson, R., Tsapogas, P., Akerlund, M., Lagergren, A., Gisler, R., and Sigvardsson, M. (2004). Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor. The Journal of biological chemistry *279*, 17905-17913.
Mao, D.Y., Watson, J.D., Yan, P.S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W.W., Farnham, P.J., Huang, T.H., and Penn, L.Z. (2003). Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. Curr Biol *13*, 882-886.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.
Marques, J.T., Devosse, T., Wang, D., Zamanian-Daryoush, M., Serbinowski, P., Hartmann, R., Fujita, T., Behlke, M.A., and Williams, B.R. (2006). A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. Nat Biotechnol *24*, 559-565.
Martin, D.I., and Orkin, S.H. (1990). Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes Dev *4*, 1886-1898.
Martin, D.I., Zon, L.I., Mutter, G., and Orkin, S.H. (1990). Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. Nature *344*, 444-447.

Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C.J., Stoughton, R., Burchard, J., Slade, D., Dai, H., *et al.* (1998). Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat Med *4*, 1293-1301.
Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., *et al.* (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. Proc Natl Acad Sci U S A *100*, 12247-12252.
Massari, M.E., Grant, P.A., Pray-Grant, M.G., Berger, S.L., Workman, J.L., and Murre, C. (1999). A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. Molecular cell *4*, 63-73.
Matsukura, S., Jones, P.A., and Takai, D. (2003). Establishment of conditional vectors for hairpin siRNA knockdowns. Nucleic acids research *31*, e77.
Matsumoto, N., Kubo, A., Liu, H., Akita, K., Laub, F., Ramirez, F., Keller, G., and Friedman, S.L. (2006). Developmental regulation of yolk sac hematopoiesis by Kruppel-like factor 6. Blood *107*, 1357-1365.
Matsumura, I., Kawasaki, A., Tanaka, H., Sonoyama, J., Ezoe, S., Minegishi, N., Nakajima, K., Yamamoto, M., and Kanakura, Y. (2000). Biologic significance of GATA-1 activities in Ras-mediated megakaryocytic differentiation of hematopoietic cell lines. Blood *96*, 2440-2450.
Medina, K.L., Pongubala, J.M., Reddy, K.L., Lancki, D.W., Dekoter, R., Kieslinger, M., Grosschedl, R., and Singh, H. (2004). Assembling a gene regulatory network for specification of the B cell fate. Developmental cell *7*, 607-617.
Medvinsky, A., and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. Cell *86*, 897-906.
Meersseman, G., Pennings, S., and Bradbury, E.M. (1992). Mobile nucleosomes--a general behavior. Embo J *11*, 2951-2959.
Mehaffey, M.G., Newton, A.L., Gandhi, M.J., Crossley, M., and Drachman, J.G. (2001). X-linked thrombocytopenia caused by a novel mutation of GATA-1. Blood *98*, 2681-2688.
Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.
Mellentin, J.D., Smith, S.D., and Cleary, M.L. (1989). lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell *58*, 77-83.
Meng, A., Tang, H., Yuan, B., Ong, B.A., Long, Q., and Lin, S. (1999). Positive and negative cis-acting elements are required for hematopoietic expression of zebrafish GATA-1. Blood *93*, 500-508.

Merika, M., and Orkin, S.H. (1995). Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. Mol Cell Biol *15*, 2437-2447.
Mermod, N., O'Neill, E.A., Kelly, T.J., and Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. Cell *58*, 741-753.
Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436-439.
Migliaccio, A.R., Rana, R.A., Sanchez, M., Lorenzini, R., Centurione, L., Bianchi, L., Vannucchi, A.M., Migliaccio, G., and Orkin, S.H. (2003). GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1low mouse mutant. J Exp Med *197*, 281-296.
Mikkola, H.K., Klintman, J., Yang, H., Hock, H., Schlaeger, T.M., Fujiwara, Y., and Orkin, S.H. (2003). Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. Nature *421*, 547-551.
Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824-827.
Milyavsky, M., Tabach, Y., Shats, I., Erez, N., Cohen, Y., Tang, X., Kalis, M., Kogan, I., Buganim, Y., Goldfinger, N., *et al.* (2005). Transcriptional programs following genetic alterations in p53, INK4A, and H-Ras genes along defined stages of malignant transformation. Cancer research *65*, 4530-4543.
Mineno, J., Okamoto, S., Ando, T., Sato, M., Chono, H., Izu, H., Takayama, M., Asada, K., Mirochnitchenko, O., Inouye, M., *et al.* (2006). The expression profile of microRNAs in mouse embryos. Nucleic acids research *34*, 1765-1771.
Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., *et al.* (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell *124*, 1283-1298.
Molkentin, J.D. (2000). The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. J Biol Chem *275*, 38949-38952.
Moore, M.A., and Metcalf, D. (1970). Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo. British journal of haematology *18*, 279-296.
Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., *et al.* (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res.

Moroni, E., Mastrangelo, T., Razzini, R., Cairns, L., Moi, P., Ottolenghi, S., and Giglioni, B. (2000). Regulation of mouse p45 NF-E2 transcription by an erythroid-specific GATA-dependent intronic alternative promoter. J Biol Chem *275*, 10567-10576.

Morris, K.V., Chan, S.W., Jacobsen, S.E., and Looney, D.J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science (New York, NY *305*, 1289-1292.

Mouthon, M.A., Bernard, O., Mitjavila, M.T., Romeo, P.H., Vainchenker, W., and Mathieu-Mahul, D. (1993). Expression of tal-1 and GATA-binding proteins during human hematopoiesis. Blood *81*, 647-655.

Mukhopadhyay, M., Teufel, A., Yamashita, T., Agulnick, A.D., Chen, L., Downs, K.M., Schindler, A., Grinberg, A., Huang, S.P., Dorward, D.*, et al.* (2003). Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. Development (Cambridge, England) *130*, 495-505.

Murre, C., McCaw, P.S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell *56*, 777-783.

Myers, J.W., Jones, J.T., Meyer, T., and Ferrell, J.E., Jr. (2003). Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. Nat Biotechnol *21*, 324-328.

Myers, L.C., and Kornberg, R.D. (2000). Mediator of transcriptional regulation. Annu Rev Biochem *69*, 729-749.

Nagy, P.L., Griesenbeck, J., Kornberg, R.D., and Cleary, M.L. (2002). A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. Proc Natl Acad Sci U S A *99*, 90-94.

Neale, G.A., Rehg, J.E., and Goorha, R.M. (1995). Ectopic expression of rhombotin-2 causes selective expansion of CD4-CD8- lymphocytes in the thymus and T-cell tumors in transgenic mice. Blood *86*, 3060-3071.

Nelson, C.J., Santos-Rosa, H., and Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell *126*, 905-916.

Ngo, V.N., Davis, R.E., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, L.T., Dave, S., Yang, L., Powell, J.*, et al.* (2006). A loss-of-function RNA interference screen for molecular targets in cancer. Nature *441*, 106-110.

Nichols, K.E., Crispino, J.D., Poncz, M., White, J.G., Orkin, S.H., Maris, J.M., and Weiss, M.J. (2000). Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. Nature genetics *24*, 266-270.

Nicolis, S., Bertini, C., Ronchi, A., Crotta, S., Lanfranco, L., Moroni, E., Giglioni, B., and Ottolenghi, S. (1991). An erythroid specific enhancer upstream to the gene encoding the cell-type specific transcription factor GATA-1. Nucleic Acids Res *19*, 5285-5291.

Nilsson, R., Bajic, V.B., Suzuki, H., di Bernardo, D., Bjorkegren, J., Katayama, S., Reid, J.F., Sweet, M.J., Gariboldi, M., Carninci, P., *et al.* (2006). Transcriptional network dynamics in macrophage activation. Genomics *88*, 133-142.

Nishikawa, K., Kobayashi, M., Masumi, A., Lyons, S.E., Weinstein, B.M., Liu, P.P., and Yamamoto, M. (2003). Self-association of Gata1 enhances transcriptional activity in vivo in zebra fish embryos. Mol Cell Biol *23*, 8295-8305.

Nishimura, S., Takahashi, S., Kuroha, T., Suwabe, N., Nagasawa, T., Trainor, C., and Yamamoto, M. (2000). A GATA box in the GATA-1 gene hematopoietic enhancer is a critical element in the network of GATA factors and sites that regulate this gene. Mol Cell Biol *20*, 713-723.

Nottingham, W.T., Jarratt, A., Burgess, M., Speck, C.L., Cheng, J.F., Prabhakar, S., Rubin, E.M., Li, P.S., Sloane-Stanley, J., Kong, A.S.J., *et al.* (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood *110*, 4188-4197.

Nourse, J., Mellentin, J.D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D., and Cleary, M.L. (1990). Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell *60*, 535-545.

Nunez, E., Kwon, Y.S., Hutt, K.R., Hu, Q., Cardamone, M.D., Ohgi, K.A., Garcia-Bassets, I., Rose, D.W., Glass, C.K., Rosenfeld, M.G., *et al.* (2008). Nuclear receptor-enhanced transcription requires motor- and LSD1-dependent gene networking in interchromatin granules. Cell *134*, 189.

O'Neil, J., Shank, J., Cusson, N., Murre, C., and Kelliher, M. (2004). TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB. Cancer Cell *5*, 587-596.

Oberley, M.J., Inman, D.R., and Farnham, P.J. (2003). E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9. J Biol Chem *278*, 42466-42476.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Gordon, W., Danford, T.W., MacIsaac, K.D., Rolfe, P.A., Conboy, C.M., Gifford, D.K., and Fraenkel, E. (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat Genet *39*, 730-732.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., *et al.* (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378-1381.

Ogilvy, S., Ferreira, R., Piltz, S.G., Bowen, J.M., Gottgens, B., and Green, A.R. (2007). The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. Mol Cell Biol *27*, 7206-7219.

Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S., Chen, H., Skiena, S., Futcher, B., and Leatherwood, J. (2005). The cell cycle-regulated genes of Schizosaccharomyces pombe. PLoS Biol *3*, e225.

Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Developmental biology *216*, 671-680.
Onizuka, T., Endo, S., Hirano, M., Kanai, S., and Akiyama, H. (2002). Design of a fluorescent electrophoretic mobility shift assay improved for the quantitative and multiple analysis of protein-DNA complexes. Bioscience, biotechnology, and biochemistry *66*, 2732-2734.
Ono, Y., Fukuhara, N., and Yoshie, O. (1997). Transcriptional activity of TAL1 in T cell acute lymphoblastic leukemia (T-ALL) requires RBTN1 or -2 and induces TALLA1, a highly specific tumor marker of T-ALL. J Biol Chem *272*, 4576-4581.
Ono, Y., Fukuhara, N., and Yoshie, O. (1998). TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3. Mol Cell Biol *18*, 6939-6950.
Onodera, K., Takahashi, S., Nishimura, S., Ohta, J., Motohashi, H., Yomogida, K., Hayashi, N., Engel, J.D., and Yamamoto, M. (1997a). GATA-1 transcription is controlled by distinct regulatory mechanisms during primitive and definitive erythropoiesis. Proc Natl Acad Sci U S A *94*, 4487-4492.
Onodera, K., Yomogida, K., Suwabe, N., Takahashi, S., Muraosa, Y., Hayashi, N., Ito, E., Gu, L., Rassoulzadegan, M., Engel, J.D., *et al.* (1997b). Conserved structure, regulatory elements, and transcriptional regulation from the GATA-1 gene testis promoter. Journal of biochemistry *121*, 251-263.
Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W., Cowley, S.M., Yost, C., *et al.* (2003). Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes Dev *17*, 1101-1114.
Pabo, C.O., and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem *61*, 1053-1095.
Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes & development *16*, 948-958.
Pal, S., Cantor, A.B., Johnson, K.D., Moran, T.B., Boyer, M.E., Orkin, S.H., and Bresnick, E.H. (2004). Coregulator-dependent facilitation of chromatin occupancy by GATA-1. Proc Natl Acad Sci U S A *101*, 980-985.
Palomero, T., Odom, D.T., O'Neil, J., Ferrando, A.A., Margolin, A., Neuberg, D.S., Winter, S.S., Larson, R.S., Li, W., Liu, X.S., *et al.* (2006). Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. Blood *108*, 986-992.

Parry, D.H., Xu, J., and Ruvkun, G. (2007). A whole-genome RNAi Screen for C. elegans miRNA pathway genes. Curr Biol *17*, 2013-2022.
Pebernard, S., and Iggo, R.D. (2004). Determinants of interferon-stimulated gene induction by RNAi vectors. Differentiation; research in biological diversity *72*, 103-111.
Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D.*, et al.* (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499-502.
Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H., and Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. Development (Cambridge, England) *121*, 163-172.
Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257-260.
Pokholok, D.K., Hannett, N.M., and Young, R.A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Molecular cell *9*, 799-809.
Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517-527.
Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell *86*, 47-57.
Privalsky, M.L. (2004). The role of corepressors in transcriptional regulation by nuclear hormone receptors. Annual review of physiology *66*, 315-360.
Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.
Pulford, K., Lecointe, N., Leroy-Viard, K., Jones, M., Mathieu-Mahul, D., and Mason, D.Y. (1995). Expression of TAL-1 proteins in human tissues. Blood *85*, 675-684.
Qiu, Y., Sharma, A., and Stein, R. (1998). p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. Mol Cell Biol *18*, 2957-2964.
Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D.*, et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.
Rekhtman, N., Radparvar, F., Evans, T., and Skoultchi, A.I. (1999). Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. Genes Dev *13*, 1398-1411.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev *16*, 245-256.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E.*, et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Reynaud, D., Ravet, E., Titeux, M., Mazurier, F., Renia, L., Dubart-Kupperschmitt, A., Romeo, P.H., and Pflumio, F. (2005). SCL/TAL1 expression level regulates human hematopoietic stem cell self-renewal and engraftment. Blood *106*, 2318-2328.

Rippe, K., von Hippel, P.H., and Langowski, J. (1995). Action at a distance: DNA-looping and initiation of transcription. Trends Biochem Sci *20*, 500-506.

Robb, L., Elwood, N.J., Elefanty, A.G., Kontgen, F., Li, R., Barnett, L.D., and Begley, C.G. (1996). The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. Embo J *15*, 4123-4129.

Robb, L., Lyons, I., Li, R., Hartley, L., Kontgen, F., Harvey, R.P., Metcalf, D., and Begley, C.G. (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.

Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K., and Young, R.A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. Molecular cell *16*, 199-209.

Robertson, S.M., Kennedy, M., Shannon, J.M., and Keller, G. (2000). A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor SCL/tal-1. Development *127*, 2447-2459.

Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S.K., Wang, A., Suka, N., and Grunstein, M. (2002). Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. Cell *109*, 437-446.

Robzyk, K., Recht, J., and Osley, M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science *287*, 501-504.

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. (1998). Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nature biotechnology *16*, 939-945.

Rylski, M., Welch, J.J., Chen, Y.Y., Letting, D.L., Diehl, J.A., Chodosh, L.A., Blobel, G.A., and Weiss, M.J. (2003). GATA-1-mediated proliferation arrest during erythroid maturation. Mol Cell Biol *23*, 5031-5042.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nature reviews *7*, 437-447.
Salghetti, S.E., Caudy, A.A., Chenoweth, J.G., and Tansey, W.P. (2001). Regulation of transcriptional activation domain function by ubiquitin. Science *293*, 1651-1653.
Sanchez, M., Gottgens, B., Sinclair, A.M., Stanley, M., Begley, C.G., Hunter, S., and Green, A.R. (1999). An SCL 3' enhancer targets developing endothelium together with embryonic and adult haematopoietic progenitors. Development *126*, 3891-3904.
Sanchez, M.J., Bockamp, E.O., Miller, J., Gambardella, L., and Green, A.R. (2001). Selective rescue of early haematopoietic progenitors in Scl(-/-) mice by expressing Scl under the control of a stem cell enhancer. Development *128*, 4815-4827.
Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res *32*, D91-94.
Sassone-Corsi, P., Mizzen, C.A., Cheung, P., Crosio, C., Monaco, L., Jacquot, S., Hanauer, A., and Allis, C.D. (1999). Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. Science *285*, 886-891.
Scacheri, P.C., Rozenblatt-Rosen, O., Caplen, N.J., Wolfsberg, T.G., Umayam, L., Lee, J.C., Hughes, C.M., Shanmugam, K.S., Bhattacharjee, A., Meyerson, M., *et al.* (2004). Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 1892-1897.
Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (New York, NY *270*, 467-470.
Scherf, M., Klingenhoff, A., and Werner, T. (2000). Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. J Mol Biol *297*, 599-606.
Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R., Kempe, D., Lottspeich, F., Sanger, H.L., and Wassenegger, M. (1998). Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato. The Plant cell *10*, 2087-2101.
Schlabach, M.R., Luo, J., Solimini, N.L., Hu, G., Xu, Q., Li, M.Z., Zhao, Z., Smogorzewska, A., Sowa, M.E., Ang, X.L., *et al.* (2008). Cancer proliferation gene discovery through functional genomics. Science (New York, NY *319*, 620-624.
Schmitter, D., Filkowski, J., Sewer, A., Pillai, R.S., Oakeley, E.J., Zavolan, M., Svoboda, P., and Filipowicz, W. (2006). Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. Nucleic acids research *34*, 4801-4815.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.
Schug, J., Overton, G.C. (1997). TESS: Transcriptional Element Search Software on the WWW. Technical Report CBIL-TR-1997-1001-v0.0. . Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania.
Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet *34*, 166-176.
Sen, G., Wehrman, T.S., Myers, J.W., and Blau, H.M. (2004). Restriction enzyme-generated siRNA (REGS) vectors and libraries. Nature genetics *36*, 183-189.
Sertil, O., Kapoor, R., Cohen, B.D., Abramova, N., and Lowry, C.V. (2003). Synergistic repression of anaerobic genes by Mot3 and Rox1 in Saccharomyces cerevisiae. Nucleic Acids Res *31*, 5831-5837.
Shabalina, S.A., and Spiridonov, N.A. (2004). The mammalian transcriptome and the function of non-coding DNA sequences. Genome Biol *5*, 105.
Shannon, M.F., and Rao, S. (2002). Transcription. Of chips and ChIPs. Science *296*, 666-669.
Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet *31*, 64-68.
Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.
Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell *19*, 857-864.
Shimizu, R., Trainor, C.D., Nishikawa, K., Kobayashi, M., Ohneda, K., and Yamamoto, M. (2007). GATA-1 self-association controls erythroid development in vivo. J Biol Chem *282*, 15862-15871.
Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S.*, et al.* (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med *8*, 68-74.
Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T.*, et al.* (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A *100*, 15776-15781.
Shivdasani, R.A., Fujiwara, Y., McDevitt, M.A., and Orkin, S.H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. Embo J *16*, 3965-3973.

Shivdasani, R.A., Mayer, E.L., and Orkin, S.H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature *373*, 432-434.
Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., *et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.
Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell *107*, 465-476.
Silberstein, L., Sanchez, M.J., Socolovsky, M., Liu, Y., Hoffman, G., Kinston, S., Piltz, S., Bowen, M., Gambardella, L., Green, A.R., *et al.* (2005). Transgenic analysis of the stem cell leukemia +19 stem cell enhancer in adult and embryonic hematopoietic and endothelial cells. Stem Cells *23*, 1378-1388.
Silva, J.M., Li, M.Z., Chang, K., Ge, W., Golding, M.C., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R., *et al.* (2005). Second-generation shRNA libraries covering the mouse and human genomes. Nature genetics *37*, 1281-1288.
Silva, J.M., Marran, K., Parker, J.S., Silva, J., Golding, M., Schlabach, M.R., Elledge, S.J., Hannon, G.J., and Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. Science (New York, NY *319*, 617-620.
Silver, L., and Palis, J. (1997). Initiation of murine embryonic erythropoiesis: a spatial analysis. Blood *89*, 1154-1164.
Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat Biotechnol *17*, 974-978.
Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H., and Williams, B.R. (2003). Activation of the interferon system by short-interfering RNAs. Nature cell biology *5*, 834-839.
Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem *72*, 449-479.
Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr Biol *10*, 169-178.
Smith, E.M., Gisler, R., and Sigvardsson, M. (2002). Cloning and characterization of a promoter flanking the early B cell factor (EBF) gene indicates roles for E-proteins and autoregulation in the control of EBF expression. J Immunol *169*, 261-270.
Song, E., Zhu, P., Lee, S.K., Chowdhury, D., Kussman, S., Dykxhoorn, D.M., Feng, Y., Palliser, D., Weiner, D.B., Shankar, P., *et al.* (2005). Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. Nat Biotechnol *23*, 709-717.

Song, S.H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Molecular cell *28*, 810-822.
Sonnichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Bettencourt, P., Cassin, E., *et al.* (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature *434*, 462-469.
Spicuglia, S., Kumar, S., Yeh, J.H., Vachez, E., Chasson, L., Gorbatch, S., Cautres, J., and Ferrier, P. (2002). Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. Mol Cell *10*, 1479-1487.
Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature *435*, 637-645.
Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., *et al.* (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA (New York, NY *9*, 493-501.
Stielow, B., Sapetschnig, A., Kruger, I., Kunert, N., Brehm, A., Boutros, M., and Suske, G. (2008). Identification of SUMO-dependent chromatin-associated transcriptional repression components by a genome-wide RNAi screen. Mol Cell *29*, 742-754.
Stopka, T., Amanatullah, D.F., Papetti, M., and Skoultchi, A.I. (2005). PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. Embo J *24*, 3712-3723.
Studitsky, V.M., Clark, D.J., and Felsenfeld, G. (1994). A histone octamer can step around a transcribing polymerase without leaving the template. Cell *76*, 371-382.
Sun, X.H., and Baltimore, D. (1991). An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell *64*, 459-470.
Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., *et al.* (2001). Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome research *11*, 677-684.
Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Dev Biol *294*, 525-540.
Szekvolgyi, L., Balint, B.L., Imre, L., Goda, K., Szabo, M., Nagy, L., and Szabo, G. (2006). Chip-on-beads: flow-cytometric evaluation of chromatin immunoprecipitation. Cytometry A *69*, 1086-1091.
Tabara, H., Grishok, A., and Mello, C.C. (1998). RNAi in C. elegans: soaking in the genome sequence. Science (New York, NY *282*, 430-431.
Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. Nat Genet *22*, 281-285.

Tavernarakis, N., Wang, S.L., Dorovkov, M., Ryazanov, A., and Driscoll, M. (2000). Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. Nature genetics *24*, 180-183.

Thomson, S., Clayton, A.L., Hazzalin, C.A., Rose, S., Barratt, M.J., and Mahadevan, L.C. (1999). The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. Embo J *18*, 4779-4793.

Thut, C.J., Chen, J.L., Klemm, R., and Tjian, R. (1995). p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. Science *267*, 100-104.

Timmons, L., Court, D.L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. Gene *263*, 103-112.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854.

Torres, T.T., Metta, M., Ottenwalder, B., and Schlotterer, C. (2008). Gene expression profiling by massively parallel sequencing. Genome Res *18*, 172-177.

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.

Tsai, S.F., Martin, D.I., Zon, L.I., D'Andrea, A.D., Wong, G.G., and Orkin, S.H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature *339*, 446-451.

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

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.

Utley, R.T., Ikeda, K., Grant, P.A., Cote, J., Steger, D.J., Eberharter, A., John, S., and Workman, J.L. (1998). Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. Nature *394*, 498-502.

Valverde-Garduno, V., Guyot, B., Anguita, E., Hamlett, I., Porcher, C., and Vyas, P. (2004). Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification. Blood *104*, 3106-3116.

van de Wetering, M., Oving, I., Muncan, V., Pon Fong, M.T., Brantjes, H., van Leenen, D., Holstege, F.C., Brummelkamp, T.R., Agami, R., and Clevers, H. (2003). Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. EMBO reports *4*, 609-615.

van Eekelen, J.A., Bradley, C.K., Gothert, J.R., Robb, L., Elefanty, A.G., Begley, C.G., and Harvey, A.R. (2003). Expression pattern of the stem cell leukaemia gene in the CNS of the embryonic and adult mouse. Neuroscience *122*, 421-436.
van Haaften, G., Romeijn, R., Pothof, J., Koole, W., Mullenders, L.H., Pastink, A., Plasterk, R.H., and Tijsterman, M. (2006). Identification of conserved pathways of DNA-damage response and radiation protection by genome-wide RNAi. Curr Biol *16*, 1344-1350.
van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nature biotechnology *18*, 424-428.
Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science (New York, NY *270*, 484-487.
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.
Vilar, J.M., and Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Current opinion in genetics & development *15*, 136-144.
Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res *35*, D88-92.
Visvader, J., Begley, C.G., and Adams, J.M. (1991). Differential expression of the LYL, SCL and E2A helix-loop-helix genes within the hemopoietic system. Oncogene *6*, 187-194.
Visvader, J.E., Fujiwara, Y., and Orkin, S.H. (1998). Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development. Genes Dev *12*, 473-479.
Visvader, J.E., Mao, X., Fujiwara, Y., Hahm, K., and Orkin, S.H. (1997). The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation. Proceedings of the National Academy of Sciences of the United States of America *94*, 13707-13712.
Vitelli, L., Condorelli, G., Lulli, V., Hoang, T., Luchetti, L., Croce, C.M., and Peschle, C. (2000). A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. Mol Cell Biol *20*, 5330-5342.
Vogel, M.J., Guelen, L., de Wit, E., Peric-Hupkes, D., Loden, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.K., and van Steensel, B. (2006). Human heterochromatin proteins form large domains containing KRAB-ZNF genes. Genome Res *16*, 1493-1504.
Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nature protocols *2*, 1467-1478.

Vokes, S.A., Ji, H., McCuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh, W.J., Davidson, E.H., Wong, W.H., and McMahon, A.P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development *134*, 1977-1989.

Vyas, P., Ault, K., Jackson, C.W., Orkin, S.H., and Shivdasani, R.A. (1999a). Consequences of GATA-1 deficiency in megakaryocytes and platelets. Blood *93*, 2867-2875.

Vyas, P., McDevitt, M.A., Cantor, A.B., Katz, S.G., Fujiwara, Y., and Orkin, S.H. (1999b). Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. Development (Cambridge, England) *126*, 2799-2811.

Wadman, I., Li, J., Bash, R.O., Forster, A., Osada, H., Rabbitts, T.H., and Baer, R. (1994). Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. Embo J *13*, 4831-4839.

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

Wakabayashi, J., Yomogida, K., Nakajima, O., Yoh, K., Takahashi, S., Engel, J.D., Ohneda, K., and Yamamoto, M. (2003). GATA-1 testis activation region is essential for Sertoli cell-specific expression of GATA-1 gene in transgenic mouse. Genes Cells *8*, 619-630.

Wall, L., deBoer, E., and Grosveld, F. (1988). The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes Dev *2*, 1089-1100.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.H., Perlin, J.R., Leonelli, L., Sonbuchner, L.S., McDonald, C.H., Cook, R.G., Dou, Y.*, et al.* (2004b). Human PAD4 regulates histone arginine methylation levels via demethylimination. Science *306*, 279-283.

Warren, A.J., Colledge, W.H., Carlton, M.B., Evans, M.J., Smith, A.J., and Rabbitts, T.H. (1994). The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. Cell *78*, 45-57.

Waterhouse, P.M., Wang, M.B., and Lough, T. (2001). Gene silencing as an adaptive defence against viruses. Nature *411*, 834-842.

Weber, F., de Villiers, J., and Schaffner, W. (1984). An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell *36*, 983-992.

Wechsler, J., Greene, M., McDevitt, M.A., Anastasi, J., Karp, J.E., Le Beau, M.M., and Crispino, J.D. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nature genetics *32*, 148-152.

Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., *et al.* (2006). A global map of p53 transcription-factor binding sites in the human genome. Cell *124*, 207-219.

Weinberg, M.S., Villeneuve, L.M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z.X., Riggs, A.D., Rossi, J.J., and Morris, K.V. (2006). The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. RNA (New York, NY *12*, 256-262.

Weinmann, A.S., Bartley, S.M., Zhang, T., Zhang, M.Q., and Farnham, P.J. (2001). Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol Cell Biol *21*, 6820-6832.

Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H., and Farnham, P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev *16*, 235-244.

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

Weiss, M.J., and Orkin, S.H. (1995). Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. Proc Natl Acad Sci U S A *92*, 9623-9627.

Welch, J.J., Watts, J.A., Vakoc, C.R., Yao, Y., Wang, H., Hardison, R.C., Blobel, G.A., Chodosh, L.A., and Weiss, M.J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. Blood *104*, 3136-3147.

Wells, J., Yan, P.S., Cechvala, M., Huang, T., and Farnham, P.J. (2003). Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. Oncogene *22*, 1445-1460.

West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. Genes Dev *16*, 271-288.

Westbrook, T.F., Martin, E.S., Schlabach, M.R., Leng, Y., Liang, A.C., Feng, B., Zhao, J.J., Roberts, T.M., Mandel, G., Hannon, G.J., *et al.* (2005). A genetic screen for candidate tumor suppressors identifies REST. Cell *121*, 837-848.

White, K.P., Rifkin, S.A., Hurban, P., and Hogness, D.S. (1999). Microarray analysis of Drosophila development during metamorphosis. Science (New York, NY *286*, 2179-2184.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Williams, B.R. (1997). Role of the double-stranded RNA-activated protein kinase (PKR) in cell regulation. Biochem Soc Trans *25*, 509-513.
Williams, R.R., Azuara, V., Perry, P., Sauer, S., Dvorkina, M., Jorgensen, H., Roix, J., McQueen, P., Misteli, T., Merkenschlager, M.*, et al.* (2006). Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. J Cell Sci *119*, 132-140.
Wolfrum, C., Shi, S., Jayaprakash, K.N., Jayaraman, M., Wang, G., Pandey, R.K., Rajeev, K.G., Nakayama, T., Charrise, K., Ndungo, E.M.*, et al.* (2007). Mechanisms and optimization of in vivo delivery of lipophilic siRNAs. Nat Biotechnol *25*, 1149-1157.
Wolstein, O., Silkov, A., Revach, M., and Dikstein, R. (2000). Specific interaction of TAFII105 with OCA-B is involved in activation of octamer-dependent transcription. J Biol Chem *275*, 16459-16465.
Wong, J.M., and Bateman, E. (1994). TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. Nucleic Acids Res *22*, 1890-1896.
Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F.*, et al.* (2003). Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol Cell *11*, 267-274.
Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K.*, et al.* (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS biology *3*, e7.
Wright, W.E., Binder, M., and Funk, W. (1991). Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol Cell Biol *11*, 4104-4110.
Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357-2360.
Xu, Z., Huang, S., Chang, L.S., Agulnick, A.D., and Brandt, S.J. (2003). Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. Mol Cell Biol *23*, 7585-7599.
Xu, Z., Meng, X., Cai, Y., Koury, M.J., and Brandt, S.J. (2006). Recruitment of the SWI/SNF protein Brg1 by a multiprotein complex effects transcriptional repression in murine erythroid progenitors. The Biochemical journal *399*, 297-304.
Yamada, Y., Pannell, R., Forster, A., and Rabbitts, T.H. (2000). The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice. Proceedings of the National Academy of Sciences of the United States of America *97*, 320-324.

Yamada, Y., Warren, A.J., Dobson, C., Forster, A., Pannell, R., and Rabbitts, T.H. (1998). The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proceedings of the National Academy of Sciences of the United States of America *95*, 3890-3895.
Yamaguchi, Y., Ackerman, S.J., Minegishi, N., Takiguchi, M., Yamamoto, M., and Suda, T. (1998). Mechanisms of transcription in eosinophils: GATA-1, but not GATA-2, transactivates the promoter of the eosinophil granule major basic protein gene. Blood *91*, 3447-3458.
Yamaguchi, Y., Nishio, H., Kishi, K., Ackerman, S.J., and Suda, T. (1999). C/EBPbeta and GATA-1 synergistically regulate activity of the eosinophil granule major basic protein promoter: implication for C/EBPbeta activity in eosinophil gene expression. Blood *94*, 1429-1439.
Yang, D., Buchholz, F., Huang, Z., Goga, A., Chen, C.Y., Brodsky, F.M., and Bishop, J.M. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *99*, 9942-9947.
Yoder, M.C., Hiatt, K., and Mukherjee, P. (1997). In vivo repopulating hematopoietic stem cells are present in the murine yolk sac at day 9.0 postcoitus. Proc Natl Acad Sci U S A *94*, 6776-6780.
Yokomizo, T., Takahashi, S., Mochizuki, N., Kuroha, T., Ema, M., Wakamatsu, A., Shimizu, R., Ohneda, O., Osato, M., Okada, H., *et al.* (2007). Characterization of GATA-1(+) hemangioblastic cells in the mouse embryo. Embo J *26*, 184-196.
Yomogida, K., Ohtani, H., Harigae, H., Ito, E., Nishimune, Y., Engel, J.D., and Yamamoto, M. (1994). Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells. Development (Cambridge, England) *120*, 1759-1766.
Yoo, J.Y., Kim, J.H., Kim, J., Huang, J.H., Zhang, S.N., Kang, Y.A., Kim, H., and Yun, C.O. (2008). Short hairpin RNA-expressing oncolytic adenovirus-mediated inhibition of IL-8: effects on antiangiogenesis and tumor growth inhibition. Gene therapy *15*, 635-651.
Yu, C., Cantor, A.B., Yang, H., Browne, C., Wells, R.A., Fujiwara, Y., and Orkin, S.H. (2002a). Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. J Exp Med *195*, 1387-1395.
Yu, C., Niakan, K.K., Matsushita, M., Stamatoyannopoulos, G., Orkin, S.H., and Raskind, W.H. (2002b). X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. Blood *100*, 2040-2045.
Zeller, K.I., Zhao, X., Lee, C.W., Chiu, K.P., Yao, F., Yustein, J.T., Ooi, H.S., Orlov, Y.L., Shahab, A., Yong, H.C., *et al.* (2006). Global mapping of c-Myc binding sites and target gene networks in human B cells. Proc Natl Acad Sci U S A *103*, 17834-17839.

Zeng, Y., Wagner, E.J., and Cullen, B.R. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. Mol Cell *9*, 1327-1333.
Zenzie-Gregory, B., Khachi, A., Garraway, I.P., and Smale, S.T. (1993). Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. Mol Cell Biol *13*, 3841-3849.
Zhang, P., Zhang, X., Iwama, A., Yu, C., Smith, K.A., Mueller, B.U., Narravula, S., Torbett, B.E., Orkin, S.H., and Tenen, D.G. (2000). PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood *96*, 2641-2648.
Zhang, W., Yang, H., Kong, X., Mohapatra, S., San Juan-Vergara, H., Hellermann, G., Behera, S., Singam, R., Lockey, R.F., and Mohapatra, S.S. (2005a). Inhibition of respiratory syncytial virus infection with intranasal siRNA nanoparticles targeting the viral NS1 gene. Nature medicine *11*, 56-62.
Zhang, Y., Payne, K.J., Zhu, Y., Price, M.A., Parrish, Y.K., Zielinska, E., Barsky, L.W., and Crooks, G.M. (2005b). SCL expression at critical points in human hematopoietic lineage commitment. Stem Cells *23*, 852-860.
Zhao, H., and Dean, A. (2004). An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res *32*, 4903-4919.
Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., *et al.* (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nature genetics *38*, 1341-1347.
Zhu, B., Zheng, Y., Pham, A.D., Mandal, S.S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005). Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell *20*, 601-611.
Zhu, P., Zhou, W., Wang, J., Puc, J., Ohgi, K.A., Erdjument-Bromage, H., Tempst, P., Glass, C.K., and Rosenfeld, M.G. (2007). A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. Mol Cell *27*, 609-621.
Zhuang, Y., Jackson, A., Pan, L., Shen, K., and Dai, M. (2004). Regulation of E2A gene expression in B-lymphocyte development. Molecular immunology *40*, 1165-1177.
Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. Cell *79*, 875-884.
Zimmermann, T.S., Lee, A.C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M.N., Harborth, J., Heyes, J.A., Jeffs, L.B., John, M., *et al.* (2006). RNAi-mediated gene silencing in non-human primates. Nature *441*, 111-114.

Zon, L.I., Tsai, S.F., Burgess, S., Matsudaira, P., Bruns, G.A., and Orkin, S.H. (1990). The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. Proc Natl Acad Sci U S A *87*, 668-672.

Zon, L.I., Yamaguchi, Y., Yee, K., Albee, E.A., Kimura, A., Bennett, J.C., Orkin, S.H., and Ackerman, S.J. (1993). Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription. Blood *81*, 3234-3241.

Zon, L.I., Youssoufian, H., Mather, C., Lodish, H.F., and Orkin, S.H. (1991). Activation of the erythropoietin receptor promoter by transcription factor GATA-1. Proc Natl Acad Sci U S A *88*, 10638-10641.