Chapter 6

Further characterisation of putative target genes of members of the SCL erythroid complex

6.1 Introduction

The previous Chapters described two experimental approaches, expression profiling of siRNA-induced knockdown and ChIP-on-chip for identifying transcriptional targets of the SCL erythroid complex. Both methods provided different types of information and are complimentary to each other. The gene expression profiling studies provided data on the changes in expression of downstream target genes while the ChIP-on-chip analysis determine the transcription factor binding to *cis*-regulatory elements. Thus, both studies facilitate a better understanding of gene expression controlled by the SCL erythroid complex.

6.1.1 Expression profiling of siRNA-induced knockdown of the SCL erythroid complex

The siRNA-induced knockdown of members of the SCL erythroid complex in combination with expression analyses on a genome-wide scale provided a platform for the study of regulation of target genes. Four main caveats from this study should be re-iterated here:

1. Both direct and indirect (downstream) target genes were identified when each of the five TFs were knocked down using siRNA. Which of these are direct or indirect targets cannot be distinguished purely by expression analysis alone (unless a rigorous time-course Affymetrix GeneChip analysis was performed).
2. The efficiency of knockdown using siRNA has a great impact on the results obtained in the subsequent expression profiling. Ultimately the outcomes of changes of expression are likely to rely on how the removal of a substantial amount of the TF in the nucleus influences its binding to the *cis*-regulatory elements.
3. The overall mode of regulation of target genes can be inferred from Affymetrix expression analysis. Based on changes in expression, it is possible to determine which genes are activated or repressed by the TF. However, complex regulation in transcriptional networks (feedback loops, auto-regulation etc.) may mean the overall effect on expression does not relate directly to the effect that an individual TF is having.

4. Expression profiling alone does not provide information on the regulatory elements and all the protein components required for changes in gene expression to be induced.

6.1.2 ChIP-on-chip study of the SCL erythroid complex

The ChIP-on-chip analyses described in Chapter 5 using antibodies against each TF allowed us to study the interaction between transcription factors and their DNA interacting partners *in vivo*. Unlike expression analysis with Affymetrix GeneChips, it provides direct evidence on the DNA-protein interactions between a TF and its target genes. Three main caveats from this study should be re-iterated here:

1. ChIP-on-chip assays provide no information on whether TF binding events actually activate or repress the transcription of genes and how they induce such regulation.
2. As the ChIP assays can identify any DNA sequences in close proximity to the protein under study, it is not possible to tell whether the TF interacts with the DNA directly, through other proteins or protein complexes in co-operative interactions, or via DNA looping.
3. The array used in Chapter 5 was a transcription factor promoter array. Therefore, the data obtained from these ChIP-on-chip studies can only identify a small portion of genes regulated through their promoters but not mediated through other regulatory elements (unless those elements are in close contact with the promoters).

6.1.3 Auto-regulation of the SCL erythroid complex

From the results obtained in Chapter 4 and 5, three different levels of auto-regulatory pattern of members of the SCL erythroid complex were observed. At the first level, a transcription factor was found to bind to its own promoter and/or enhancer. This was observed for SCL in the ChIP-on-chip study where it bound to its own enhancer and promoter. GATA1 has also been shown to bind to the G1HE, the double GATA site in promoter IE and the intronic enhancer intron-SP *in vivo* (Valverde-Garduno et al., 2004). At the second level of auto-regulation, the whole SCL erythroid complex directly regulates expression of its members. SCL and ETO2, both of which are members of the SCL erythroid complex, were shown to be bound by the whole SCL erythroid complex. At the third level, one member regulated another. From the gene expression profiling study in Chapter 4, E2A was found to be activated by both GATA1 and SCL, LMO2 was activated by E2A and GATA1, and LDB1 was activated by GATA1 (note that such expression changes may be direct or indirect). Furthermore, based on ChIP-on-chip and ChIP-qPCR, GATA1 binds to the promoter of LMO2. These three levels of auto-regulation ensure multiple levels of control over the expression of each member of the complex, and thus they tightly control the expression of their own genes.

Understanding the complexities of this auto-regulation is essential for understanding the transcriptional cascades controlled by the SCL erythroid complex.

6.1.4 Regulation of the LYL1 gene

The ChIP-on-chip analysis identified LYL1 as a direct target gene of the whole SCL erythroid complex. LYL1 is a bHLH protein and has overlapping expression pattern in the erythroid and myeloid lineages and in ascular tissues with SCL in mouse (Visvader et al., 1991). Its expression is initiated slightly later than SCL during haematopoietic specification, beginning during haemangioblast differentiation (Chan et al., 2007). Therefore, LYL1 is thought to be a functional paralogue of SCL (Chapter 1, section 1.4.2.1).

The promoters of LYL1 and SCL have similar structure with two GATA sites located in close proximity. However, no E-box/GATA composite motifs were found in their promoters. SCL, however, has a canonical E-box/GATA composite motif in its +51 enhancer sequence – suggesting that the +51 region may mediate the binding of the SCL erythroid complex. Regulation of the gene may therefore be achieved through the interaction of +51 with its cognate promoter – thus facilitated the detection of the erythroid complex on the promoter using ChIP-on-chip. Considering the coordinated expression pattern of SCL and LYL1, LYL1 is therefore an interesting candidate for studying off-promoter binding of the SCL erythroid complex.

6.1.5 Generation of transcription networks

Ultimately, a more complete understanding of the role of the SCL erythroid complex and its target genes would come from an integration of the datasets presented in this thesis. This would result in a transcriptional regulatory network with the SEC at its core. To aid in the integration of these datasets, a number of computational programmes have been developed for building and visualising gene regulation and expression patterns. These include CellDesigner, Cytoscape and BioTapestry (Longabaugh et al., 2005; Oda et al., 2004; Shannon et al., 2003). All of these softwares simplify the representation of transcription regulatory networks using graphical interfaces and are open source packages freely available to the scientific community. CellDesigner (www.CellDesigner.org) was first described in the modelling of a comprehensive molecular interaction pattern in macrophage based on data found in the published literature (Oda et al., 2004). This software provides a wide range of interaction symbols for the drawing of regulatory networks. It also supports representation of multiple compartments such as the cell membrane and the nucleus. However, in this software, genes are treated as other ‘biomolecules’ (a CellDesigner terminology) and cannot be distinguished from each other. Cytoscape (http://www.cytoscape.org/) is another software for simulating protein-DNA, protein-protein and genetic networks (Shannon et al., 2003).

It allows the end-user to integrate interaction networks with expression data. Users can also customise the properties of interaction symbols. Like CellDesigner, however, DNA sequences are not presented explicitly in Cytoscape. This is a drawback, given that a computational representation that specifically describes transcription factor interactions with the *cis*-regulatory DNA elements is required for modelling transcription networks (Figure 6.1). BioTapestry (www.biotapestry.org) was developed to support this kind of representation of gene regulation (Longabaugh et al., 2005). BioTapestry also allows users to include time-course expression data in an interactive interface in addition to other features that BioTapestry, CellDesigner and Cytoscape all provide.

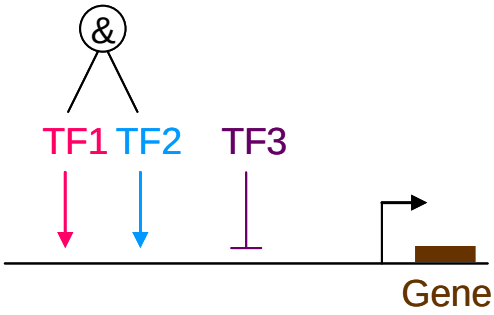


Figure 6.1. A computational representation of interactions between transcription factors and a gene. The black horizontal line represents DNA. The black arrow indicates the transcription start site and the DNA to the left of the arrow represents upstream sequences while the brown box, to the right, represents a gene. TFs 1, 2 and 3 are transcription factors binding to upstream regulatory elements. Coloured arrows indicate activating activities while purple blunt arrow indicated repressing activity. The & symbol shows that TF1 and 2 are both required for activation.

Previous studies have demonstrated the integration of various experimental data for the generation of regulatory networks in haematopoietic development (Chapter 1, section 1.4.2.6) (Swiers et al., 2006). However, the studies performed were based on mouse and a comprehensive study of the SCL erythroid complex was not described. The data obtained in this thesis allowed us to have a more in-depth understanding of the transcription regulatory network of this complex in human haematopoiesis.

6.2 Aims of this chapter

The results obtained in Chapter 4 and Chapter 5 generated different datasets describing different means by which to identify target genes of the SCL erythroid complex. These datasets were compared, evaluated and discussed. Additional experiments were also performed to explain the results obtained from the previous Chapters and further characterise putative target genes of the SCL erythroid complex.

The aims of the work presented in this Chapter were:

1. To compare the putative target genes identified in the expression analyses using Affymetrix GeneChips with the targets identified with ChIP-on-chip and ChIP-qPCR analysis.
2. To study the expression of putative target genes identified in ChIP-on-chip during perturbations of members of the SCL erythroid complex in siRNA-mediated knockdown time-course studies.
3. To further examine aspects of auto-regulation of the SCL erythroid complex observed from the datasets from Chapters 4 and 5. This will be accomplished by looking at the expression level of members of the SCL eythroid complex during perturbations of each members of the SCL erythroid complex in siRNA-mediated knockdown time-course studies.
4. To observe the kinetics of TF binding under the conditions used for Affymetrix analysis. This will be done by characterising the effect of siRNA-mediated GATA1 knockdown on the binding efficiency of GATA1 to promoters.
5. To study the off-promoter binding and regulation of SCL and LYL1 by various TFs in the SCL erythroid complex.
6. To generate a transcriptional regulatory network during erythroid development based on the data obtained in Chapters 4 and 5 and in this Chapter.

6.3 Results

6.3.1 Further characterisation of putative target genes identified in Affymetrix and ChIP-on-chip studies

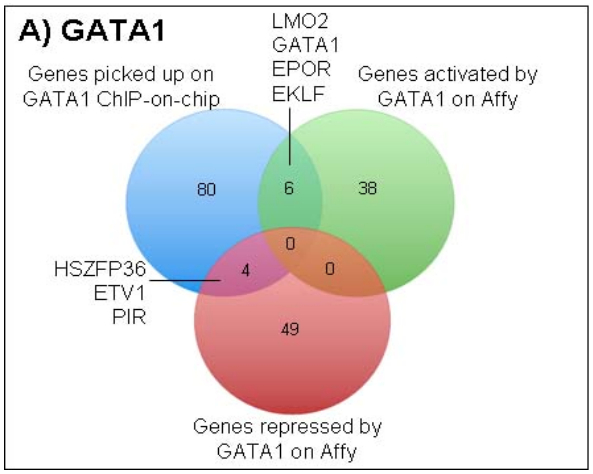
6.3.1.1 Comparison between putative target genes of Affymetrix expression and ChIP-on-chip studies

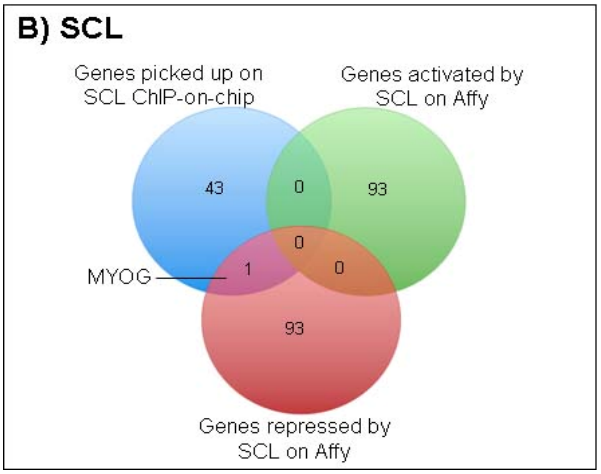
The putative direct target genes of the SCL erythroid complex which were identified in the ChIP-on-chip studies and those validated by ChIP-qPCR in Chapter 5 were compared with the activated or repressed gene lists obtained in the Affymetrix expression studies in Chapter 4. Only the targets for a given TF which were represented on both the promoter array and on the Affymetrix array were used in the comparisons. To increase the level of overlap between the lists, the promoter targets from the ChIP-on-chip study analysed by method B (which included all targets identified by method A plus additional ones) were used in the comparisons. Targets identified by both ChIP-on-chip or ChIP-qPCR, and which showed changes in Affymetrix expression analysis of siRNA-mediated TF knockdowns, were likely to be *bona fide* direct targets for the relevant TF.

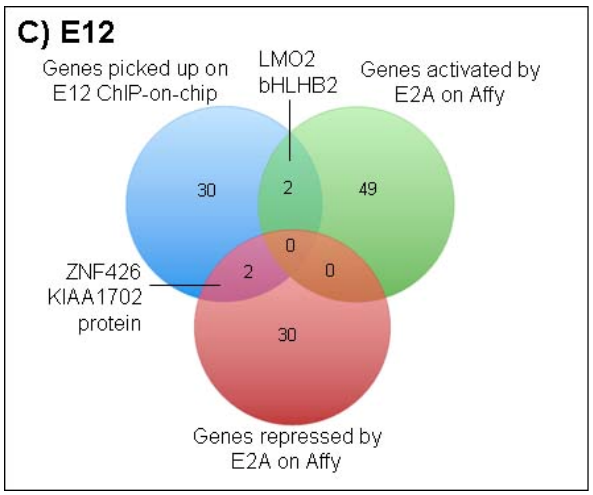
Only a small number of genes were found to be identified in both the Affymetrix expression studies and the ChIP-on-chip studies. In the case of GATA1, 10 target genes were found in both analyses,

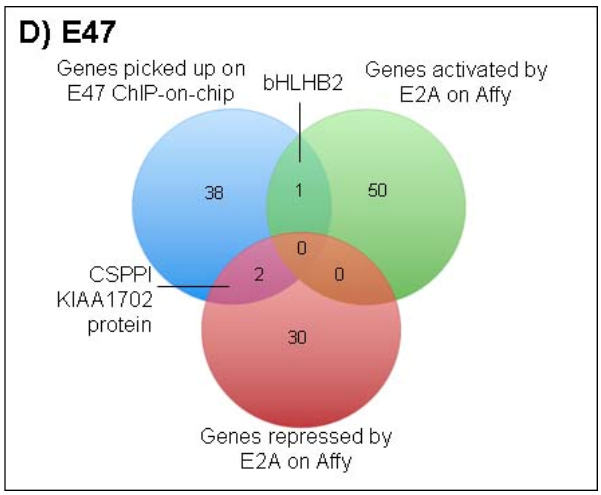
from a total of 90 possible ChIP-on-chip targets and 97 possible Affymetrix targets (Figure 6.2 A). Activation of EKLF by GATA1 was also confirmed in an expression time-course study of siRNA-mediated GATA1 knockdown in K562 cells (Section 6.3.1.2). Only one target gene of SCL (MYOG) was confirmed in both Affymetrix and ChIP-on-chip analyses (Figure 6.2 B). In the case of E2A, ChIP-on-chip targets for either variant E12 or E47 were compared separately against the Affymetrix expression analyses of knockdown by the E2A siRNAs (Figure 6.2 C and D). In total, five targets were identified by both ChIP-on-chip and Affymetrix analysis, out of a total of 34 (E12) or 41 (E47) ChIP-on-chip targets and 81 Affymetrix targets. Two of these targets, bHLHB2 and KIAA1702, were found to be common to both E12 and E47 as ChIP-on-chip targets. No genes were identified as overlapping genes in the ChIP study and the Affymetrix analysis for LDB1 and LMO2 (Figure 6.2 E and F), further suggesting that both of these proteins may be dispensable from the SCL erythroid complex without having consequences at the level of expression (section 4.5.1). The genes co-regulated by SCL, GATA1 and E2A and present on the promoter array were also compared with the putative targets identified in the ChIP-on-chip by SCL, GATA1 or E2A (Figure 6.2 G). No genes were found to be overlapping between these two categories.

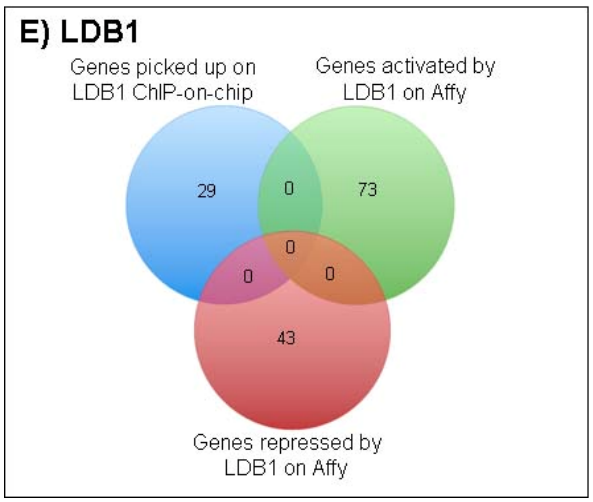
A second way of comparing the data was made by looking at genes identified in any of the Affymetrix GeneChip knockdown experiments with target promoters identified in any of the ChIP-on-chip assays (Figure 6.2 H). This would allow for genes which may have been identified with a particular ChIP-on-chip assay, but not with its corresponding Affymetrix experiment and vise versa, to be identified. Using method B of ChIP-on-chip analyses, 37 target genes were found in both analyses, from a total of 196 possible ChIP-on-chip targets and 331 possible Affymetrix targets. This number of overlapping genes is significantly higher than that of total number of overlapping genes when the five transcription factors were investigated independently (18 target genes). With either analysis, this would suggest that only approximately 5-11% of *bona fide* target genes of members of the SCL erythroid complex, actually change in measureable levels of expression during knockdown. Furthermore, this would argue that the less stringent method B ChIP-on-chip analysis may allow for more direct targets to be identified.

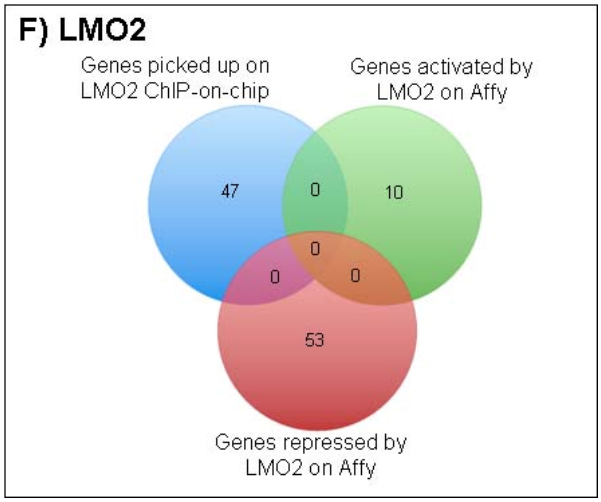

A) GATA1
LMO2
GATA1
EPOR
EKLF
Genes picked up on
GATA1 ChIP-on-chip
Genes activated by
GATA1 on Affy
80
6
38
0
HSZFP36
ETV1
PIR
4
0
49
Genes repressed by
GATA1 on Affy


B) SCL
Genes picked up on
SCL ChIP-on-chip
Genes activated by
SCL on Affy
43
0
93
0
MYOG
1
0
93
Genes repressed by
SCL on Affy


C) E12
Genes picked up on
E12 ChIP-on-chip
LMO2
bHLHB2
Genes activated by
E2A on Affy
30
2
49
0
ZNF426
KIAA1702
protein
2
0
30
Genes repressed by
E2A on Affy


D) E47
Genes picked up on
E47 ChIP-on-chip
bHLHB2
Genes activated by
E2A on Affy
38
1
50
0
CSPPI
KIAA1702
protein
2
0
30
Genes repressed by
E2A on Affy


E) LDB1
Genes picked up on
LDB1 ChIP-on-chip
Genes activated by
LDB1 on Affy
29
0
73
0
0
0
43
Genes repressed by
LDB1 on Affy


F) LMO2
Genes picked up on
LMO2 ChIP-on-chip
Genes activated by
LMO2 on Affy
47
0
10
0
0
0
53
Genes repressed by
LMO2 on Affy

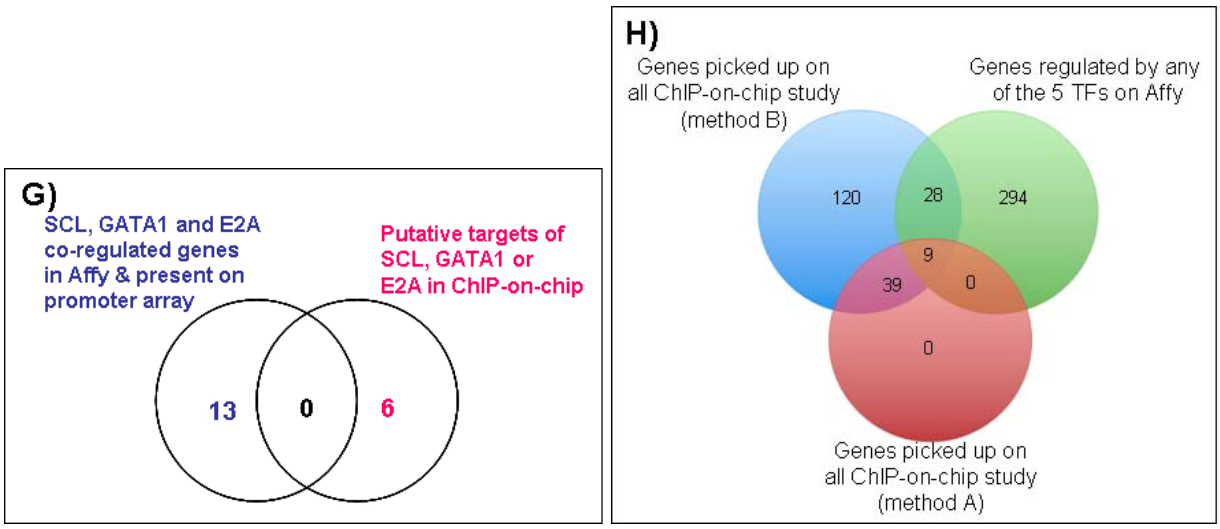


Figure 6.2. Venn diagram comparison of putative target genes identified in Affymetrix expression studies of siRNA-mediated knockdowns and ChIP-on-chip studies of the SCL erythroid complex. Numbers shown in the Venn diagrams are numbers of probes (either Affymetrix probe sets or promoter array elements). Some of the overlapping genes are labelled in the Venn diagram. Panel A: Venn diagram for GATA1; panel B: Venn diagram for SCL; panel C: Venn diagram of E12 ChIP targets compared with E2A siRNA knockdown; panel D: Venn diagram of E47 ChIP targets compared with E2A siRNA knockdown; panel E: Venn diagram of LDB1; panel F: Venn diagram of LMO2; panel G: Venn diagram of SCL, GATA1 and E2A co-regulated genes identified by Affymetrix analysis and the putative target gene promoters identified by either of SCL, GATA1 or E2A in ChIP-on-chip; panel H: Venn diagram of genes regulated by any one of the TF identified by Affymetrix analysis and the putative target gene promoters identified by any one of the five TFs in ChIP-on-chip (analysis method A or B).

6.3.1.2 Expression time-course study of the putative target genes during siRNA knockdown of the SCL erythroid complex

From the analysis described above, there was a low correlation between the Affymetrix data and the ChIP-on-chip/ChIP-qPCR data. There are many possible reasons for this, all of which are discussed in section 6.4.1. One of these reasons, which will be addressed in this section, is that the effect of the knockdown on the expression of the ChIP targets may not be manifested at the time point which was analysed in the Affymetrix analysis (i.e. 24 hours). Thus, it is necessary to examine the effects of siRNA-mediated TF knockdown across a time-course to determine when changes in expression occur for the target genes. To this end, time course studies of up to 48 hours were performed for the knockdown of GATA1, E2A, SCL, LDB1 and LMO2. A range of target genes identified by ChIP-on-chip/ChIP-qPCR were analysed in this expression study - these included targets identified by one TF only, by several, or by all five members of the SCL eythroid complex. The changes in expression of the putative target genes identified by ChIP-on-chip/ChIP-PCR analysis were investigated by quantitative PCR at 12, 24, 36 and 48 hours after knockdown. Except in the case of SCL, where only one siRNA was used, two siRNAs were used for all the other TFs (in a manner similar to that used for the Affymetrix analysis). Two independent biological replicates of each

siRNA were performed for each time-course experiment and qPCR validation of putative target genes (i.e., 4 datapoints). In the qPCR, the normalisation was performed against the luciferase knockdown. The % of mRNA remaining for the putative target genes after siRNA transfection was calculated based on the mean of the four datapoints derived for each target gene (Figure 6.3). The expression level of the TF being knocked down was also included in Figure 6.3 as a positive control for each experiment. Virtually all 14 putative target genes identified in the ChIP-on-chip studies showed changes in expression, to some degree, in the knockdown of all five members of the SCL erythroid complex.

The expression level of four housekeeping genes (β-ACTIN, RPL16, GAPDH and β-TUBULIN) were also monitored in each experiment. The expression of these four housekeeping genes was not expected to change substantially in the time course and therefore they were used as a baseline to determine a significant expression change for the target genes. At each time point, the standard deviation and mean of fold change in the four housekeeping genes in the TF knockdown compared to the luciferase knockdown were calculated. An expression fold change with two standard deviations above or below the mean expression of these four genes at each time point was chosen as the cut-off to determine statistically significant changes in expression in the putative target genes (Table 6.1).

siRNA knockdown		Cut-off in % of mRNA remained after siRNA knockdown			
		12h	24h	36h	48h
GATA1	Down-regulation	54.91	50.35	64.90	65.08
	Up-regulation	182.11	198.61	154.08	153.66
E2A	Down-regulation	67.61	45.63	59.03	79.11
	Up-regulation	147.91	219.16	169.40	126.40
SCL	Down-regulation	42.44	57.81	57.65	58.28
	Up-regulation	235.62	172.99	173.45	171.59
LDB1	Down-regulation	62.98	45.83	63.84	75.57
	Up-regulation	158.79	218.21	156.64	132.34
LMO2	Down-regulation	78.52	60.36	53.01	46.51
	Up-regulation	127.36	165.68	188.65	215.01

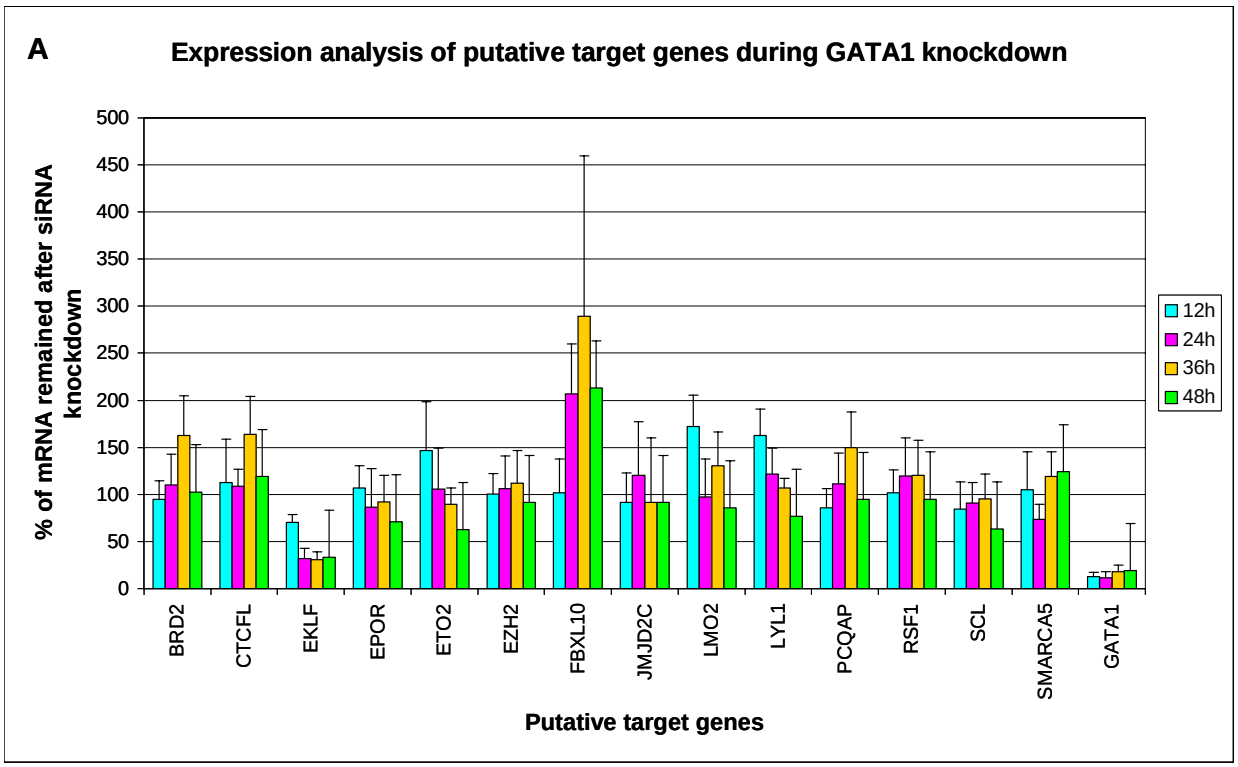
Table 6.1. Cut-off of fold increase or decrease for each siRNA knockdown study. The table shows the percentage of mRNA remained after siRNA knockdown determined by the cut-off calculated for each time point. Both the percentages in mRNA remained for up- or down-regulations were shown.

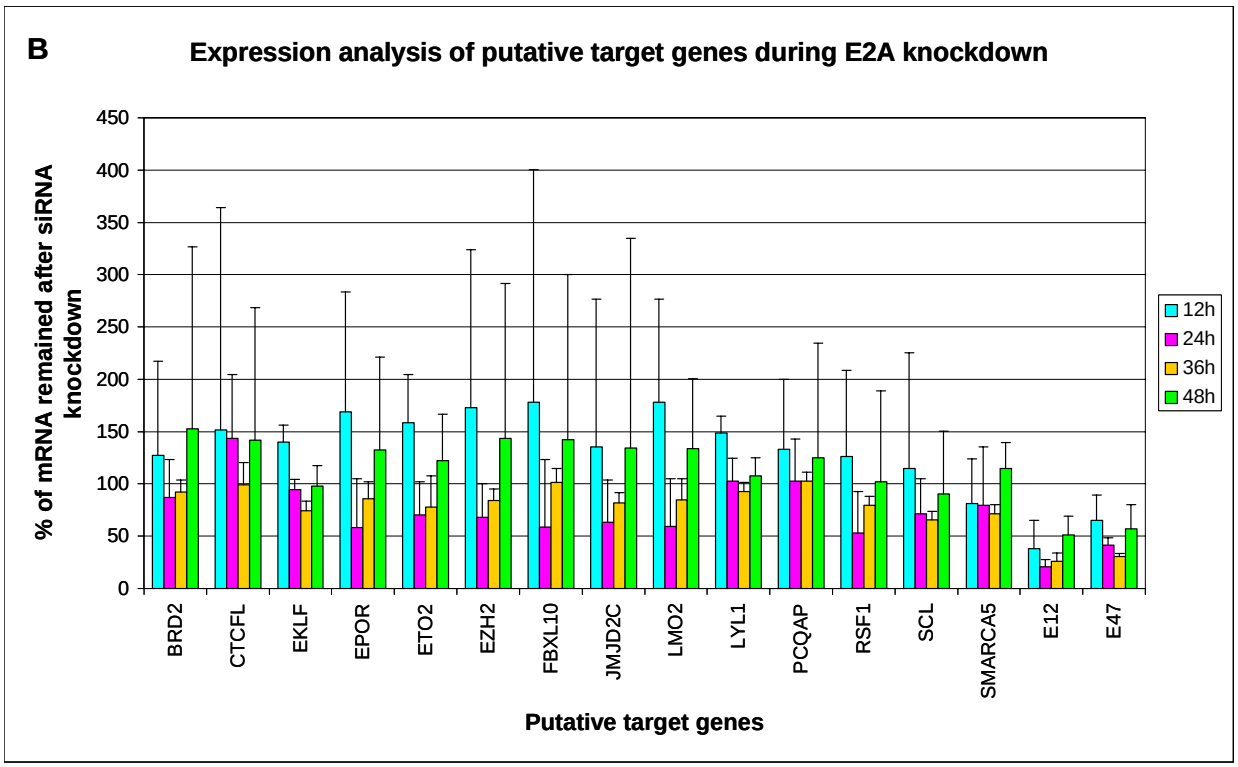
The expression changes of each of the putative target genes in the time-course knockdown study are shown in Table 6.2 and Figure 6.3 and summarised below. At one or more time points in any of the five TF knockdowns, the following interpretations were made:

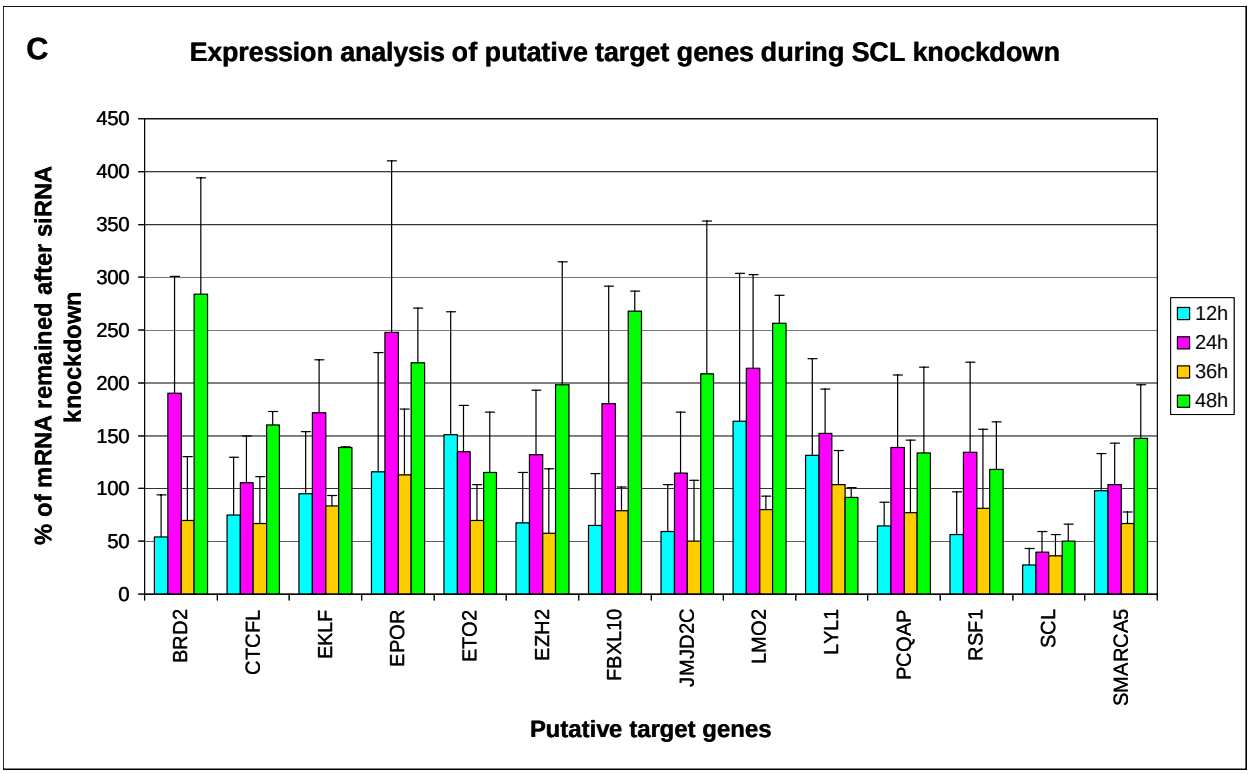
- BRD2 is activated by GATA1, E2A and SCL while repressed by LDB1 and LMO2.

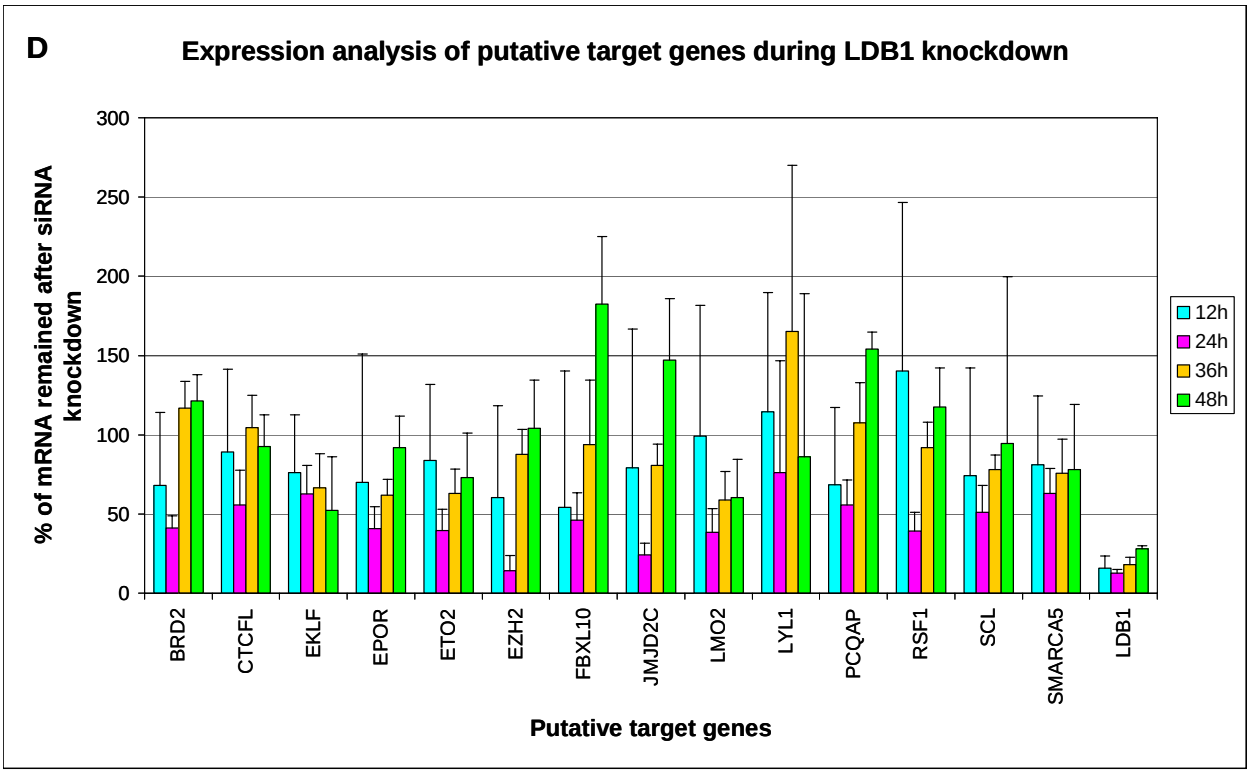
- CTCFL is activated by GATA1 and repressed by E2A.
- EKLF is activated by GATA1 and LDB1.
- EPOR is activated by E2A and SCL and repressed by LDB1.
- ETO2 is activated by GATA1 and LDB1 and repressed by E2A.
- EZH2 is activated by E2A and repressed by LDB1 and LMO2. It was also down-regulated at 36 hour and up-regulated at 48 hour for SCL knockdown. Such fluctuation in expression makes it hard to determine the mode of regulation by SCL.
- FBXL10 is activated by GATA1, E2A and SCL while repressed by LMO2. It was also down-regulated at 12 hour and up-regulated at 48 hour for LDB1 knockdown. Such fluctuation in expression makes it hard to determine the mode of regulation by LDB1.
- JMJD2C is activated by E2A while repressed by LMO2. It was also down-regulated at 36 hour and up-regulated at 48 hour for SCL knockdown and down-regulated at 24 hour and up-regulated at 48 hour LDB1 knockdown. Such fluctuation in expression makes it hard to determine the mode of regulation by SCL and LDB1.
- LMO2 is activated by E2A and SCL while repressed by and LDB1.
- LYL1 is repressed by E2A, LDB1 and LMO2.
- RSF1 is activated by LDB1.
- SCL is activated by GATA1.
- SMARCA5 is activated by LMO2.

Whilst the Affymetrix experiment only detected changes in three genes at the 24 hour time point, the time course qPCR analysis described here detected changes in all 14 target genes for at least one time point in at least one knockdown experiment (63 expression changes in total). Three genes (EZH2, FBXL10 and JMJD2C) showed complex patterns of up- and down-regulation. Furthermore, out of a total of 63 significant expression changes amongst all 14 targets, only 17 of them occur at the time point selected for the Affymetrix experiment (27%), while the rest occur at earlier or later time points. This indicates that the majority of measurable expression changes induced by siRNA knockdown occurred at different time points than the 24 hour time point studied in the Affymetrix GeneChip analyses.


A
Expression analysis of putative target genes during GATA1 knockdown
% of mRNA remained after siRNA knockdown
500
450
400
350
300
250
200
150
100
50
0
12h
24h
36h
48h
BRD2
CTCFL
EKLF
EPOR
ETO2
EZH2
FBXL10
JMJD2C
LMO2
LYL1
PCQAP
RSF1
SCL
SMARCA5
GATA1
Putative target genes


B
Expression analysis of putative target genes during E2A knockdown
% of mRNA remained after siRNA knockdown
450
400
350
300
250
200
150
100
50
0
12h
24h
36h
48h
BRD2
CTCFL
EKLF
EPOR
ETO2
EZH2
FBXL10
JMJD2C
LMO2
LYL1
PCQAP
RSF1
SCL
SMARCA5
E12
E47
Putative target genes


C
Expression analysis of putative target genes during SCL knockdown
% of mRNA remained after siRNA knockdown
450
400
350
300
250
200
150
100
50
0
BRD2
CTCFL
EKLF
EPOR
ETO2
EZH2
FBXL10
JMJD2C
LMO2
LYL1
PCQAP
RSF1
SCL
SMARCA5
Putative target genes
12h
24h
36h
48h


D
Expression analysis of putative target genes during LDB1 knockdown
% of mRNA remained after siRNA knockdown
300
250
200
150
100
50
0
BRD2
CTCFL
EKLF
EPOR
ETO2
EZH2
FBXL10
JMJD2C
LMO2
LYL1
PCQAP
RSF1
SCL
SMARCA5
LDB1
Putative target genes
12h
24h
36h
48h

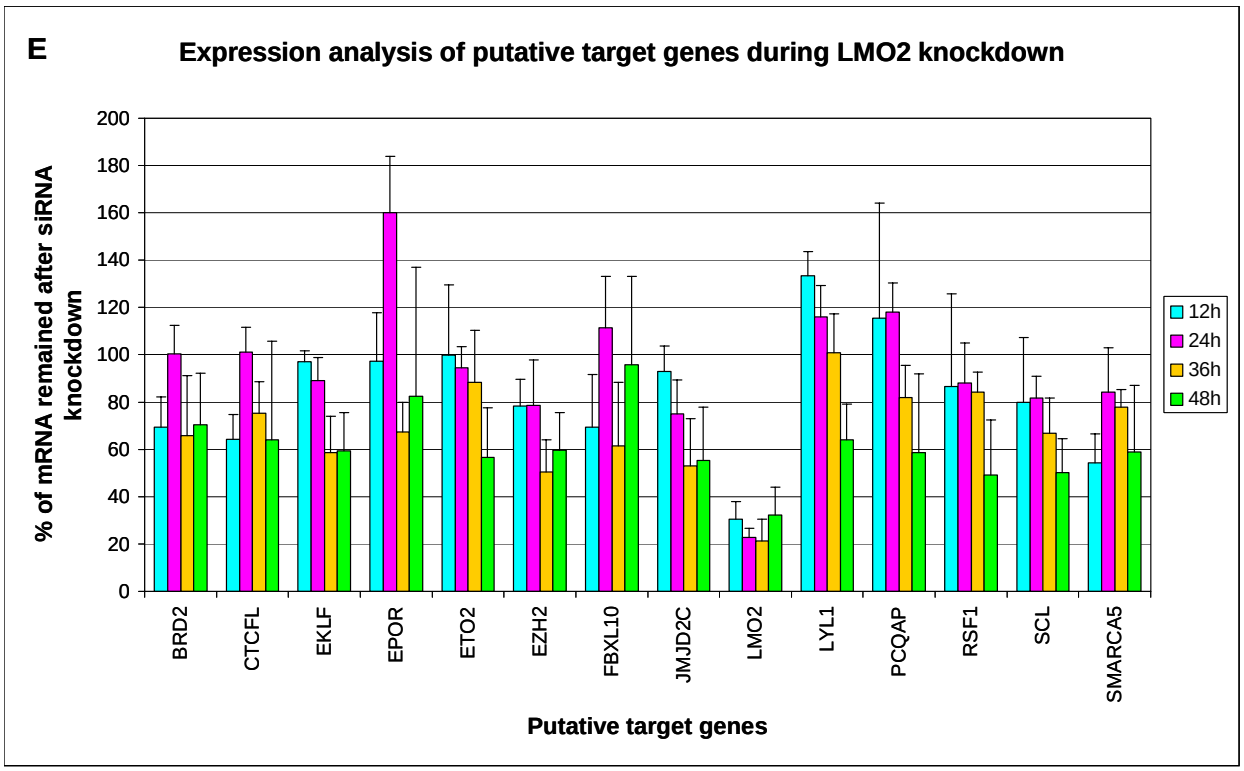


Figure 6.3. Time-course expression analyses of putative target genes during the siRNA knockdown of the SCL erythroid complex. Panel A: GATA1 knockdown; panel B: E2A knockdown; panel C: SCL knockdown; panel D: LDB1 knockdown; panel E: LMO2 knockdown. Y-axis: % of mRNA remaining after siRNA transfection normalised to luciferase siRNA transfection. These percentages are the mean values of 4 datapoints (2 biological replicates for two siRNAs). X-axis: putative target genes. Colour bars indicating the time points are shown on the right of the histograms.

Putative target gene	% of mRNA remained after siRNA knockdown																			
	GATA1 KD				E2A KD				SCL KD				LDB1 KD				LMO2 KD			
	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h
BRD2	95	110	163	103	127	87	92	153	54	190	70	284	68	41	117	122	69	100	66	70
CTCFL	113	109	164	119	152	143	99	142	75	105	67	160	89	56	104	92	64	101	75	64
EKLF	71	32	31	33	140	95	74	98	95	172	83	139	76	63	67	52	97	89	59	60
EPOR	107	87	92	71	169	58	86	132	116	248	113	219	70	41	62	92	97	160	67	83
ETO2	146	106	90	63	158	70	78	122	151	135	70	115	84	40	63	73	100	94	88	57
EZH2	100	107	112	91	173	68	84	144	68	132	58	198	60	14	88	104	78	79	50	60
FBXL10	102	207	289	213	178	59	101	142	65	180	79	268	54	46	94	182	70	111	61	96
JMJD2C	92	120	92	91	135	63	82	134	59	115	50	208	79	24	81	147	93	75	53	55
LMO2	172	97	131	86	178	59	85	133	163	214	80	256	99	39	59	60	31	23	21	32
LYL1	162	122	107	77	149	103	93	108	131	152	104	92	114	76	165	86	133	116	101	64
PCQAP	86	112	150	95	133	103	103	125	64	139	77	134	69	56	107	154	115	118	82	59
RSF1	102	120	120	95	126	53	79	102	56	134	81	118	140	39	92	117	87	88	84	49
SCL	85	91	95	63	115	71	66	90	28	40	36	50	74	51	78	94	80	82	67	50
SMARCA5	105	74	119	124	81	80	71	115	98	104	67	148	81	63	76	78	54	84	78	59

No significant change
Down-regulated below cut-off
Up-regulated above cut-off
TF being knocked down by siRNA

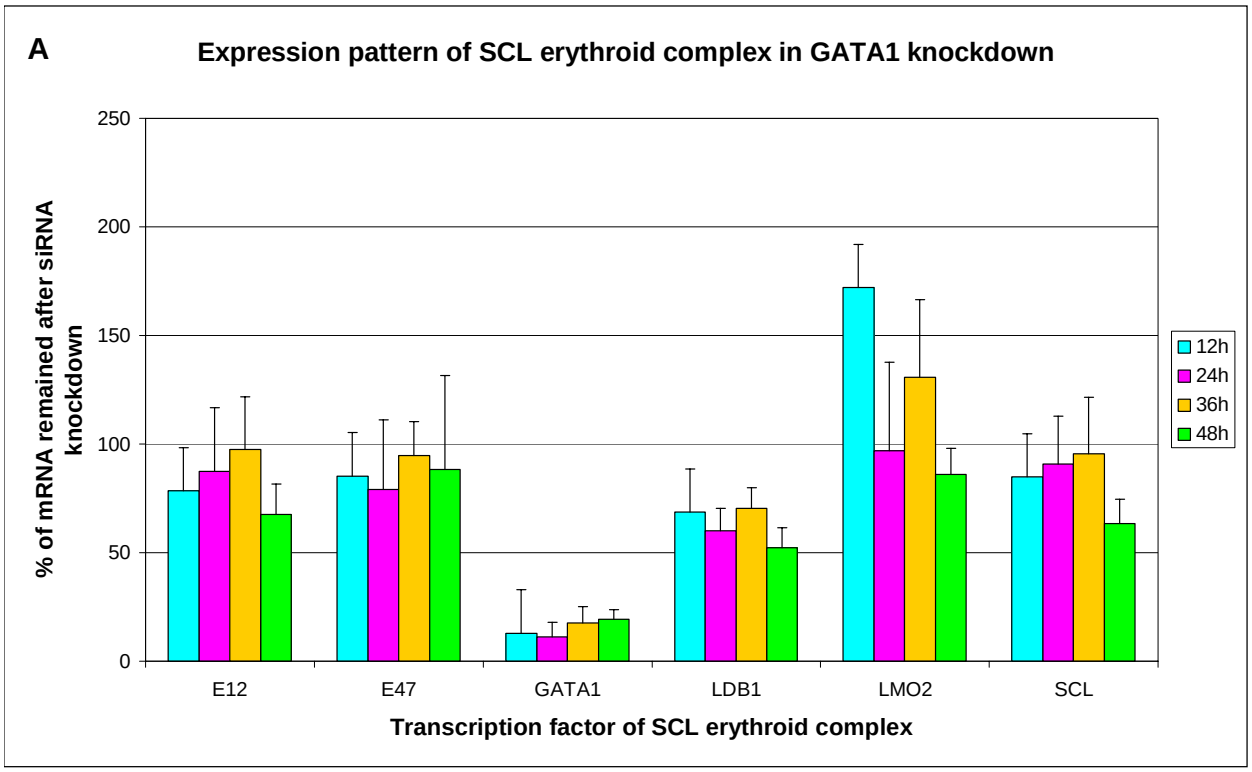
Table 6.2. Putative target genes showing statistically significant change in expression. The percentage of mRNA remained after siRNA knockdown (KD) at 12, 24, 36 and 48 hour time points are shown for each putative target gene. The green boxes indicate down-regulated genes having a change in expression below the cut-off determined in Table 6.2. The red boxes indicate up-regulated genes having a change in expression above the cut-off. The grey boxes the change in expression of the transcription factor being knocked down by siRNA.

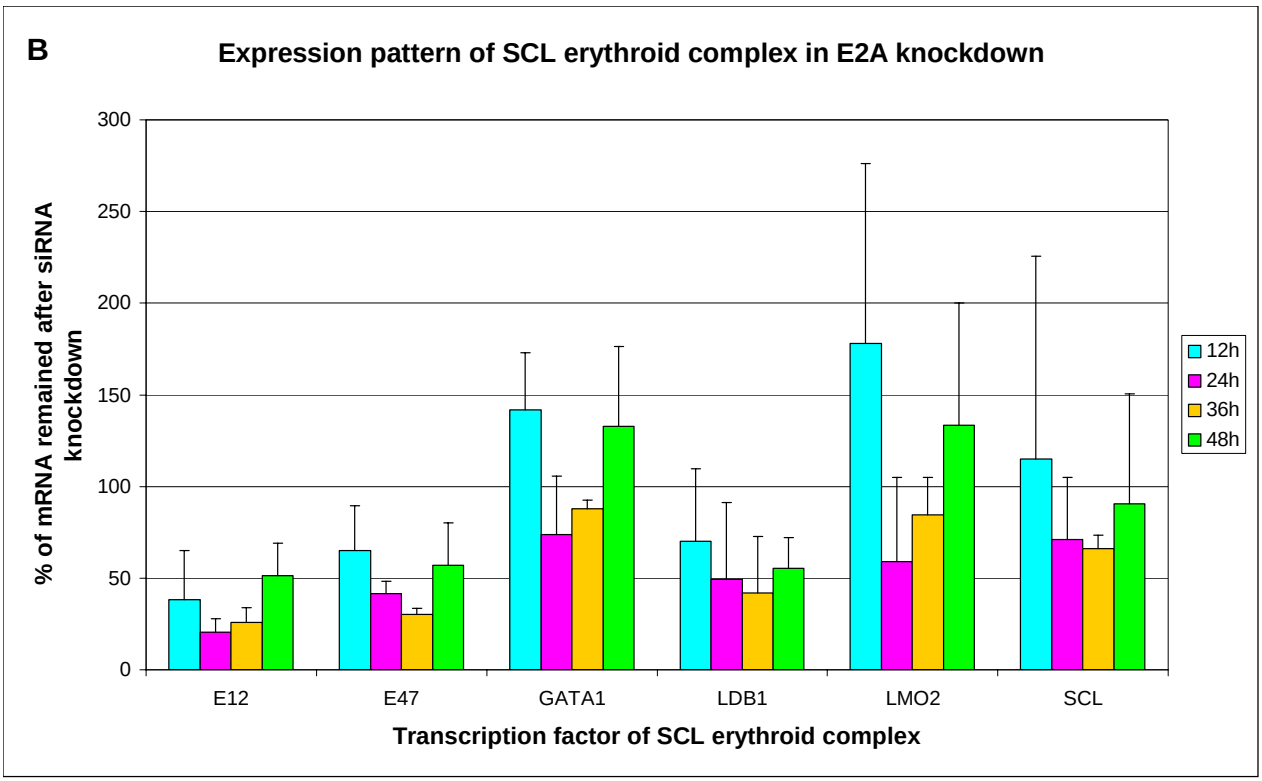
6.3.1.3 Further evidence for auto-regulation of the SCL erythroid complex at the level of gene expression

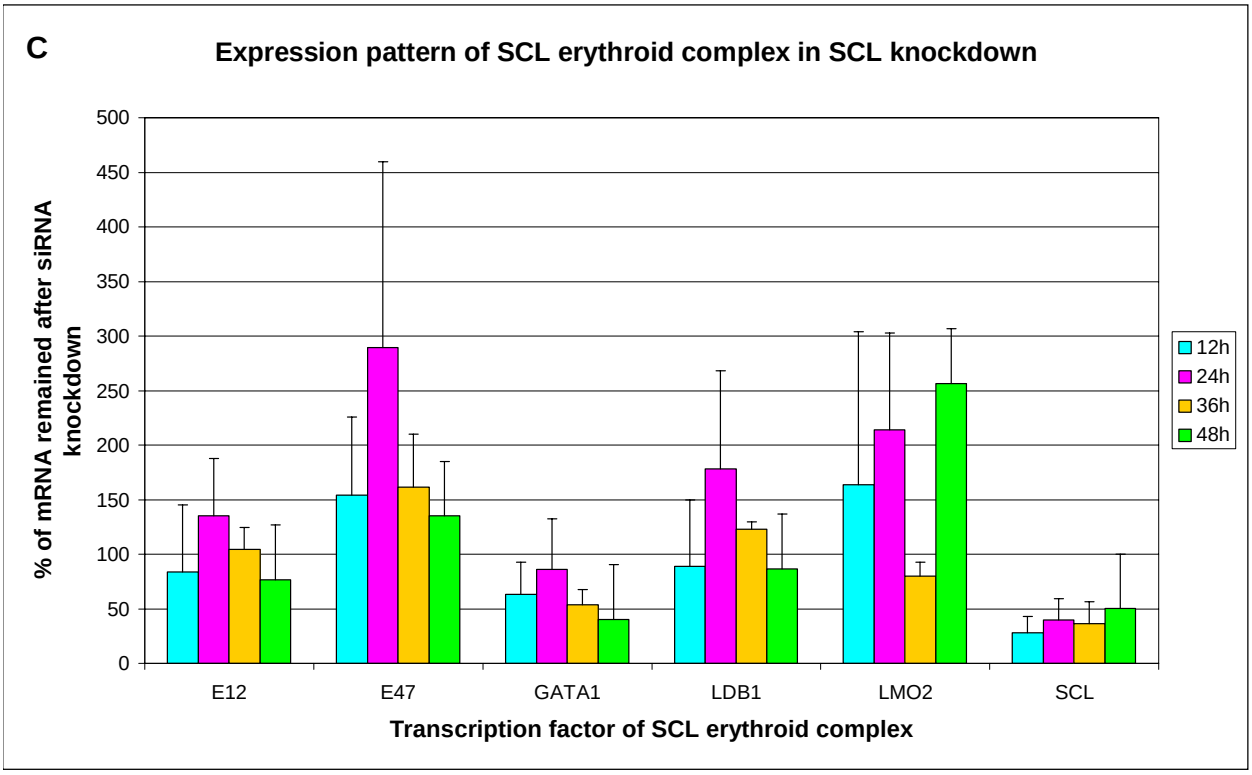
There was evidence presented in Chapters 4 and 5 which suggested that members of the SCL eyrthroid complex were self-regulated by the whole complex, or members therein. To understand this auto-regulation in more detail, the effects of siRNA-mediated knockdown of members of the complex was assessed on the expression levels for each member of the complex. The expression patterns for each member was studied by qPCR in 48 hour time-course experiments as described above for the target gene analysis (Figure 6.4). The expression of other TFs in the complex changed when one of the TFs was being silenced. The same fold change cut-off described in section 6.3.1.2 (Table 6.1) was used to determine statistically significant change in expression of the transcription factors during knockdown (Table 6.3). The data can be summarised as follows:

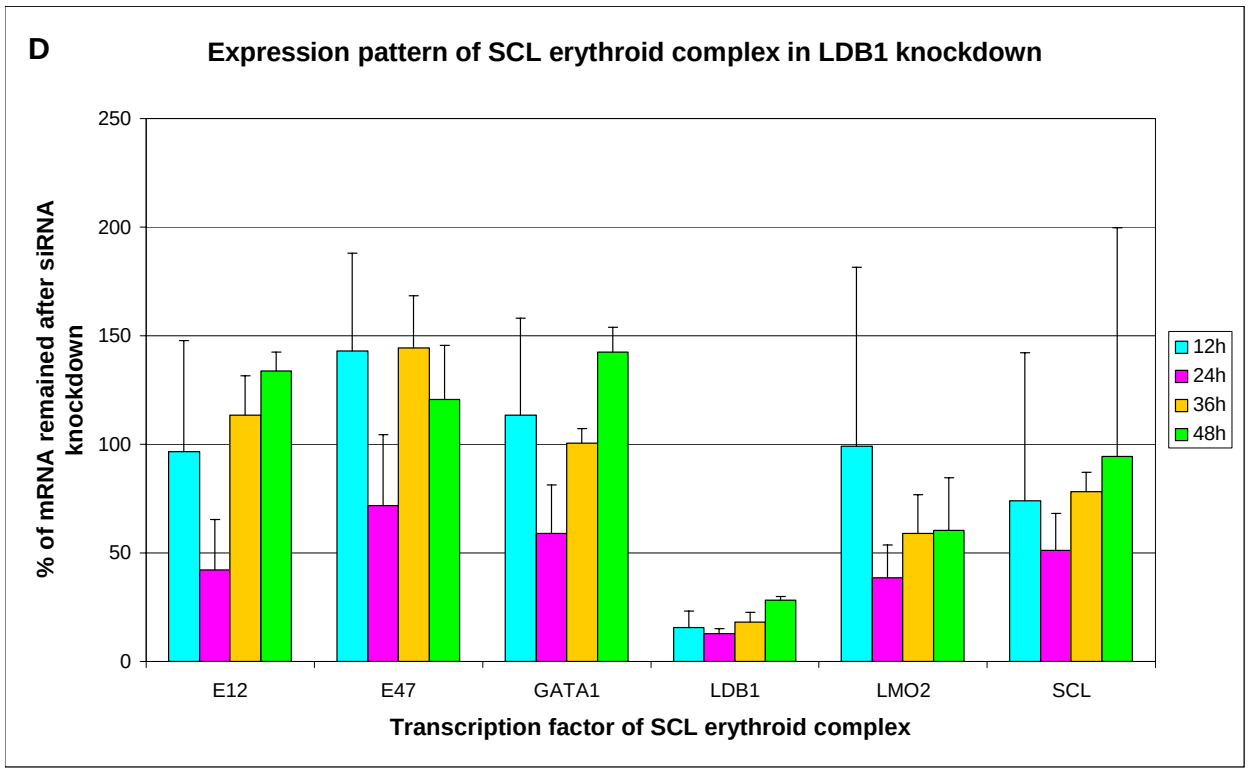
- In the knockdown of GATA1, LDB1 and SCL showed significant down-regulation at the 48 hour time point. Curiously, SCL, a direct target of the whole SCL erythroid complex (Chapter 5) only showed a significant expression change with the knockdown of GATA1 (and none of the other TFs).
- In the knockdown of E2A, significant up-regulation was observed for GATA1 at the 48 hour time point and for LMO2 at the 12 and 48 hour time points. Down-regulation was observed in LDB1 at the 36 and 48 hour time point.
- In the knockdown of SCL, significant up-regulation was also observed for E47 and LDB1 at the 24 hour time point and for LMO2 at the 24 and 48 hour time point. Down-regulation was observed in GATA1 at the 36 and 48 hour time point.
- In the knockdown of LDB1, significant up-regulation was also observed for GATA1 at the 48 hour time point. Expression of E12 was shown to be significantly down-regulated at 24 hour and up-regulated at 48 hour. Down-regulation was observed in LMO2 at the 24, 36 and 48 hour time point.
- In the knockdown of LMO2, significant down-regulation was also observed for E47 at the 12 hour time point.

The results shown here demonstrate that the knockdown of each TF in the complex, affects the expression of other members of the complex in a variety of ways involving both up and down regulation and combinations of both through time. This data further delineates the various modes of auto-regulation which are involved in modulating levels of each member of the TF complex. Such effects would compound the issue of identifying target genes for each member of the complex using knockdown analysis, since expression changes associated with the changing levels of other TFs in the complex, would also be reflected in the final Affymetrix GeneChip analyses.




B
Expression pattern of SCL erythroid complex in E2A knockdown
% of mRNA remained after siRNA knockdown
300
250
200
150
100
50
0
E12
E47
GATA1
LDB1
LMO2
SCL
Transcription factor of SCL erythroid complex
12h
24h
36h
48h


C
Expression pattern of SCL erythroid complex in SCL knockdown
% of mRNA remained after siRNA knockdown
500
450
400
350
300
250
200
150
100
50
0
E12
E47
GATA1
LDB1
LMO2
SCL
Transcription factor of SCL erythroid complex
12h
24h
36h
48h



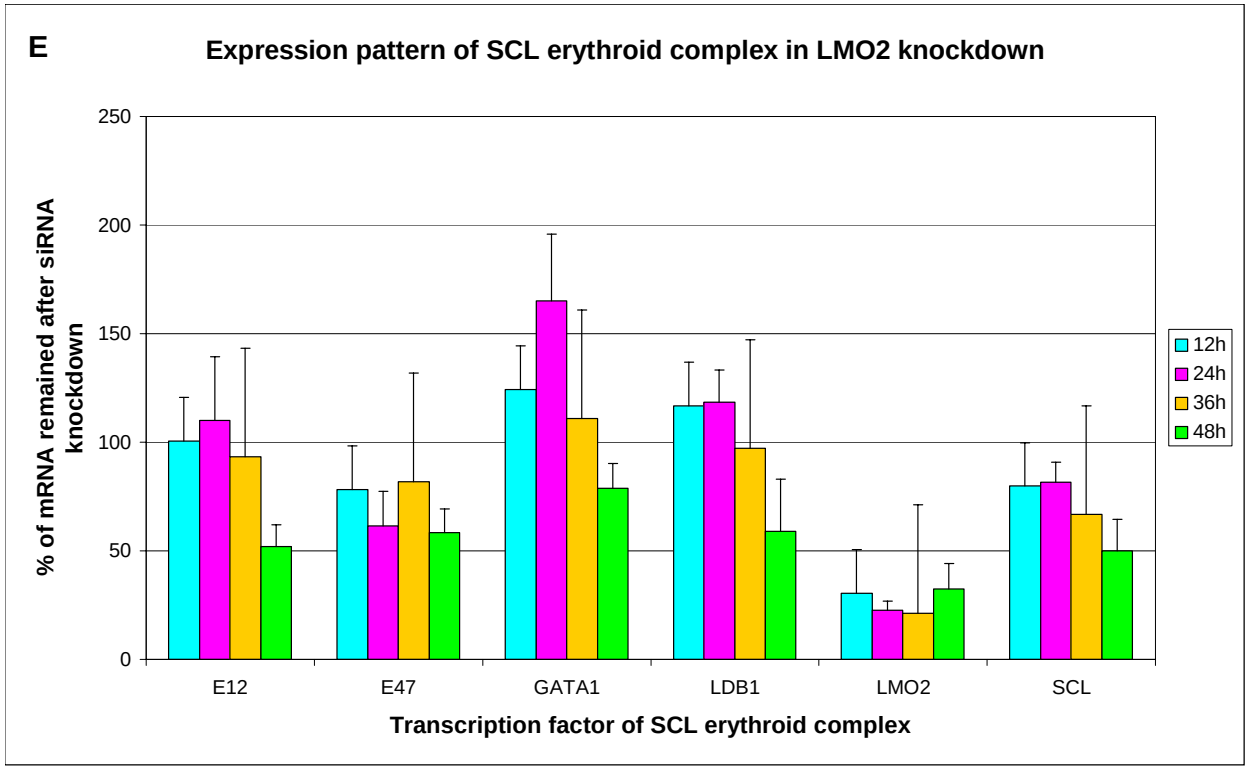


Figure 6.4. Time-course expression analyses for each of the five members of the SCL erythroid complex during siRNA knockdown of members of the complex. Y-axis: % of mRNA remaining after siRNA transfection normalised to luciferase knockdown. These percentages were the mean values of 4 datapoints from two biological replicates of 2 siRNAs used per TF. X-axis: transcription factor of SCL erythroid complex. Colour bars indicating the time points are shown on the right of the histograms.

Transcription factor of SCL erythroid	% of mRNA remained after siRNA knockdown																			
	GATA1 KD				E2A KD				SCL KD				LDB1 KD				LMO2 KD			
	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h	12h	24h	36h	48h
E12	78	88	98	68	38	21	26	51	84	135	105	77	97	42	113	134	101	110	93	52
E47	85	79	95	88	65	42	30	57	154	289	162	135	143	72	144	121	78	61	82	58
GATA1	13	11	18	19	142	74	88	133	63	86	54	40	113	59	100	142	124	165	111	79
LDB1	69	60	70	52	70	49	42	55	89	178	123	87	16	13	18	28	117	118	97	59
LMO2	172	97	131	86	178	59	85	133	163	214	80	256	99	39	59	60	31	23	21	32
SCL	85	91	95	63	115	71	66	90	28	40	36	50	74	51	78	94	80	82	67	50

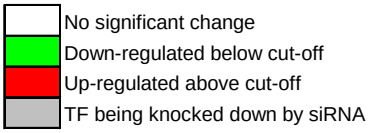


Table 6.3. Members of the SCL erythroid complex showing statistically significant change in expression. The percentage of mRNA remained after siRNA knockdown (KD) at 12, 24, 36 and 48 hour time points are shown for each of the five transcription factor. The green boxes indicate down-regulated genes having a change in expression below the cut-off determined in Table 6.1. The red boxes indicate up-regulated genes having a change in expression above the cut-off. The grey boxes the change in expression of the transcription factor being knocked down by siRNA.

6.3.1.4 ChIP-on-chip study of GATA1 knockdown

Data described in section 6.3.1.1 demonstrated that the TF binding events observed in ChIP-on-chip studies did not correlate, for the most part, with changes in expression of target genes after siRNA-induced knockdown. This may be due to differences in the experimental set-ups between the Affymetrix and ChIP-on-chip studies in combination with biological reasons. It is important to understand why these differences occurred in order to provide confidence that data derived from both experimental approaches was biologically meaningful. Experiments were designed to address these issues as described below.

The ways in which the ChIP-on-chip studies and Affymetrix experiments were conducted in this project were inherently different. For ChIP-on-chip, wild type K562 cells were used to identify targets. For Affymetrix analysis, knockdown samples were used. During knockdown, the kinetics of TF clearance from binding sites on target gene promoters during TF knockdown may not be the same for all targets – some may be removed from targets more rapidly than others. Thus, the effect of TF clearance on transcription may be different for these targets. Furthermore, TF clearance kinetics may not be directly inferred from the knockdown of the protein levels of the TFs themselves – thus, the time at which the majority of the clearance has occurred may not correspond to when the maximal knockdown of the protein was observed. Furthermore, the experimental manipulation of the knockdown cells (i.e., transfection) may affect TF binding events, further complicating the issues.

To provide some clues to why these discrepancies occurred between Affymetrix and ChIP-on-chip data, the effects of siRNA-mediated TF knockdown on the binding of TFs to promoters on the array

were monitored. This was performed for only one of the TFs - GATA1. A ChIP-on-chip study using the TF promoter array was performed for GATA1 during siRNA-mediated GATA1 knockdown at the 24 hour time point, consistent with the experimental set-up used for the Affymetrix analysis. The luciferase, GATA1a and GATA1b siRNAs were transfected into K562 cells by electroporation. After 24 hours, protein, total RNA and chromatin were extracted from the K562 cells. The knockdown of GATA1 was confirmed by quantitative PCR and western blotting for three independent biological replicates (Appendix 5). ChIP-on-chip was performed as previously described for the wild type K562 analysis.

Since the SCL locus was used as a positive control for the ChIP-on-chip studies described in Chapter 5 (and is also a key target of the entire SCL erythroid complex), the binding of GATA1 to various regulatory regions of the SCL locus during GATA1 knockdown was studied initially (Figure 6.5). The profiles of GATA1 binding after luciferase knockdown, GATA1a knockdown and GATA1b knockdown were shown to be very similar when compared to wild type K562 cells with little evidence for substantial loss of the GATA1 protein from all of the regulatory regions of SCL. In the study of GATA1 binding in wild type cells, four regions were shown to be significantly enriched: the -9/-10 enhancer, the +3 enhancer, promoter 1a and the +51 erythroid enhancer (labelled in Figure 6.5). The GATA1 enrichments of these four regions were compared in the luciferase control knockdown against the GATA1a and GATA1b knockdown (Table 6.4). The enrichment at the +51 enhancer did not change substantially after the knockdown of GATA1 whereas the enrichments for -9/-10 enhancer and promoter 1a were reduced by 15% and 24% respectively. The change for the +3 enhancer was the greatest with the enrichment decreasing by approximately 41% after siRNA-induced knockdown of GATA1.

Surprisingly, differences in fold enrichments were also observed between the wild type K562 cells and the luciferase siRNA transfected cells (Table 6.4). Fold enrichments at the +3 and +51 SCL enhancers increased by approximately 25% in the luciferase siRNA transfected cells and decreased by 11% and 14% at the -9/-10 enhancer and promoter 1a respectively.

This analysis of the SCL locus provided some initial evidence that GATA1 clearance from its binding sites does not necessarily reflect the degree of protein knockdown for GATA1 (more than 90% at the 24 hour timepoint), and that electroporation may also affect binding of GATA1 – at least at the SCL locus.

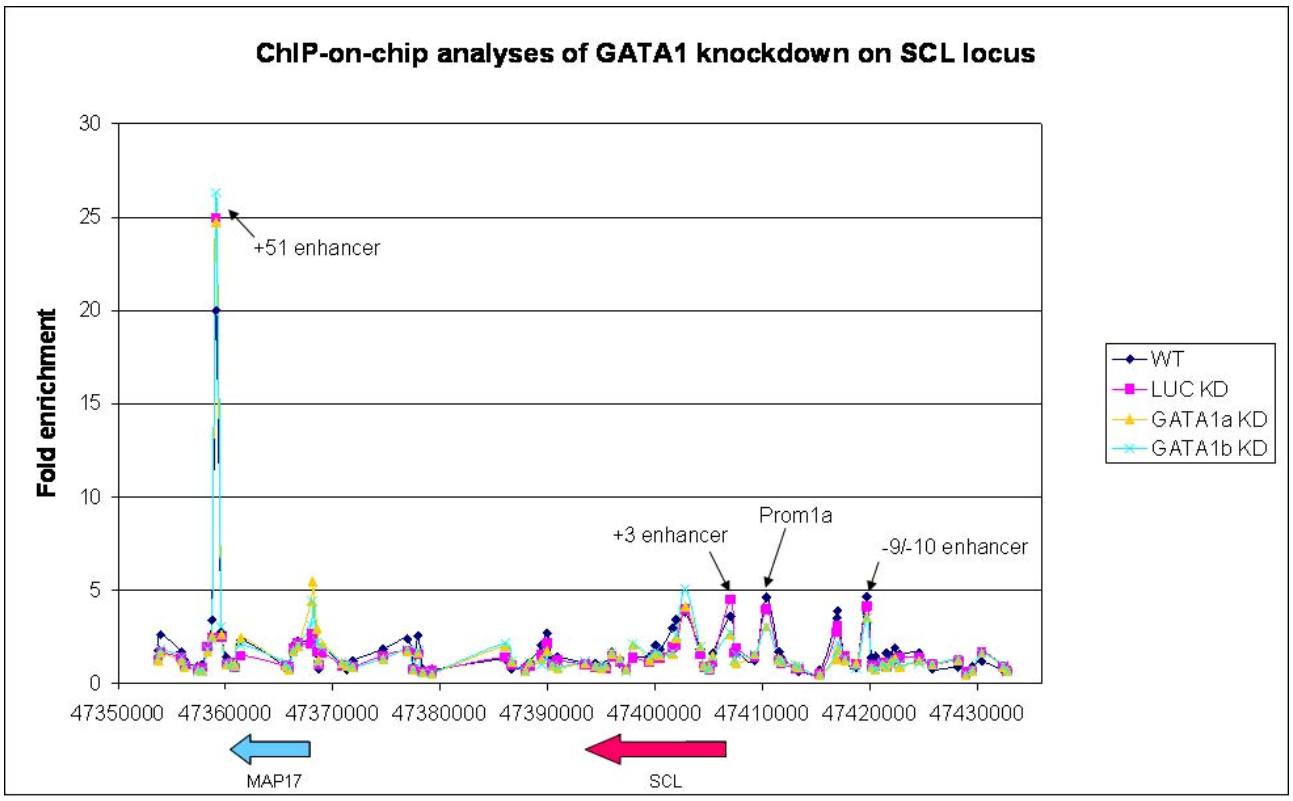


Figure 6.5. GATA1 ChIP-on-chip profile during GATA1 siRNA-mediated knockdown across the SCL locus in K562 cells. The ChIP-on-chip profiles across the SCL locus for wild type cells, luciferase, GATA1a and GATA1b siRNA transfected cells are shown. The x-axis shows the genomic coordinates across the SCL tiling path and the y-axis shows the fold enrichments. The thick-coloured arrows at the bottom of the figure show the position of the genes included on the SCL tiling path and their direction of transcription. Blue curve: profile for wild type cells; pink curve: profile for the luciferase knockdown (KD); yellow curve: profile for the GATA1a knockdown and aqua curve: profile for the GATA1b knockdown. Enhancers or promoters which showed significant enrichments are labelled by black arrows on the graph. The fold enrichments for each region were the mean of three independent biological replicates.

Regulatory elements on SCL locus	Fold enrichment				% change in enrichments in LUC KD against WT	% change in enrichments in GATA1 KD against LUC KD		
	Wild type	LUC KD	GATA1a KD	GATA1b KD	LUC vs WT	GATA1a KD	GATA1b KD	Average
+51 enhancer	20.00	24.98	24.72	26.33	24.9	-1.04	5.40	2.18
+3 enhancer	3.60	4.49	2.61	2.72	24.72	-41.87	-39.42	-40.65
Promoter 1a	4.61	3.97	3.03	3.03	-13.88	-23.68	-23.68	-23.68
-9/-10 enhancer	4.63	4.11	3.50	3.48	-11.23	-14.84	-15.33	-15.09

Table 6.4. Comparison of fold enrichments of enhancers and promoters of SCL in luciferase, GATA1a and GATA1b siRNA transfected K562 cells and wild type cells. This table shows the fold enrichment of the promoters and enhancers of SCL in luciferase, GATA1a and GATA1b siRNA transfected K562 cells and wild type cells and the percentage change for each regulatory element.

The ChIP-on-chip enrichments for the promoters of the putative target genes selected from the ChIP-on-chip studies in Chapter 5 were also investigated in the GATA1 knockdown condition (Figure 6.6 and Table 6.5). For thirteen of the target promoters, there was a reduction in the fold

enrichment in the GATA1 knockdown compared to the luciferase control knockdown with a percentage reduction of 8% to 44% for 13 of the targets - even though the overall protein level of GATA1 was reduced by 85-90%. In contrast, the fold enrichments increased for LYL1 by 13% (Figure 6.7). Differences in fold enrichment were again observed between the ChIP-on-chip performed in wild type K562 cells and the luciferase siRNA transfected cells. 9 out of the 14 promoters studied show decreases in GATA1 binding of up to 39% after luciferase siRNA transfection while the other 5 promoters showed increases of up to 68%. This further confirmed that electroporation with siRNAs may also affect binding of GATA1 to promoters.

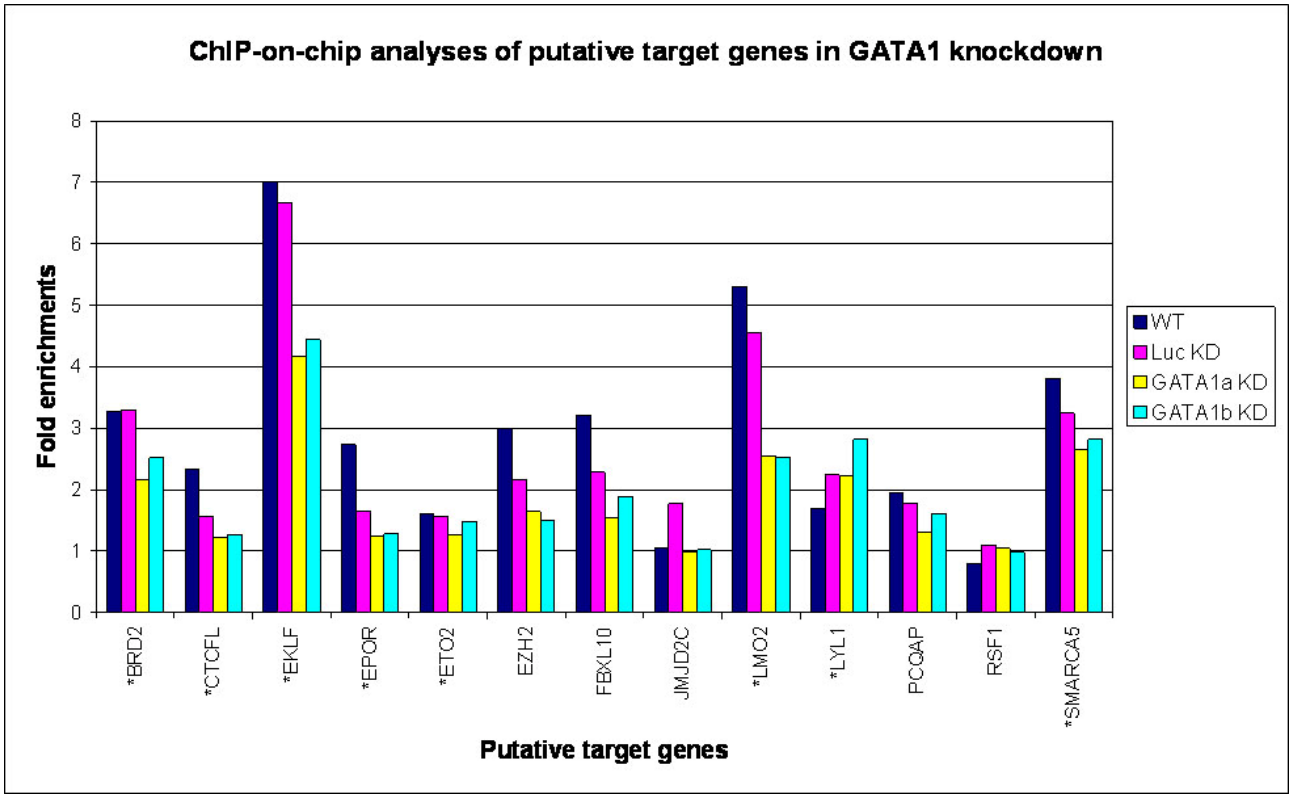


Figure 6.6. ChIP-on-chip analyses of GATA1 binding at target gene promoter during knockdown of GATA1. Histogram showed the fold enrichments for putative target genes in ChIP-on-chip studies in wild type K562 cells, the luciferase siRNA knockdown, the GATA1a siRNA knockdown and the GATA1b siRNA knockdown. Y-axis: fold enrichments. X-axis: putative target gene promoters. The ChIP-on-chip assays represented by the colour bars are shown in the key on the right. The fold enrichments for each target promoter were the averages of three independent biological replicates. The asterisk indicated genes with significant enrichments in the ChIP-on-chip (analysed by both methods A and B) and ChIP-qPCR studies of GATA1 in wild type cells in Chapter 5.

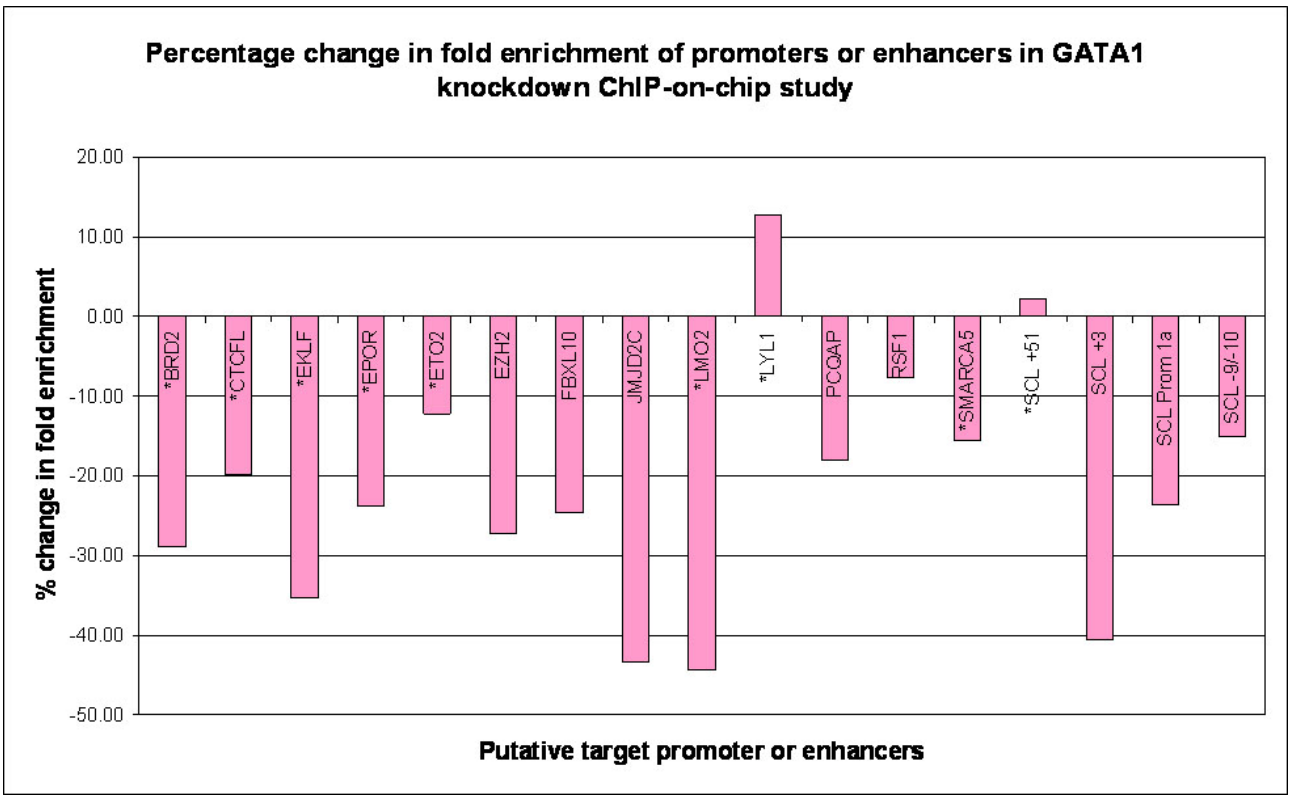


Figure 6.7. Percentage change in fold enrichment of promoters and enhancers in GATA1 knockdown ChIP-on-chip study. Histogram showed the percentage change in fold enrichments for putative target genes and the SCL promoter and enhancers in ChIP-on-chip studies in GATA1 siRNA knockdown compared to luciferase knockdown. Y-axis: percentage change in fold enrichments. X-axis: putative target gene promoters or SCL promoter and enhancers (last four bars). The asterisk indicated promoters or enhancers with significant enrichments in the ChIP-on-chip (analysed by both methods A and B) and ChIP-qPCR studies of GATA1 in wild type cells in Chapter 5.

Putative target gene	Fold enrichments				% change in enrichments in LUC KD against WT	% change in enrichments in GATA1 KD against LUC KD			% of mRNA remained after GATA1 KD at 24 hour
	WT	Luc KD	GATA1a KD	GATA1b KD	LUC vs WT	GATA1a KD	GATA1b KD	Average	Average
*BRD2	3.28	3.29	2.16	2.52	0.32	-34.47	-23.38	-28.93	110.31
*CTCFL	2.33	1.56	1.23	1.26	-33.30	-20.93	-18.81	-19.87	108.90
*EKLF	7.01	6.66	4.17	4.44	-5.02	-37.37	-33.42	-35.40	32.24
*EPOR	2.73	1.66	1.24	1.28	-39.36	-25.02	-22.65	-23.84	86.70
*ETO2	1.62	1.56	1.27	1.47	-3.41	-18.63	-5.76	-12.20	105.79
EZH2	2.99	2.16	1.64	1.50	-27.77	-23.80	-30.54	-27.17	106.56
FBXL10	3.21	2.28	1.55	1.89	-28.92	-31.95	-17.24	-24.59	206.71
JMJD2C	1.05	1.77	0.98	1.02	68.25	-44.47	-42.41	-43.44	120.38
*LMO2	5.31	4.55	2.54	2.53	-14.16	-44.18	-44.56	-44.37	97.02
*LYL1	1.69	2.24	2.22	2.82	32.30	-0.69	26.11	12.71	121.72
PCQAP	1.94	1.78	1.31	1.60	-8.36	-26.22	-10.02	-18.12	111.60
RSF1	0.81	1.10	1.05	0.98	36.13	-4.59	-10.68	-7.63	119.54
*SMARCA5	3.81	3.24	2.65	2.82	-14.95	-18.30	-12.98	-15.64	73.62

Table 6.5. Comparison of fold enrichments of putative target promoters in luciferase, GATA1a and GATA1b siRNA transfected K562 cells and wild type cells. This table shows the fold enrichment of the putative target promoters in luciferase, GATA1a and GATA1b siRNA transfected K562 cells and in wild type cells. The percentage

change in enrichment between the luciferase and GATA1 siRNAs is shown in the fifth column. The percentage of mRNA remained after GATA1 KD at the 24 hour time point from the time-course study in section 6.3.1.2 is shown in the last column. The asterisk indicated genes with significant enrichments in the ChIP-on-chip (analysed by both methods A and B) and ChIP-qPCR studies of GATA1 in wild type cells in Chapter 5.

The percentage clearance of the GATA1 protein from the promoters after GATA1 knockdown was also compared to the expression changes for these genes which were obtained from the qPCR time course studies at the 24 hour time point (section 6.3.1.2) (Table 6.5). For genes which showed the greatest percentage of GATA1 clearance (i.e. LMO2 and JMJD2C), the change in expression was not significant. However, for EKLF, the percentage clearance was 34% and there was a 70% reduction in mRNA level. This suggests that the effect, at the level of expression, of clearance of GATA1 from the promoters of target genes may be different for each target gene. For this reason, the knockdown of GATA1 would have different effects on each of its target genes – and only some may demonstrate measurable expression changes in the knockdown condition. These data again help resolve issues which relate to the inability to detect expression changes for many of the targets of the SCL erythroid complex at the 24 hour time point.

In summary, the results presented here suggest that differential rates of GATA1 clearance from target promoters in knockdown experiments (resulting from effects of GATA1 knockdown and electroporation of siRNAs) may confound attempts to identify measurable expression changes in both qPCR and Affymetrix GeneChip analysis.

6.3.1.5 Off-promoter regulation of SCL and LYL1

The ChIP-on-chip studies of Chapter 5 identified LYL1 as a target gene of four of the five members of the SCL erythroid complex – and is therefore likely to be a good candidate to be regulated by the entire complex. Expression analysis by qPCR confirmed that knockdown of each member of the complex affected the expression of LYL1, further supporting a role for the whole complex in its regulation. However, its promoter does not have an E-box/GATA composite motif which would support the role of the whole SCL erythroid complex involved in its regulation. The absence of an E-box/GATA site in its promoter is analogous to the situation found for SCL – which is thought to be a functional and structural paralogue of LYL1 at both the protein and DNA level (Chapter 1, section 1.4.2.1). SCL, however, has a canonical E-box/GATA composite motif in its +51 enhancer sequence – suggesting that the +51 region may mediate the binding of the SCL erythroid complex. Regulation of the gene may therefore be achieved through the interaction of +51 with its cognate promoter – thus facilitated the detection of the erythroid complex on the promoter using ChIP-on-chip. If such a similar situation were also true for LYL1 (and the paralogy between the two genes extended to regulation), one would expect that a downstream enhancer of LYL1 would contain an

E-box/GATA composite site. This site would therefore also bind the SCL eythroid complex and this would be detected by ChIP.

To elucidate whether this type of regulation occurs outside the promoter region of LYL1, TFBS search by TESS and TFSearch together with comparative genomic analyses was used to determine the level of paralogy between the SCL and LYL1 promoters, and to aid in the identification of a region downstream of the LYL1 promoter which showed structural hallmarks of the SCL +51 enhancer. Figure 6.8 A and B shows that the +51 enhancer contains the consensus E-box/GATA motif while promoter 1a contains two GATA sites residing close to each other. The LYL1 promoter also contains two GATA sites (Figure 6.8 C) and, at approximately 33 kb downstream of the transcription start site of LYL1, a highly conserved E-box and GATA motifs separated by 8 bases was identified (Figure 6.8 D and 6.9). ChIP-qPCR was performed to complement assays performed previously in Chapter 5 to demonstrate the binding of members of the SCL erythroid complex at the SCL and LYL1 promoters and also at the downstream regions of SCL and LYL1 (Figure 6.10). Substantial enrichments of up to 60 fold were seen in both the LYL1 promoter and the +33 region in both K562 and HEL in all five ChIP assays except for LMO2 (where the antibody used did not facilitate good enrichments for any of the experiments performed in this thesis). The enrichments at the SCL promoter were somewhat lower, but, as was previously shown by ChIP-on-chip studies, the +51 enhancer also showed large enrichments of up to 60-fold in both K562 and HEL cells. This data suggests that a putative novel regulatory element for LYL1 was identified (see also Discussion). Moreover, the regulation of both SCL and LYL1 may be similar and is likely to be mediated through interactions between distal elements and their corresponding promoters. This data also demonstrates that the SCL erythroid complex is likely to mediate interactions through other regulatory regions apart from promoters, many of which may not have been detected by TF binding on the promoter array.

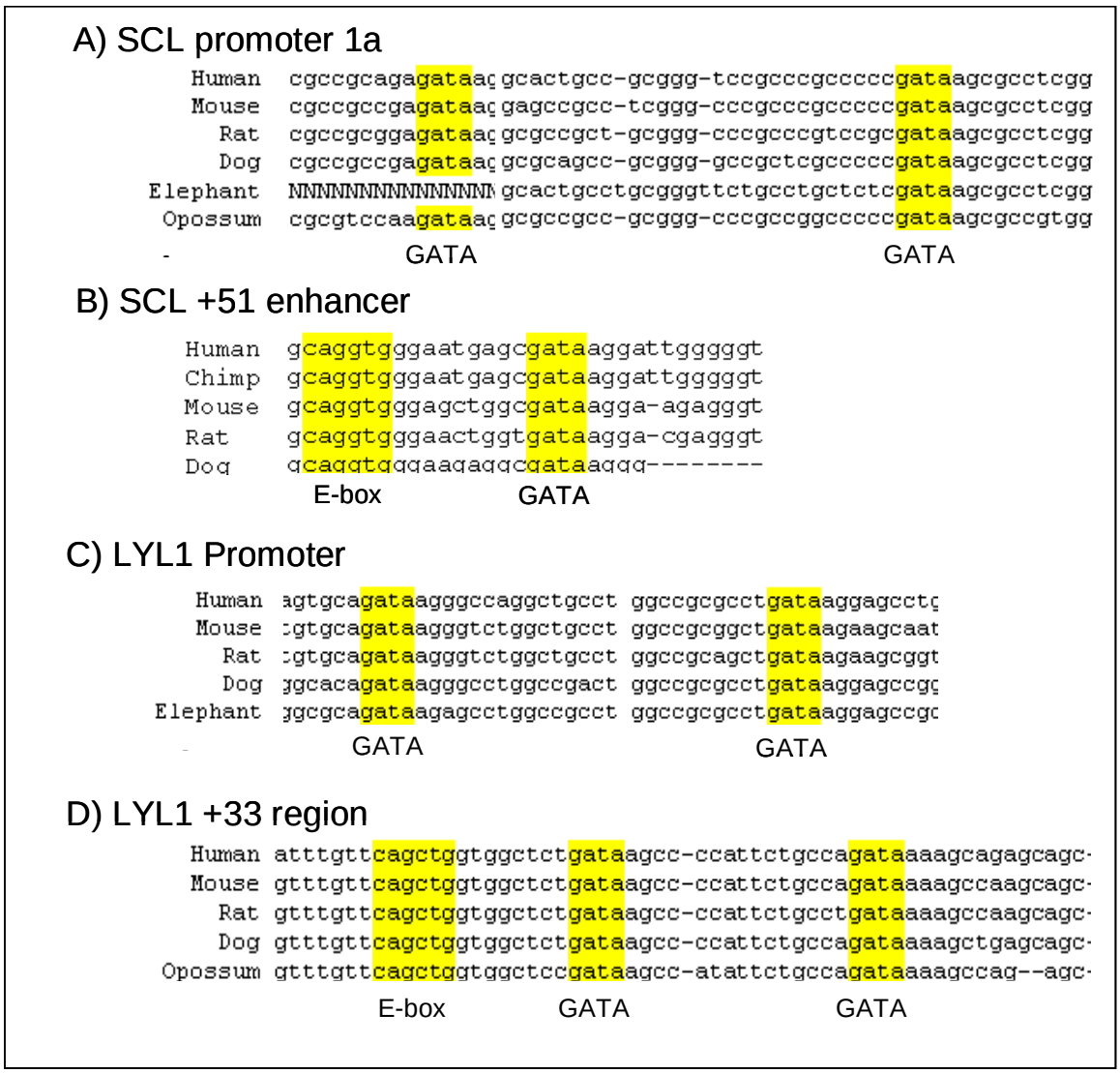


Figure 6.8. Multiple sequence alignments of transcription factor binding sites in the regulatory regions of SCL and LYL1. E-box and GATA motifs were identified by TESS and TFSearch and by viewing the conserved TFBS track on the UCSC genome browser. Multiple species sequence alignments were taken from the UCSC genome browser. Labelled in yellow are the conserved nucleotides across species in the E-box or GATA motifs. Panel A: SCL promoter 1a; panel B: SCL +51 enhancer; panel C: LYL1 putative promoter; panel D: LYL1 putative +33 enhancer.

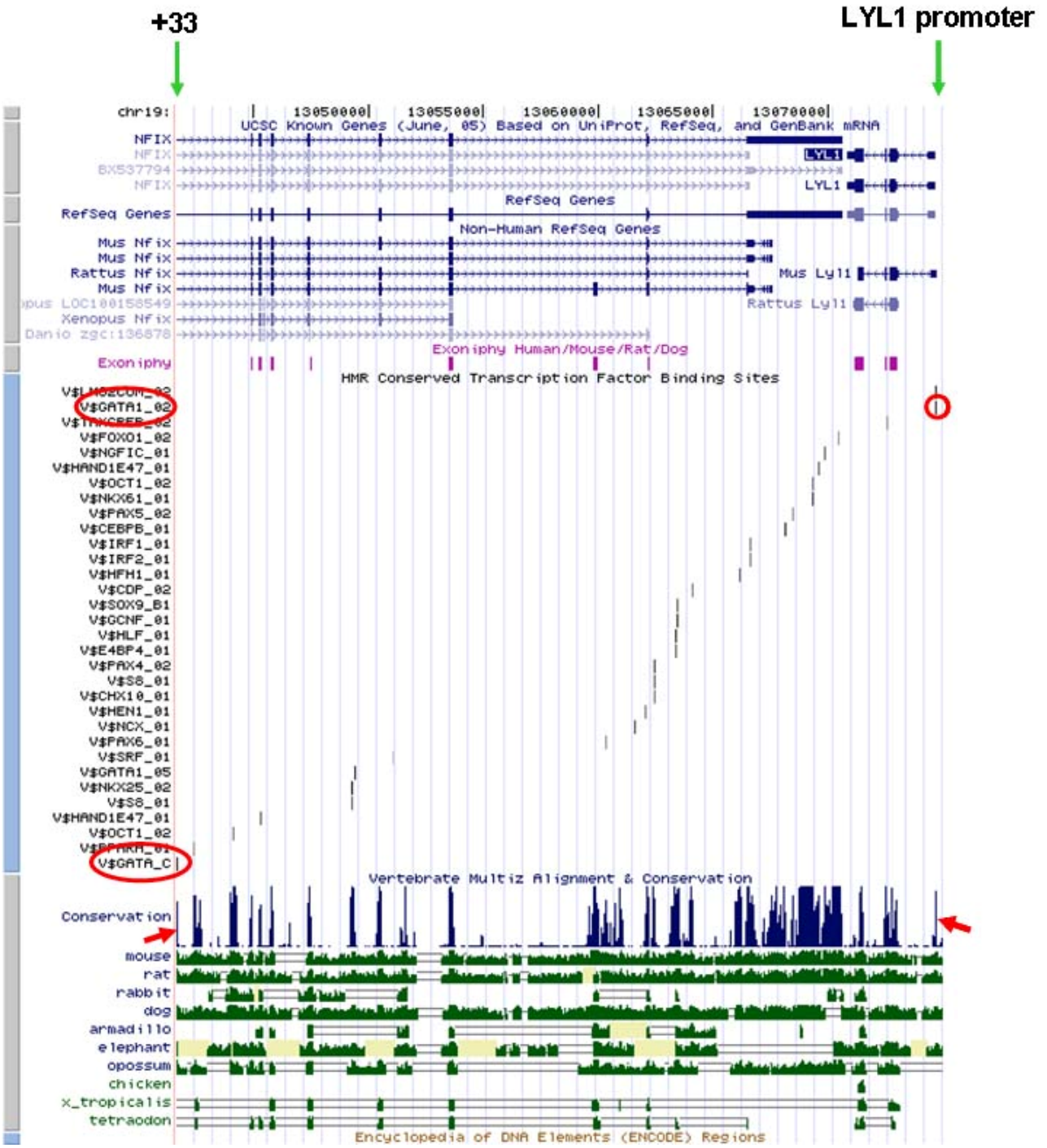


Figure 6.9. UCSC genome browser snapshot of the LYL1 promoter and putative enhancer regions. This diagram shows the LYL1 promoter and the +33 region in the UCSC genome browse. The green arrows indicate the position of the LYL1 promoter and the +33 region. The red circles show the conserved GATA sites identified by UCSC and the red arrows show the conservation of these GATA sites across species.

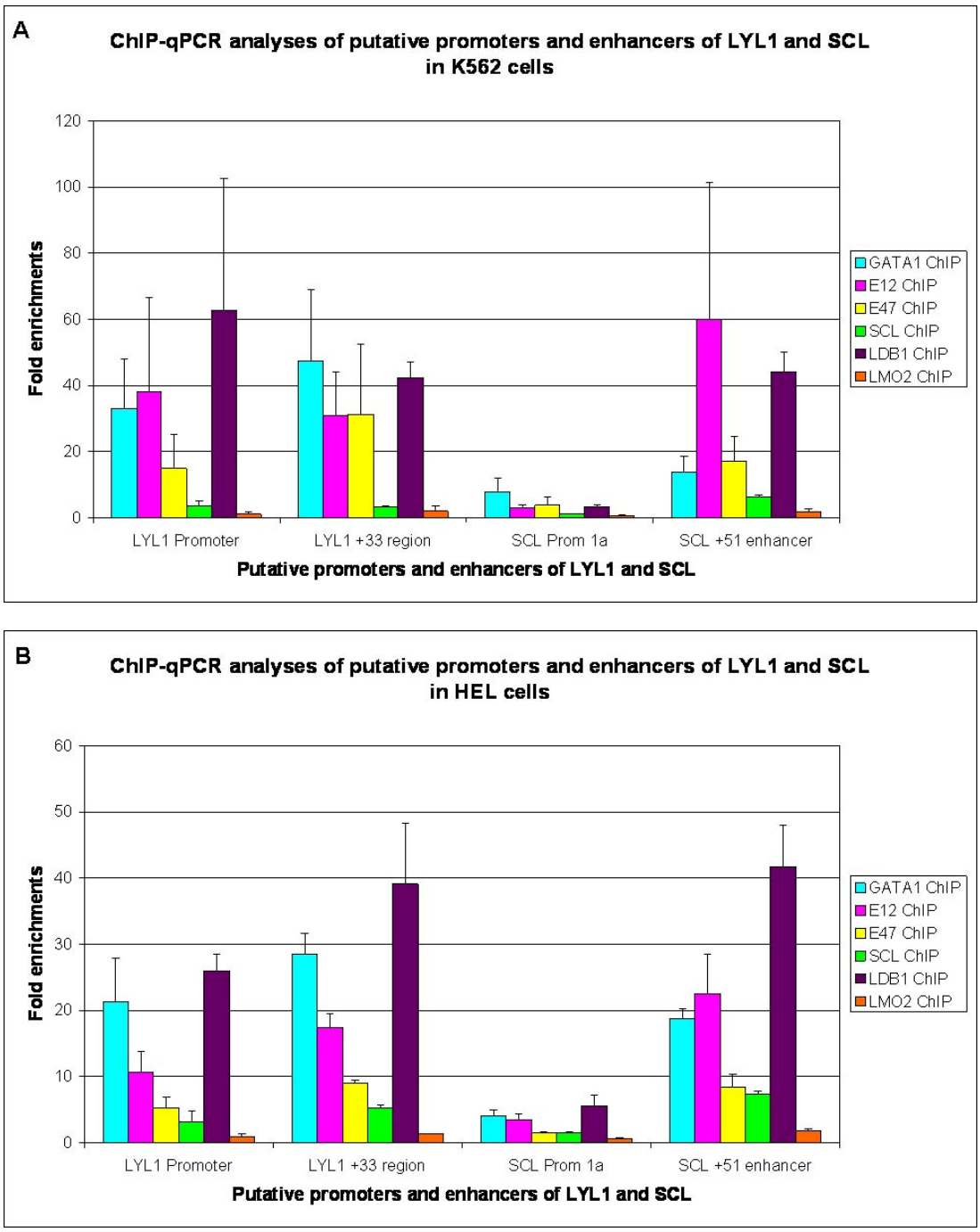


Figure 6.10. ChIP-qPCR analyses of promoters and distal regulatory elements of LYL1 and SCL in K562 and HEL cells. Histograms showed the fold enrichments of promoters and enhancers of LYL1 and SCL in ChIP-qPCR. Panel A: ChIP-qPCR in K562; panel B: ChIP-qPCR in HEL. Y-axes: fold enrichments above background. X-axes: regulatory regions of target genes. The colour-coded key for the various ChIP assays against members of the SCL erythroid complex is shown on the right of the panels. Error bars showed standard errors of two biological replicates.

6.3.2 Integration of expression and ChIP data: Derivation of transcriptional interaction networks in the erythropoeitic lineage.

The expression analyses on Affymetrix array and the ChIP-on-chip assays in the previous Chapters, together with the expression time-course studies of putative gene targets described in this Chapter, provide a wealth of biological information describing the cause and effects of regulatory interactions by members of the SCL erythroid complex in the human erythroid lineage. However, the vast amount of data produced from these experiments, and the variety of types/sources of data (array, qPCR, ChIP, two different cell lines, etc.,) makes overall interpretations difficult. To facilitate the integration of these datasets in a meaningful way, interaction network diagrams were generated using all or subsets of the data as discussed in the following sections.

6.3.2.1 Networks generation based on ChIP-on-chip data

The results obtained from the ChIP-on-chip experiments in Chapter 5 identified the promoters of 24 genes (Table 5.2) which were likely to be putative target for one or more members of the SCL erythroid complex. These 24 target genes, identified with both statistical methods A and B (summarised in Chapter 5, Table 5.2) were used for the generation of a network diagram using BioTapestry (Figure 6.11). This network diagram only shows the binding of each member of the SCL erythroid complex to the promoters of their putative target genes but does not integrate the mode of regulation involving these TFs at the transcriptional level (activation or repression).

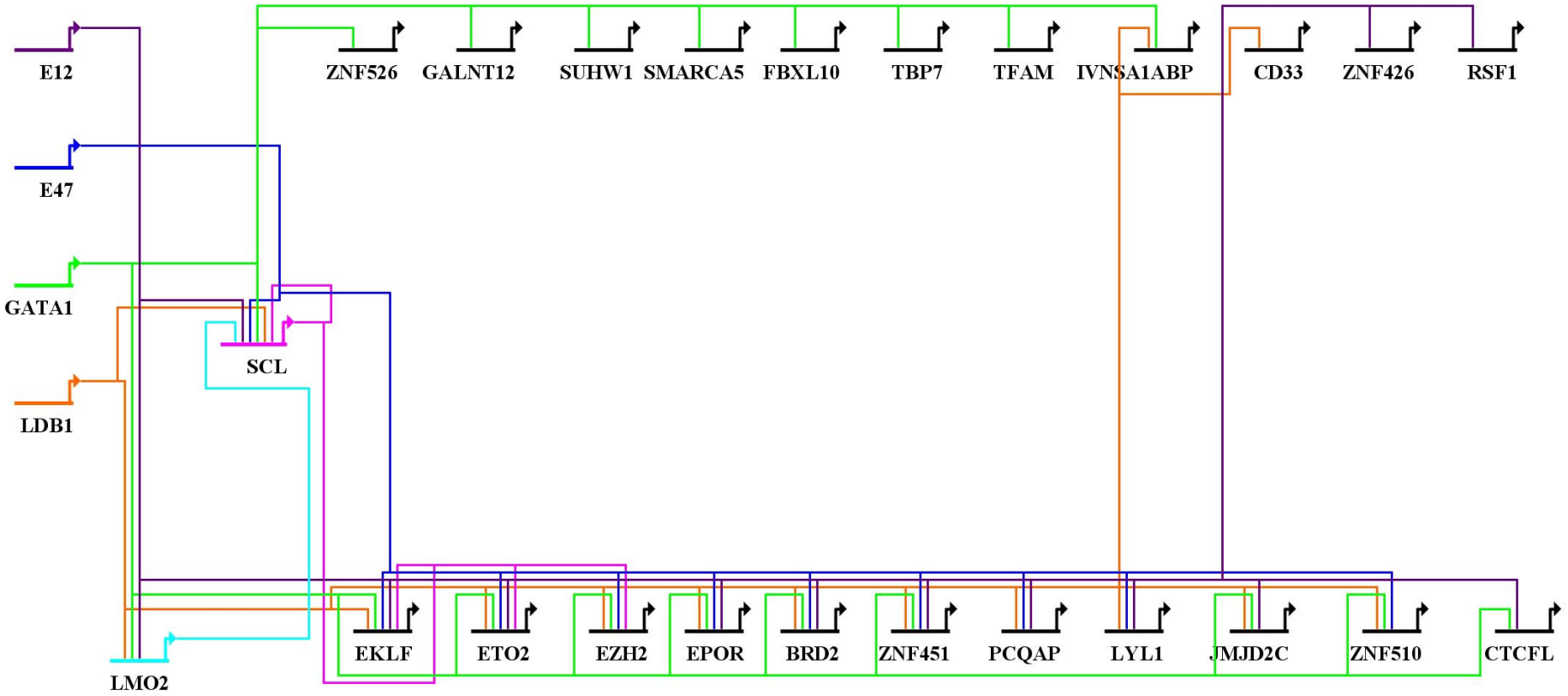
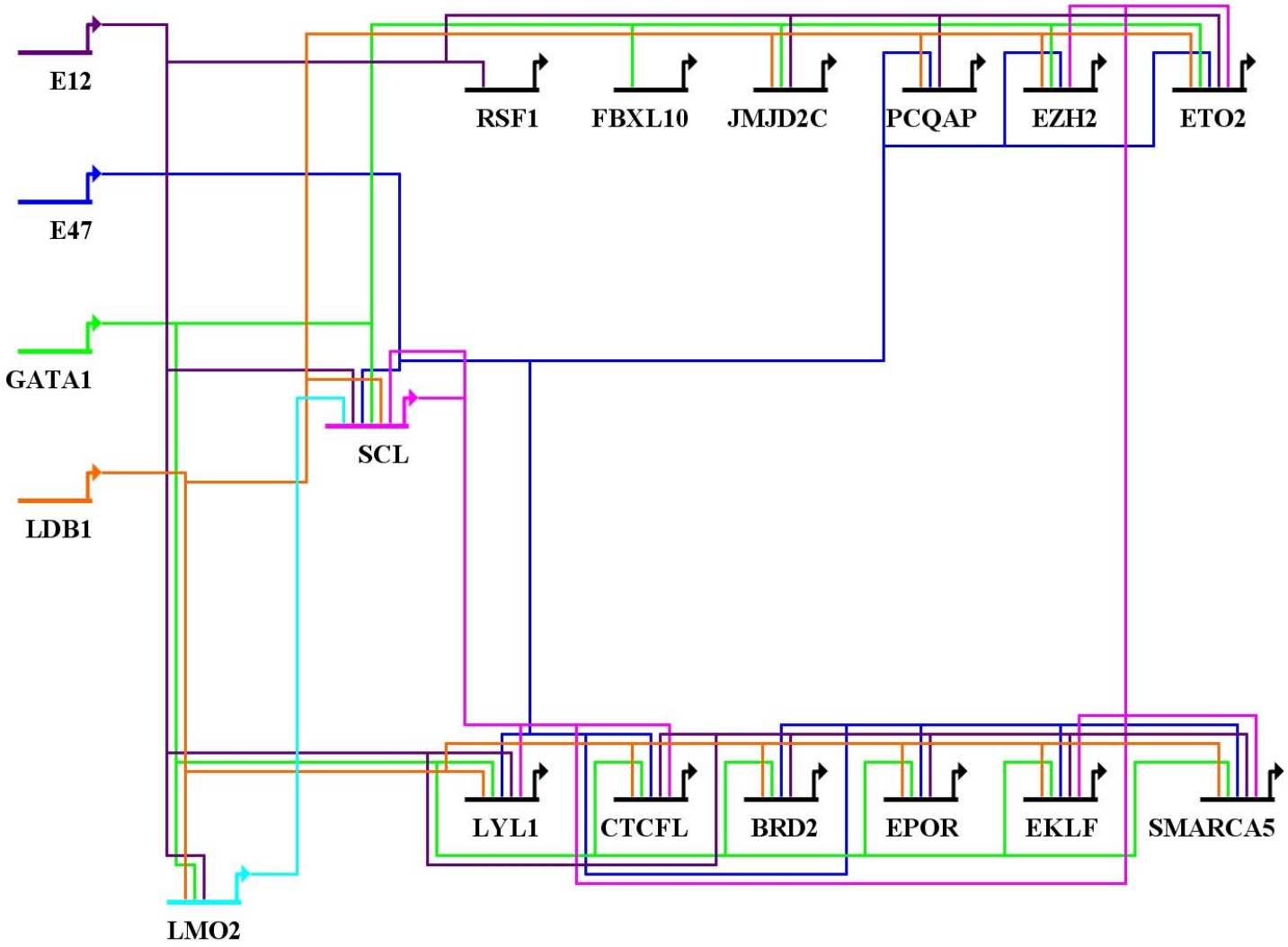


Figure 6.11. Network diagram of promoter-TF interactions of members of the SCL eyrthroid complex (GATA1, SCL, E12, E47, LDB1 and LMO2) in ChIP-on-chip analyses. Coloured bars on the left represent the 6 members (GATA1, SCL, E12, E47, LDB1 and LMO2) in the complex under study while the black bars on the right shows the targets. Note that LMO2 and SCL are also target genes. The coloured lines show the binding events between the TF and the target gene promoters; SCL: pink, GATA1: green, E12: violet, E47: blue, LDB1: orange, LMO2: aqua.

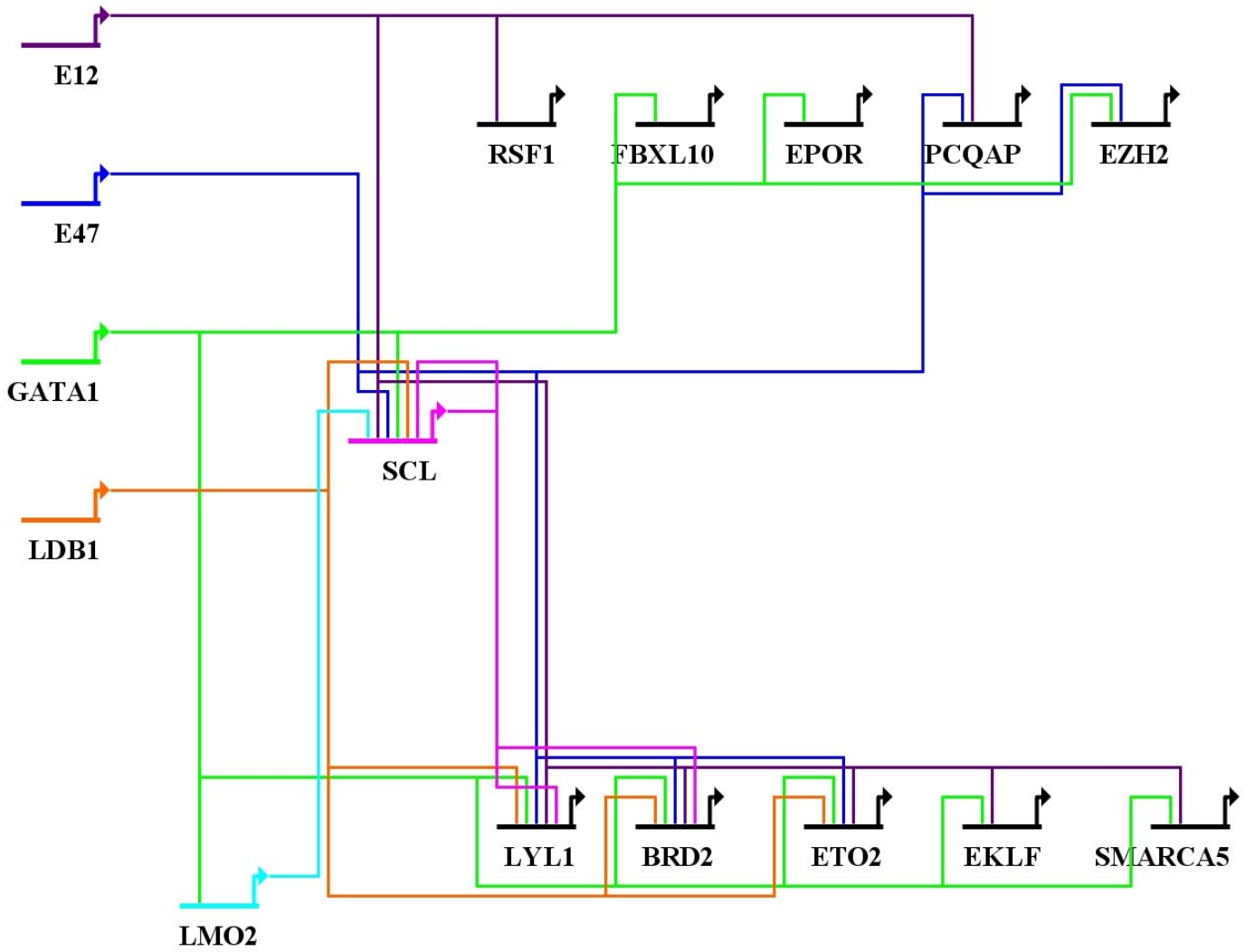
6.3.2.2 Integration of ChIP-qPCR data into networks based on ChIP-on-chip interactions

Fourteen of the target genes in the ChIP-on-chip analyses were further validated and characterised by ChIP-qPCR in K562 and HEL cells (Chapter 5, section 5.4.4.3). Additional TF binding events identified by ChIP-qPCR in K562 were incorporated into the network diagram for K562 (Figure 6.12 A). ChIP-qPCR data derived for HEL cells is shown in the network diagram in Figure 6.12 B. Since it was shown that approximately 60% of the TF-target interactions were in common between K562 and HEL (Chapter 5 section 5.4.4.3 D), the interactions detected were likely to be biologically relevant and reflect the erythropoietic lineage *in vivo*. Thus, a network diagram was also produced to reflect all of the interactions detected by ChIP-on-chip or ChIP-qPCR in either K562 or HEL as a representation of the erythroid lineage (Figure 6.12 C). The ChIP-qPCR provided additional information of transcription factor-promoter binding, than that obtained from ChIP-on-chip, as indicated by the number of novel linkages shown between the TFs and the targets in the diagrams of Figure 6.12.

A) Network based on ChIP-on-chip and ChIP-qPCR data in the K562 cell line



B) Network diagram based on ChIP-qPCR data in the HEL cell line



C) Network diagram based on ChIP-on-chip and ChIP-qPCR data in either K562 or HEL

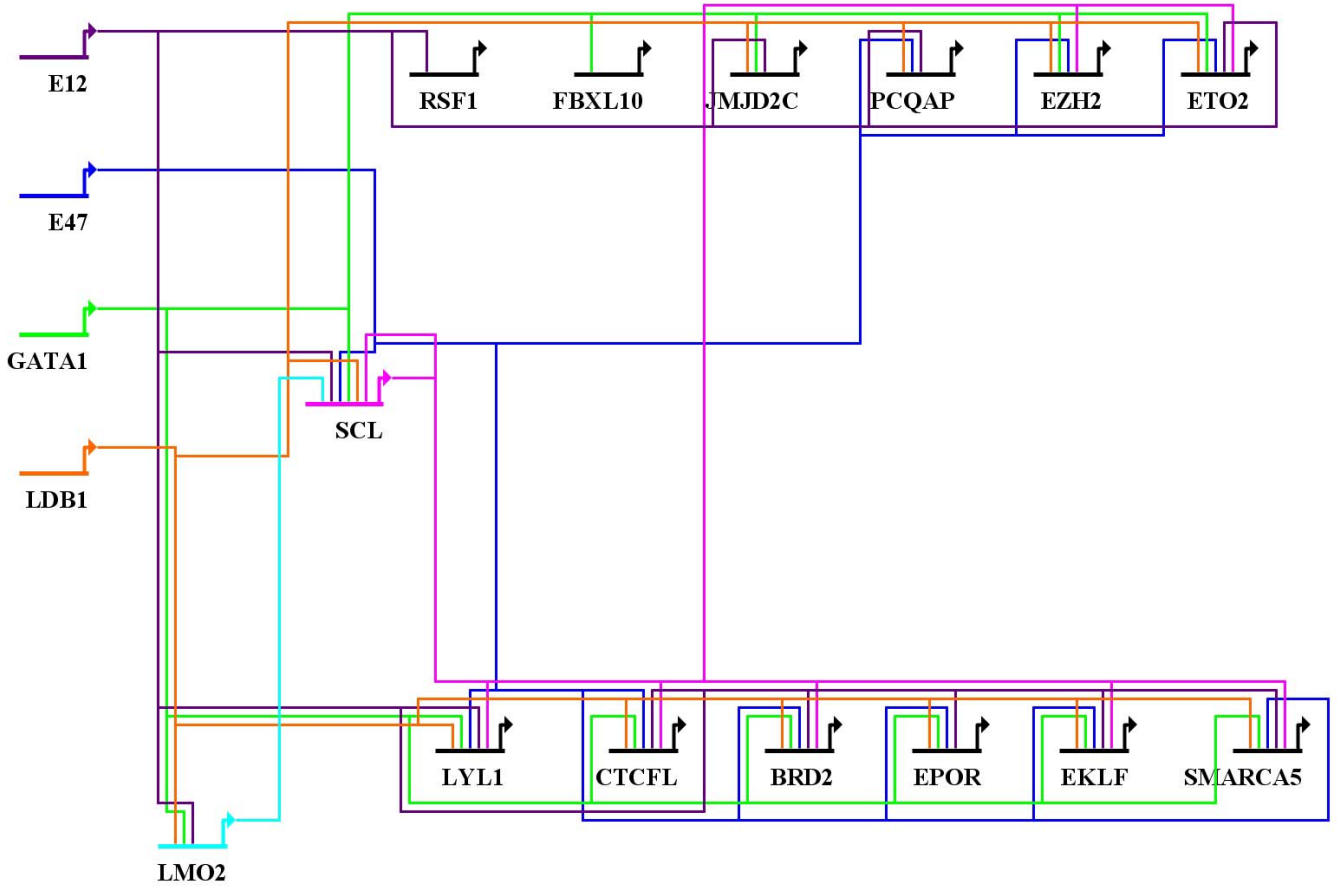


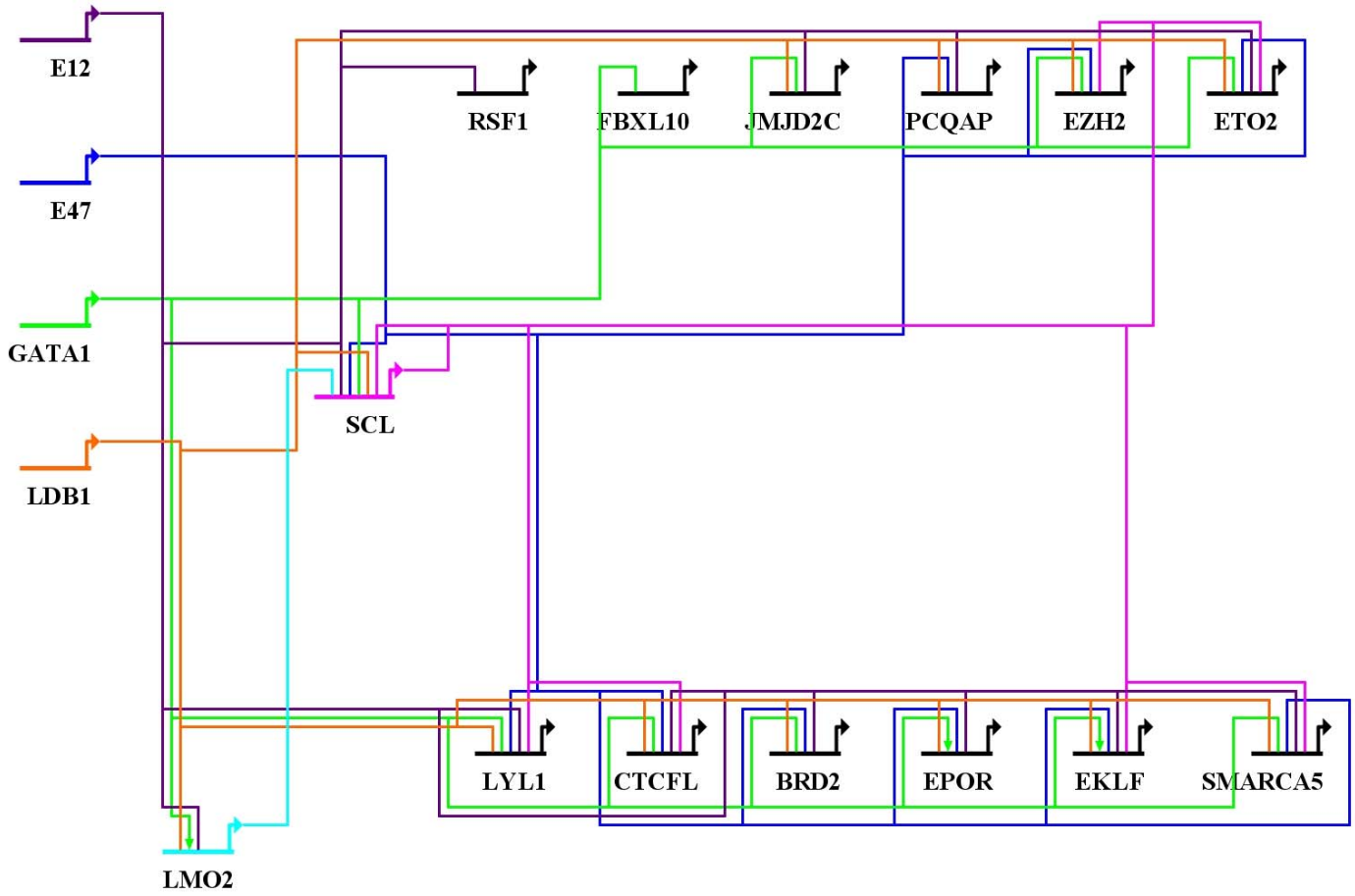
Figure 6.12. Network diagram of promoter-TF interactions of members of the SCL erythroid complex (GATA1, SCL, E12, E47, LDB1 and LMO2) based on ChIP-on-chip and ChIP-qPCR in the K562 and HEL cell lines. Coloured bars on the left represent the 6 members (GATA1, SCL, E12, E47, LDB1 and LMO2) in the complex under study while the black bars on the right shows the targets. Note that LMO2 and SCL are also target genes. The coloured lines show the binding events between the TF and the target gene promoters; SCL: pink, GATA1: green, E12: violet, E47: blue, LDB1: orange, LMO2: aqua. Panel A: the network in the K562 cell line; panel B: the network in the HEL cell line. Panel C: the putative network in the erythroid lineage (K562 and HEL cells data combined).

6.3.2.3 Integration of expression information into interaction networks

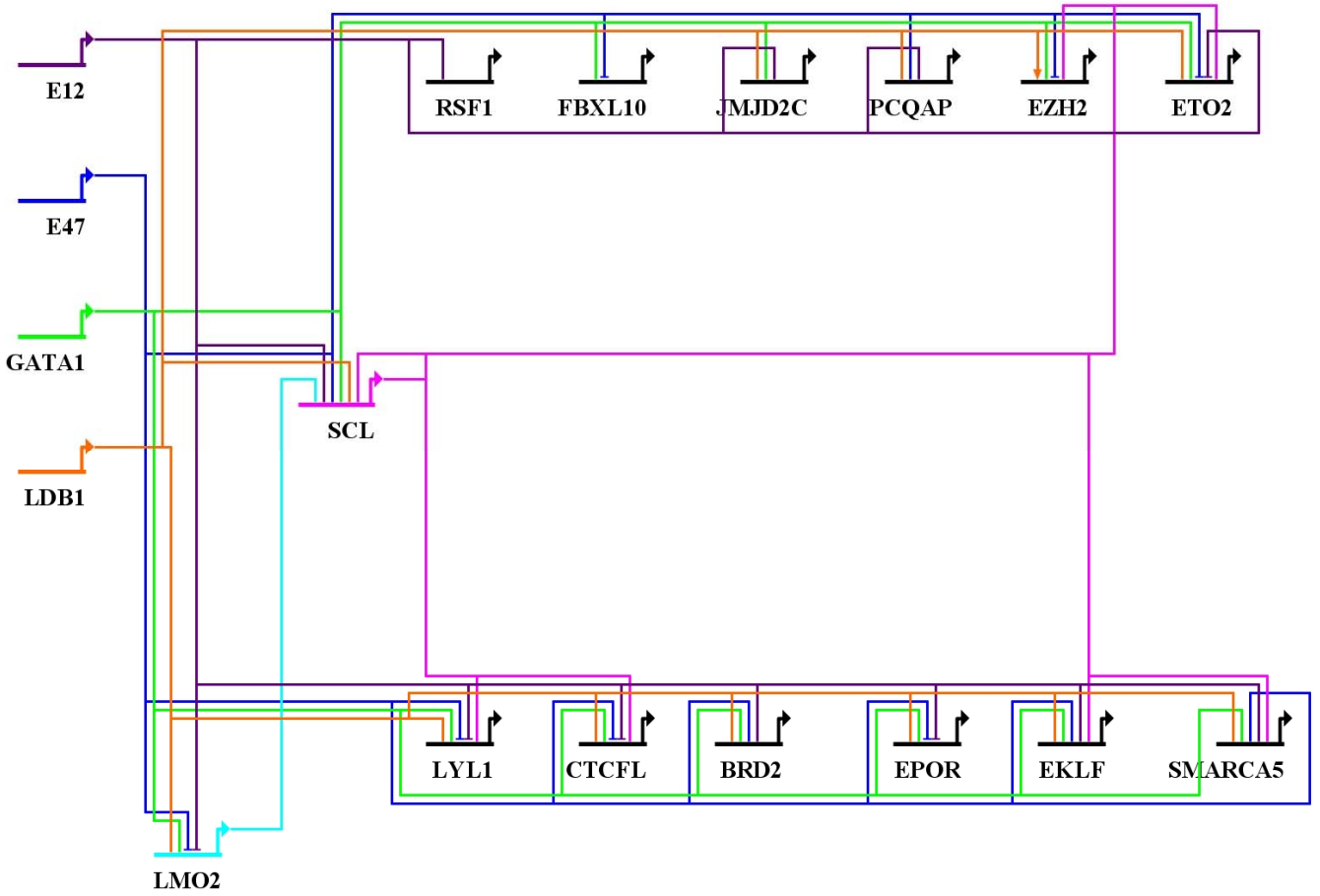
Since TF-target interactions had been confirmed at the level of ChIP-on-chip and ChIP-qPCR for the 14 targets of members of the SEC (see Chapter 5, section 5.4.4.4), it was possible to further elaborate the networks with data showing the mode of regulation of these targets (activation or repression). Thus, the siRNA-mediated TF knockdown data for these targets derived from both the Affymetrix GeneChip knockdown studies and the qPCR time-course studies (section 6.3.1.2) were incorporated into network analysis. This would provide a combined "cause and effect" network for the SCL erythroid complex, based on data described in this thesis. Initially, the Affymetrix GeneChip expression data for these target genes was integrated (Figure 6.13 A). This allowed the activation and repression information for each putative target gene to be determined based on the Affymetrix GeneChip study alone. It should also be noted that expression data was only included for direct TF-target interactions and downstream secondary effects manifested at the level of expression were not shown. Similarly, network diagrams were derived for these 14 target gene interactions with the incorporation of the qPCR time-course knockdown expression data at different time points (12, 24, 36, 48 hour) after knockdown (Figures 6.13 B, C, D, E). The cut-off fold increase or decrease in expression as described in section 6.3.1.2 was used to determine activation or repression status.

The networks generated using the Affymetrix GeneChip data and the expression time course data at the 24 hour time point were not the same (Figure 6.13 A and C). Only EKLF was found to be regulated in both studies. EPOR and LMO2 were shown to be activated by GATA1 in the Affymetrix GeneChip data but not in the expression time course. Conversely, EPOR, JMJD2C and ETO2 were shown to be activated by LDB1 while FBXL10 was shown to be repressed by GATA1 only in the expression time course.

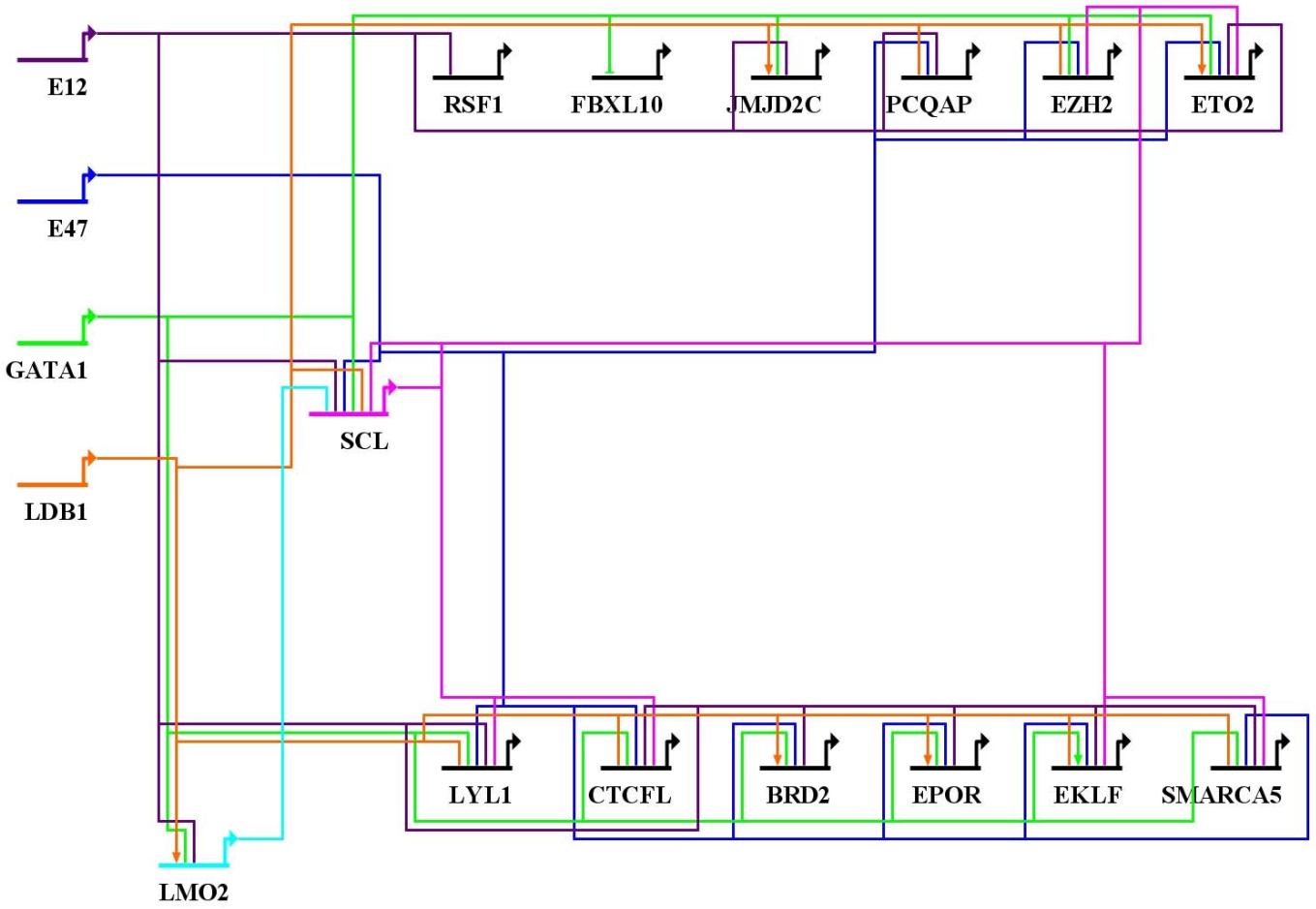
A) Network diagram based on ChIP-on-chip, ChIP-qPCR and Affymetrix expression study at the 24 hour time point in the K562 cell line



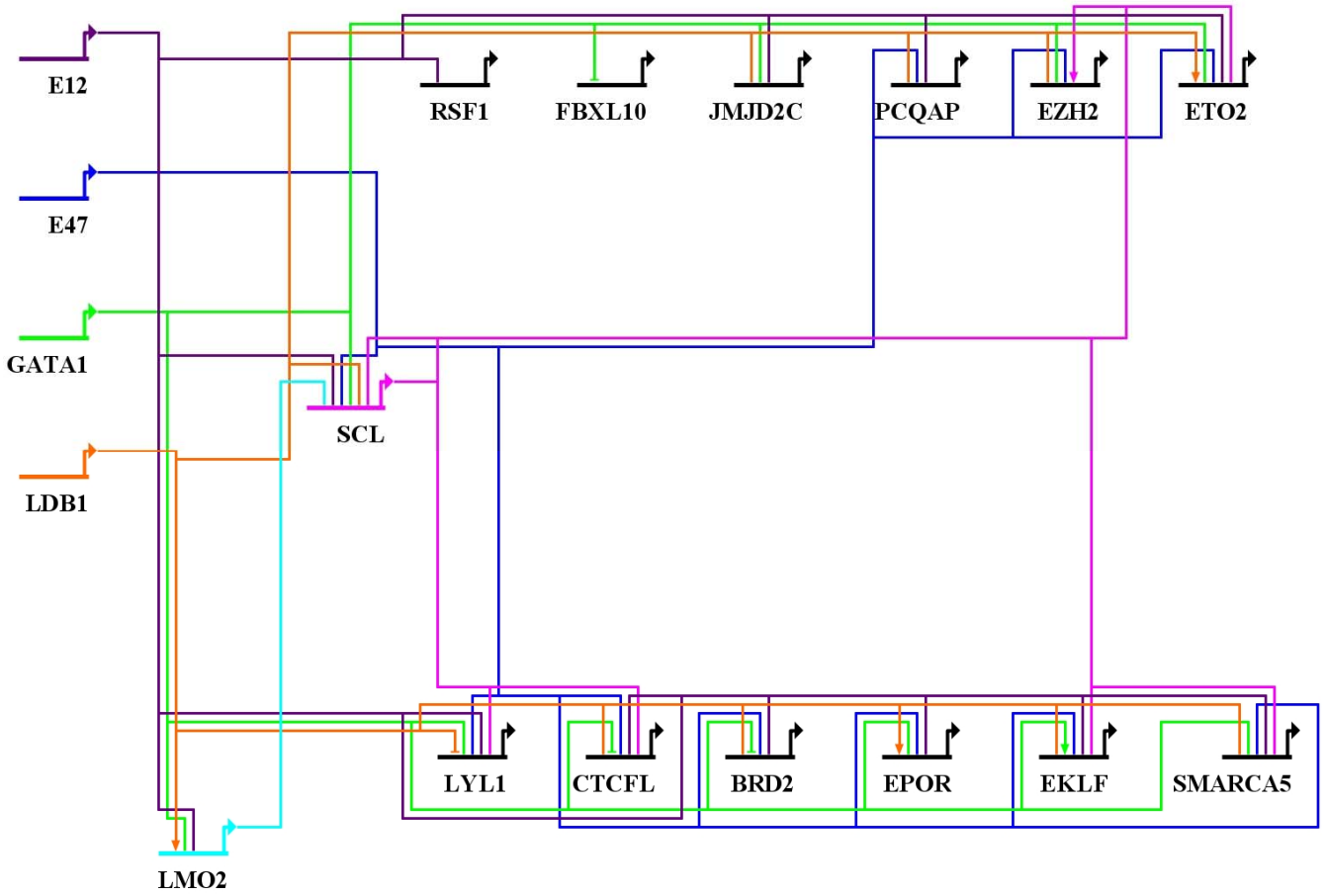
B) Network diagram based on ChIP-on-chip, ChIP-qPCR and time-course expression study at the 12 hour time point in the K562 cell line



C) Network diagram based on ChIP-on-chip, ChIP-qPCR and time-course expression study at the 24 hour time point in the K562 cell line



D) Network diagram based on ChIP-on-chip, ChIP-qPCR and time-course expression study at the 36 hour time point in the K562 cell line



E) Network diagram based on ChIP-on-chip, ChIP-qPCR and time-course expression study at the 48 hour time point in the K562 cell line

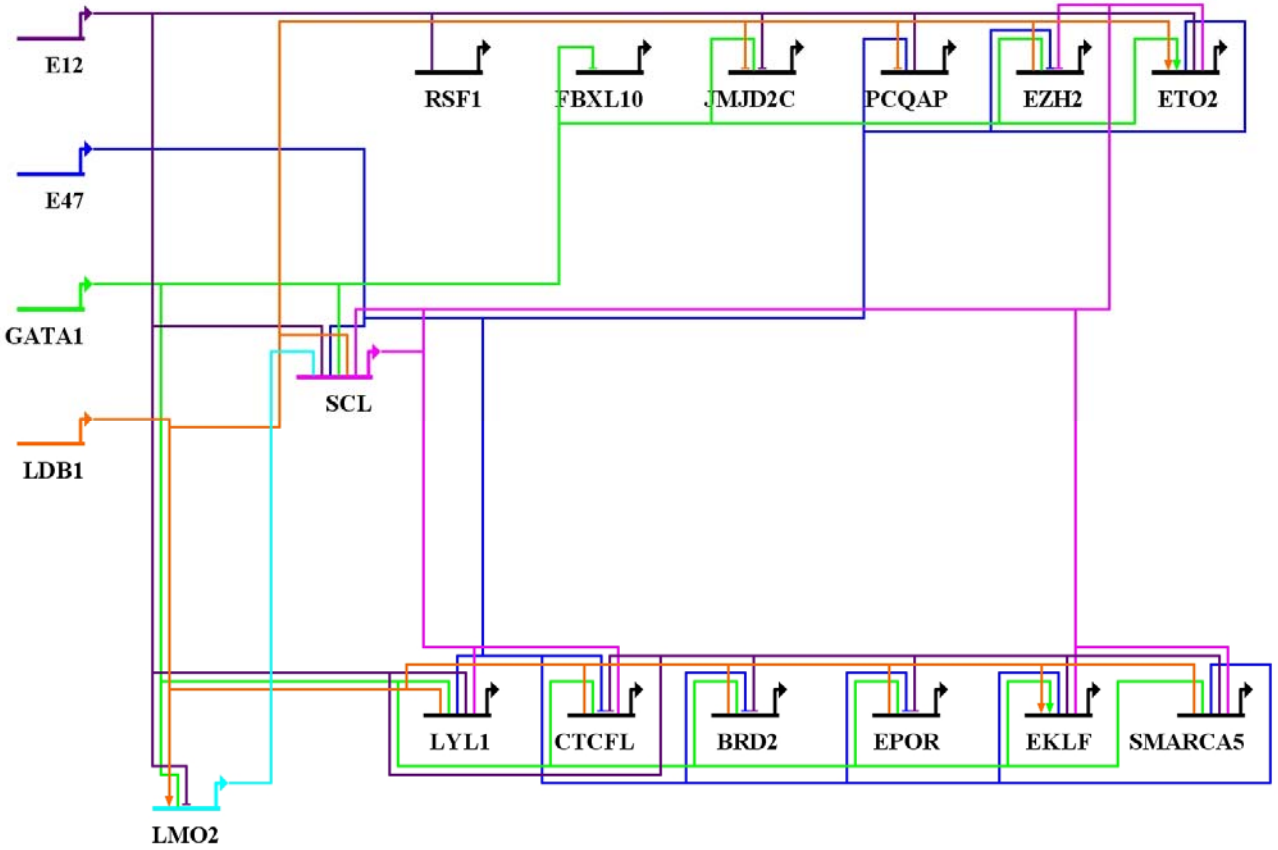


Figure 6.13. Network diagram of promoter-TF interactions of members of the SCL eyrthroid complex (GATA1, SCL, E12, E47, LDB1 and LMO2) based on ChIP-on-chip, ChIP-qPCR, Affymetrix GeneChip and expression time-course in the K562 cell line. Coloured bars on the left represent the 6 members (GATA1, SCL, E12, E47, LDB1 and LMO2) in the complex under study while the black bars on the right shows the targets. Note that LMO2 and SCL are also target genes. The coloured lines show the binding events between the TF and the target gene promoters; SCL: pink, GATA1: green, E12: violet, E47: blue, LDB1: orange, LMO2: aqua. The arrow head of these coloured lines indicate activation while the dash indicates repression. Panel A: network integrating ChIP-on-chip, ChIP-qPCR and Affymetrix GeneChip. Panel B: network integrating ChIP-on-chip, ChIP-qPCR and expression time-course at the 12 hour time point. Panel C: network integrating ChIP-on-chip, ChIP-qPCR and expression time-course at the 24 hour time point. Panel D: network integrating ChIP-on-chip, ChIP-qPCR and expression time-course at the 36 hour time point. Panel E: network integrating ChIP-on-chip, ChIP-qPCR and expression time-course at the 48 hour time point.

6.3.2.4 Networks generation with an integration of all experimental studies

To consolidate the network analysis for these 14 target genes in the K562 cell line, all ChIP interaction data (ChIP-on-chip and ChIP-qPCR) and all expression data (either from Affymetrix GeneChip or qPCR expression analysis) were integrated as shown in Figure 6.14. Direct TF-target interactions were defined as an interaction confirmed by either ChIP-on-chip or ChIP-qPCR. Up or down regulation status for any TF-target interaction was shown if an expression change above the cut-off was found in either the Affymetrix experiment or the qPCR time course study. Furthermore, expression information for this network was obtained from the 24 hour time points after siRNA-mediated knockdown of each TF. This would avoid incorporating effects related to transfection (at the 12 hour time-point) and off-target RNAi effects (36 or 48 hour time points) in the biological information used to elaborate the network. This network diagram clearly shows the multiple interactions obtained for the 8 genes (EZH2, ETO2, CTCFL, LYL1, BRD2, SCL, ELKF, SMARCA5) considered targets of the whole SCL erythroid complex.

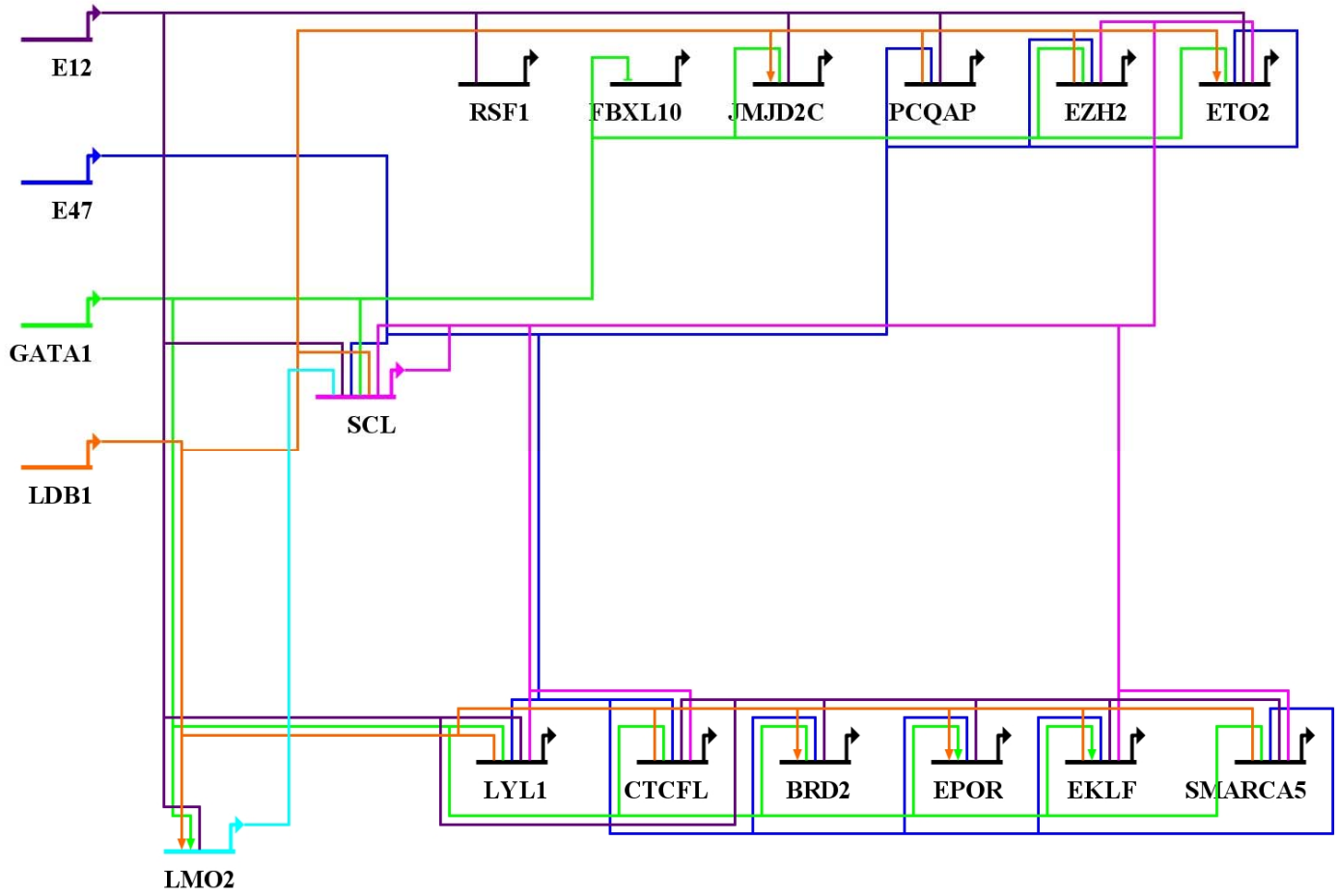


Figure 6.14. Network diagram of promoter-TF interactions of members of the SCL eyrthroid complex (GATA1, SCL, E12, E47, LDB1 and LMO2) based on ChIP-on-chip, ChIP-qPCR, Affymetrix GeneChip and expression time-course at the 24 hour time point in the K562 cells. Coloured bars on the left represent the 6 members (GATA1, SCL, E12, E47, LDB1 and LMO2) in the complex under study while the black bars on the right shows the targets. Note that LMO2 and SCL are also target genes. The coloured lines show the binding events between the TF and the target gene promoters; SCL: pink, GATA1: green, E12: violet, E47: blue, LDB1: orange, LMO2: aqua. The arrow head of these coloured lines indicates activation while the dash indicates repression.

6.3.2.5 Identification of network motifs

As mentioned in Chapter 1, a transcription network can consist of different types of network motifs which are combined and interlinked together to control gene expression (Chapter 1, section 1.2.1). In the transcription network centered around the regulation mediated by the SCL erythroid complex in K562 cells, a number of different types of these network motifs were identified based on the transcription factor-promoter binding events (Figure 6.15). These network motifs describe aspects of the overall network shown in Figure 6.14. Auto-regulation was observed for SCL (Figure 6.15 A). Feed forward loops were demonstrated for six genes. One example was that of GATA1 which regulated SCL which in turn regulated ETO2, EZH2, CTCFL, SMARCA5, EKLF and LYL1, while GATA1 itself also regulated these target genes (Figure 6.15 B). Fifteen regulator chains were identified - one example is illustrated by GATA1, LMO2 and SCL which were regulated in a series resulting in the regulation of ETO2 (Figure 6.15 C). A multiple input motif was observed where GATA1, SCL and LDB1 all worked together to control the expression of SMARCA5, LYL1 and ETO2 (Figure 6.15 D). A dense overlapping region was identified where different combinations of the 4 TFs GATA1, E12, LDB1 and E47 regulated FBXL10, JMJD2C, PCQAP and RSF1 (Figure 6.15 E). An example of single input motif was shown for GATA1 which activated SMARCA5, EPOR, ETO2, LYL1 and EKLF (Figure 6.15 F). Thus, the intricacies and multiple aspects of regulation which relate to the SCL erythroid complex, which are not apparent within any one dataset, can be deciphered by integrating datasets using network building software.

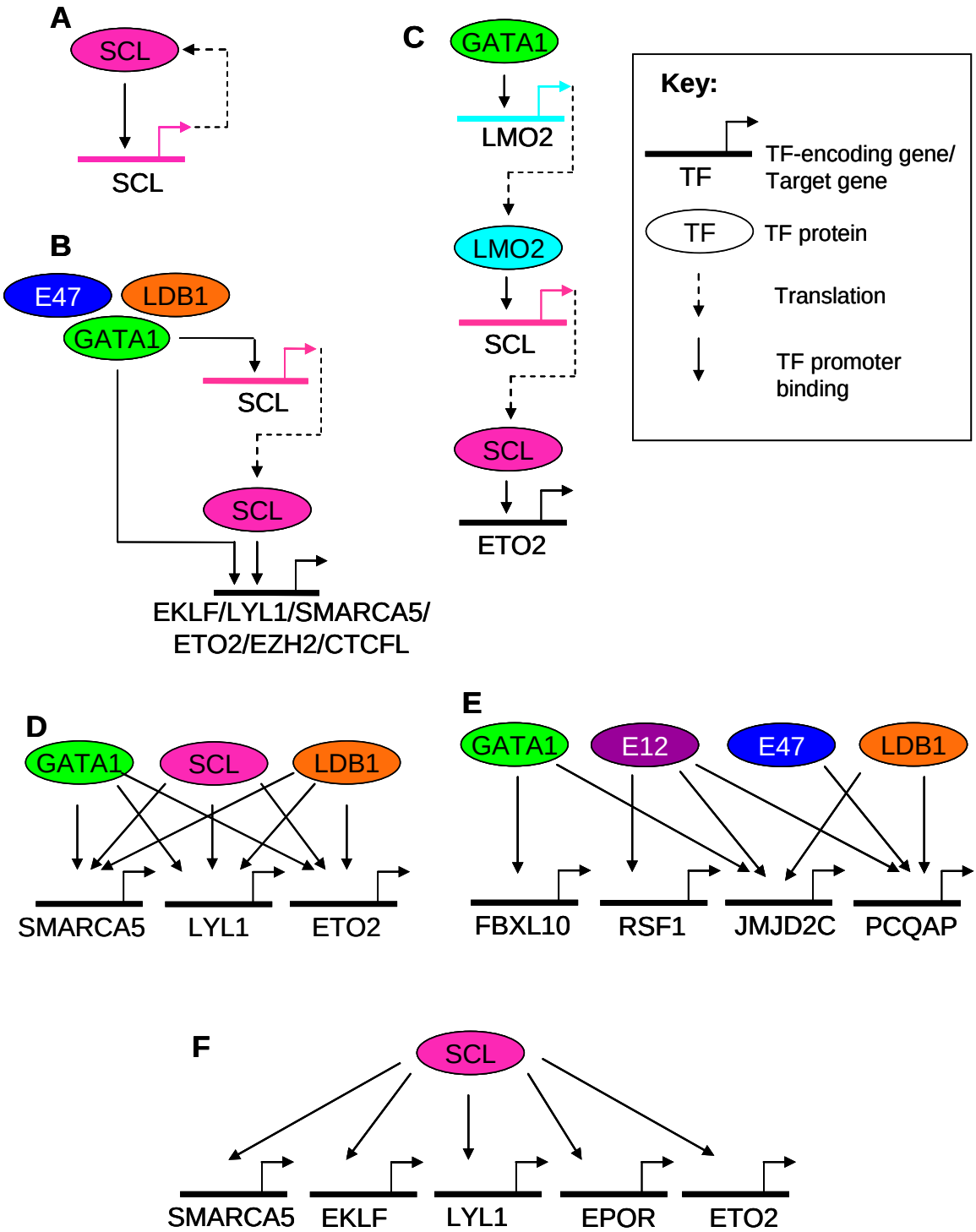


Figure 6.15. The complexities of regulation involving the SCL erythroid complex: Network motifs identified in the SCL erythroid complex transcription network in the K562 cell line. Combinations of these motifs regulate the expression patterns of target genes in the transcriptional network. The symbols and arrows descriptions are included in the key. A: auto-regulation motif. B: feed forward loop. C: regulator chain. D: multiple input motif. E: dense overlapping region. F: single input motif.

6.4 Discussion

The studies performed in this Chapter address a number of fundamental questions which describe the relationship between binding of the SCL erythroid complex to promoters and gene expression. Furthermore, the data in this Chapter helped in the interpretation of complex expression and TF binding datasets, as a means of understanding some of the issues associated with different types of data generated by different experimental approaches. The removal of transcription factor binding at promoters by siRNA knockdown was also studied to further characterise the relationship between the expression profiling by siRNA knockdown and transcription factor binding. Off-promoter binding study by transcription factors was another aspect of this Chapter to illustrate the fact that transcription factor binding is not limited to promoters. Auto-regulation of the SCL erythroid complex was further characterised in order to show that knockdown of any one member of the complex can impinge on the expression of the other members. Finally, the data obtained in this Chapter and previous Chapters were integrated to generate a network of transcriptional regulation in haematopoietic development by the SCL erythroid complex.

6.4.1 Low correlation rate between Affymetrix and ChIP-on-chip studies

A cross comparison between the data obtained in the Affymetrix analyses and the ChIP-on-chip study revealed that only a small portion of genes were found to be overlapped between these two studies (Section 6.3.1.1). A number of reasons may explain why there is low correlation between the datasets.

- **Time point studied**

The Affymetrix study only included the expression changes of one time point during the knockdown where an optimal silencing effect was observed. However, from the results obtained in the expression time-course study of the putative ChIP target genes during the knockdown of members of the SCL erythroid complex (section 6.3.1.2), many changes in expression level of the target genes occurred at other time points rather than the time point used for the Affymetrix analysis (only 27% of changes occurred at this time point). It is also possible that the change in expression of these target genes occurs immediately after the induction of knockdown and therefore they were not identified on the Affymetrix array.

- **Poised regulation**

The ChIP-on-chip analyses allowed us to identify the binding of a TF to the *cis*-regulatory elements but did not provide information about whether the factor actually regulates the genes nearby the regulatory elements. In fact, the recruitment of a TF may not necessarily correlates with the transcriptional control of its target genes. In a study of the mapping of NF-kappaB binding sites

along chromosome 22, binding was observed near a substantial number of genes whose expression was not regulated (Martone et al., 2003). These observations suggested that the recruitment of other transcription factors or co-factors, together with the chromatin modifiers, may be required to achieve a combinatorial effect on transcriptional regulation. Genes will only be expressed at a certain stage of development when all these co-factors and transcription factors are expressed and are recruited to the sites of regulation in the genome.

- **Auto-regulation of the SCL erythroid complex**

Various pieces of evidence from the Affymetrix analyses, ChIP-on-chip study and the expression time-course study all suggested auto-regulatory aspects for members of the SCL erythroid complex (see section 6.4.4). This adds to the complexity of the expression changes that may occur when one TF of the complex is perturbed – often resulting in perturbation or up-regulation of other members of the complex. This illustrates that expression of each member of the SCL erythroid complex is tightly controlled. One reason for this is that their target genes are maintained at a constant expression level even if one member of the complex is perturbed. As a result, no obvious expression changes may be detected.

- **Off-promoter regulation**

Since the TF promoter array used here had only coverage of 1 kb of the promoter regions/transcription start sites of genes, binding events occurring at other regulatory elements would not detected. The characterisation of binding events at the putative enhancers of SCL and LYL1 in section 6.3.1.5 illustrated that binding of a TF is not restricted to promoter regions. This was also shown in the mapping of NF-kappaB binding sites along chromosome 22 (Martone et al., 2003). Taken together, it is possible that a number of target genes picked up in the expression studies were regulated by TF of the SEC in *cis*-elements located outside of the promoters.

- **Indirect targets**

The Affymetrix expression analyses identified both direct target genes and secondary target genes downstream in transcriptional cascades whereas the ChIP-on-chip analyses identified only direct target genes. Depending on the role that the SCL eyrthoid complex has on transcriptional programmes (and the speed at which changes in these programmes can occur when the complex is perturbed), the number of secondary targets identified after 24 hours of siRNA-mediated knockdown may be very high. Furthermore, since magnitudes of expression changes do not correlate with whether the target is direct or secondary, top-scoring Affymetrix hits may not necessarily be direct targets. For both of these reasons, it is possible that a substantial proportion of top-scoring hits in the Affymetrix datasets could be secondary targets.

- **Effect of knockdown of TF on binding efficiency**

Although the siRNA-mediated knockdown of TFs was more than 70% of the original mRNA/protein levels in most cases, binding to DNA may not diminish in response to the knockdown - as demonstrated in the ChIP-on-chip study of GATA1. This would suggest that only a small proportion of the TF may be bound at any one time in the nucleus. In addition, the protein may need time to dissociate from their protein or DNA binding partners and degrade after the RNAi trigger is induced. The kinetics of these events may be somewhat different from the kinetics of knockdown of all of the unbound protein in the nucleus. Furthermore, there was not a correlation between GATA1 clearance from promoters and changes in gene expression – further supporting the idea that changes in GATA1 binding may not necessarily result in changes in expression of targets. Finally, electroporation itself was shown to perturb TF binding to some degree – this off-target effect would again reduce the correlation between GATA1-bound targets in wild type K562 cells, and expression changes observed in siRNA-mediated GATA1 knockdowns.

- **Microarray technology platform**

As described in Chapter 4 section 4.4.3.3, the validation rate of the differentially-expressed genes identified on the Affymetrix GeneChip by qPCR was only 37%. This indicates a high proportion of false positive targets may have been identified and is probably another reason why a low correlation was observed between the Affymetrix and ChIP-on-chip data.

- **Using siRNA knockdown to study expression**

A complete loss of the transcription factor cannot be achieved by transfecting siRNA into cells. Thus, variations in the amount of TF remaining in cells used in different replicate experiments may induce different effects on gene expression. Furthermore, as has been mentioned above, changes in target gene expression may be minimal when there is an incomplete gene knockdown due to binding of the remaining protein to its normal sites of regulation. Thus, a better way of studying downstream regulation by TFs may be to use an inducible knockout system.

6.4.2 The SCL complex transcription network

The whole SCL erythroid complex has previously been shown to regulate only three genes in human haematopoiesis: c-kit (Vitelli et al., 2000), α-globin (Anguita et al., 2004) and glycophorin A (Lahlil et al., 2004). Although researchers have tried to build a gene regulatory network of the erythroid lineage using published literature in mouse (Swiers et al., 2006), a genome-wide scale experimental study of the SCL complex in erythroid development in human has not been reported in the literature. The studies performed in this thesis provided further insights into the network of genes regulated by the SCL erythroid complex in K562 cells. Given that some of these targets were

also validated at the level of ChIP-qPCR in a second eyrthroid cell line (HEL), it is likely that this network is representative of events which occur in the human erythroid lineage. This project represents the first integrated approach to delineate the regulatory network controlled by the SCL erythroid complex on a genome-wide scale.

Eight additional direct target genes were identified for the whole SCL erythroid complex in this study where four of them are related to haematopoietic development (EKLF, ETO2, LYL1 and SCL) and the other four are related to chromatin structure, modification or remodelling (BRD2, CTCFL, SMARCA5 and EZH2) (see Chapter 5, section 5.5.5). According to the network diagram in Figure 6.15, EPOR, EKLF, BRD2 and ETO2 are likely to be activated while the mode of regulation for the others is unknown. The auto-regulatory role of the SCL erythroid complex as described in section 6.4.1 is likely to help modulate the expression of these target genes, which in turn control the expression of other erythroid-specific genes by direct transcription factor binding or chromatin remodelling in the erythroid lineage. The relationship between the network of genes controlled by the SCL erythroid complex and the wider network of gene regulation in erythroid development awaits further study.

6.5 Conclusions

The results presented in this Chapter illustrated the relationship between transcription factor binding and *cis*-regulatory elements, as well as the effect of knockdown on binding and release. The autoregulatory role of the SCL erythroid complex was further characterised. Despite the complexity of the datasets, a transcription network integrating all studies in this and previous Chapters was generated for the SCL erythroid complex for the first time in human erythroid cells.

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.
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.
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.
Longabaugh, W.J., Davidson, E.H., and Bolouri, H. (2005). Computational representation of developmental genetic regulatory networks. Developmental biology *283*, 1-16.
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.
Oda, K., Kimura, T., Matsuoka, Y., Funahashi, A., Muramatsu, M., and Kitano, H. (2004). Molecular Interaction Map of Macrophage. AfCS Research Reports *2*.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.
Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Dev Biol *294*, 525-540.
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.
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.
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.