Chapter 7

Further Characterization of Regulatory Regions From Across the SCL Locus

7.1 Introduction

In chapter 6, a consensus histone code was proposed for the SCL region. This code suggested that regulatory regions across the SCL locus were associated with a set of histone modifications which could define their function and activity in K562 cell line. In chapter 5, ChIP-chip datasets were also described for various histone modifications in three other cell lines: Jurkat, HL60 and HPB-ALL. The purpose of surveying the distribution of histone modifications in these cell lines was to elucidate differences and/or similarities and relate these results with the expression and regulation of SCL accordingly.

Understanding differences in the regulation of SCL in various cell types is important for understanding SCL biology in both normal and disease states. This is particularly important where inappropriate expression of SCL in T-cells results in T-cell acute lymphoblastic leukaemia (T-ALL). In cases of T-ALL, the inappropriate expression of SCL has most often been linked to chromosomal rearrangements (deletions or translocations) at the 5' or 3' end of the SCL gene (Bernard et al. 1991; Brown et al. 1990; Xia et al. 1992). It has also been reported that several established leukaemic T-cell lines, such as Jurkat, exhibit high levels of SCL expression in the absence of any apparent SCL genomic alteration (Bernard et al. 1992; Leroy-Viard et al. 1994). In Jurkat, transcription is mainly initiated from SCL p^{1b}, and to a much lesser extent from p^{EXON4}. The expression was found to be mono-allelic (Leroy-Viard et al. 1994), thus suggesting that a *cis*-acting mechanism might be responsible for activating the single allele. However, to date, the mechanism of SCL activation in Jurkat is not known.

Of the two DNase I hypersensitive sites identified in Jurkat at the SCL locus (Leroy-Viard et al. 1994), the SCL -7 region exhibits a strong DNase I sensitivity and was found to be located near one of the SCL breakpoints reported in SIL-SCL deletions resulting in T-ALL (Breit et al. 1993); this suggests that this genomic region might be implicated in the *cis*-activation of SCL transcription (Leroy-Viard et al. 1994). However, it is also possible that in Jurkat (or other T-cell lines with no genomic rearrangements at the SCL locus),

other regulatory regions at the SCL locus, may be involved in the inappropriate expression of SCL.

This further emphasizes the fact that, in order to fully understand the regulation of SCL in different cell types, it is important to elucidate and characterize the full complement of regulatory regions across the SCL locus - with respect to their activity, function, role in transcriptional regulation of SCL and their associated DNA-protein interactions. To this end, the results presented in this chapter provide additional clues to the biological activity of these regulatory elements.

7.2 Aims of this chapter

The aims of the work reported in this chapter were:

To provide characterizations of known regulatory regions across the SCL locus with respect to data obtained from this study and from other sources where appropriate.

i) To test the proposed histone code with respect to the known regulatory regions, with known activities and functions across the SCL locus in SCL expressing and non-expressing cell lines (i.e. K562, Jurkat, HL60 and HPB-ALL).

ii) To further characterize novel regions identified at the SCL locus.

iii) To identify regulatory elements involved in the *cis*-activation of the SCL gene in Jurkat.

Results

7.3 Generating TreeView visualization profiles for the four cell lines

ChIP-chip datasets for the seven key histone modifications of the SCL histone code (Chapter 6) were used to generate TreeView (Eisen et al. 1998) visualization profiles for the four human cell lines K562, Jurkat, HL60 and HPB-ALL (similar to those shown in Figure 6.2). For the purposes of the work presented in this chapter, these TreeView profiles depicted the various histone modification with respect to their genomic distribution across the SCL locus and were not clustered. In summary, the key histone modifications included were:

i) H3 K79 trimethylation (H3 K79Me3),

ii) H3 K4 monomethylation (H3 K4Me1),

iii) H3 K4 dimethylation (H3 K4Me2) (two datasets were included for this histone modification),

iv) H3 K4 trimethylation (H3 K4Me3),

v) H3 K9 dimethylation (H3 K9Me2) (this dataset was not available for HL60),

vi) H4 K16 acetylation (H4 K16Ac), and

vii) H3 K9 acetylation (H3 K9Ac).

In chapter 5, the assay for H3 K9 acetylation was shown not to yield significant enrichments in Jurkat, suggesting that there may be a cell type specific difference with respect to this histone modification (see section 5.6.3.1). Similarly, H3 K9 dimethylation did not show enrichments across the cytochrome P450 genes to indicate the location of regions of silent chromatin (data not shown). Given that the other three cell lines analyzed here all showed H3 K9 acetylation and dimethylation profiles consistent with the proposed histone code, this suggests that Jurkat may not utilise modifications at lysine 9 residue of histone H3 in the same way as other cell types. However, for consistency, data for both of these ChIP-chip assays are included in the Treeview profiles shown in this chapter. These visualizations are described with respect to each of the regulatory regions across the SCL locus in the subsequent sections of this chapter.

7.4 The SIL promoter and the SIL -10 region

In Figure 7.1, TreeView profiles at the 5' end of the SIL gene and the upstream SIL -10 region are shown for the four human cell lines. In the figure, the known and well characterized SIL promoter (Colaizzo-Anas and Aplan 2003) has been annotated as p^{SILa}.

According to the consensus histone code, the SIL p^{SILa} promoter shows features of an active promoter in all the four cell lines. These features include hyper-acetylation of H3 K9, hyper di- and trimethylation of H3 K4, hypo-acetylation of H4 K16, and hypo monomethylation of H3 K4. In addition, trimethylation of H3 K79 was seen at the region immediately downstream of the promoter region in all cell lines. It should be noted that hypo dimethylation of H3 K9 and hyper-acetylation of H3 K9 were not very apparent in Jurkat for reasons discussed above (see section 7.3). In the K562 cell line, binding interactions for p300, CBP, HDAC2 (all three cofactors have HAT and HDAC activity) and $Taf_{II}250$ (component of the pre-initiation complex) were seen at the SIL p^{SILa}, which provided further support for the activity of this promoter in this cell line (see chapter 5, sections 5.13, 5.14).

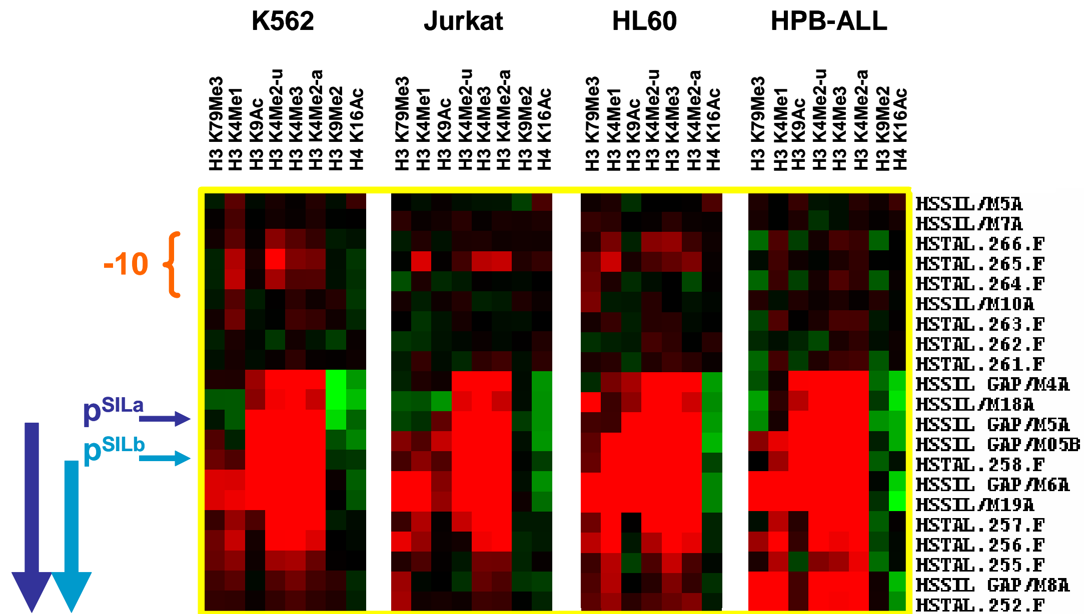

Figure 7.1: TreeView profiles at the 5' end of the SIL gene and the SIL -10 region in four human cell lines. The cell line corresponding to each profile is shown at the top. The well known and characterized SIL promoter (Colaizzo-Anas and Aplan 2003) is annotated as p^{SILa} and another region with potential promoter function has been annotated as p^{SILb}. The thin blue horizontal arrows represent the genomic positions of p^{SILa} and p^{SILb} with respect to the array elements of the human SCL tiling array. The thick blue vertical arrows at the left represent the direction of transcripts originating from p^{SILa} and p^{SILb} (see text). The location of the SIL -10 region is also shown in red at the left of the figure. The profiles were generated using only the key seven histone modifications which are listed along the x-axis at the top of the figure. Two datasets were included for H3 K4 dimethylation in each cell line. The names of the array elements representing the genomic regions are listed along the y-axis (extreme right). Each box in the profiles represents an array element for the corresponding histone modification and the red/black/green colour indicates enrichment/baseline levels/depletion for that modification respectively.

The 5'end of the SIL gene showed enrichments or depletions for various modifications across a region spanning the upstream -1 region to almost +3/+4 region. From the ChIP-chip profiles for various histone modifications, three distinct peaks of enrichments were observed in all four cell lines at -1, p^{SILa} and +1/+2 (the last of which is annotated as p^{SILb}) and are shown in Figure 7.2. The p^{SILb} region was also observed in the TreeView profiles, as a distinct region downstream of p^{SILa}, and exhibited some of the features of being an active regulatory element. These included hyper di- and trimethylation and hypo monomethylation of H3 K4, hypo dimethylation and hyper acetylation of H3 K9, and hypo acetylation of H4 K16. This data suggests that this region may be an alternative SIL promoter. On further investigation, it was found that in the ENSEMBL database (www.ensembl.org), three transcripts for SIL have been annotated - the 5' end of one of these transcripts (annotated as 'SIL' in Figure 7.2, panel E) maps to the p^{SILb} region.

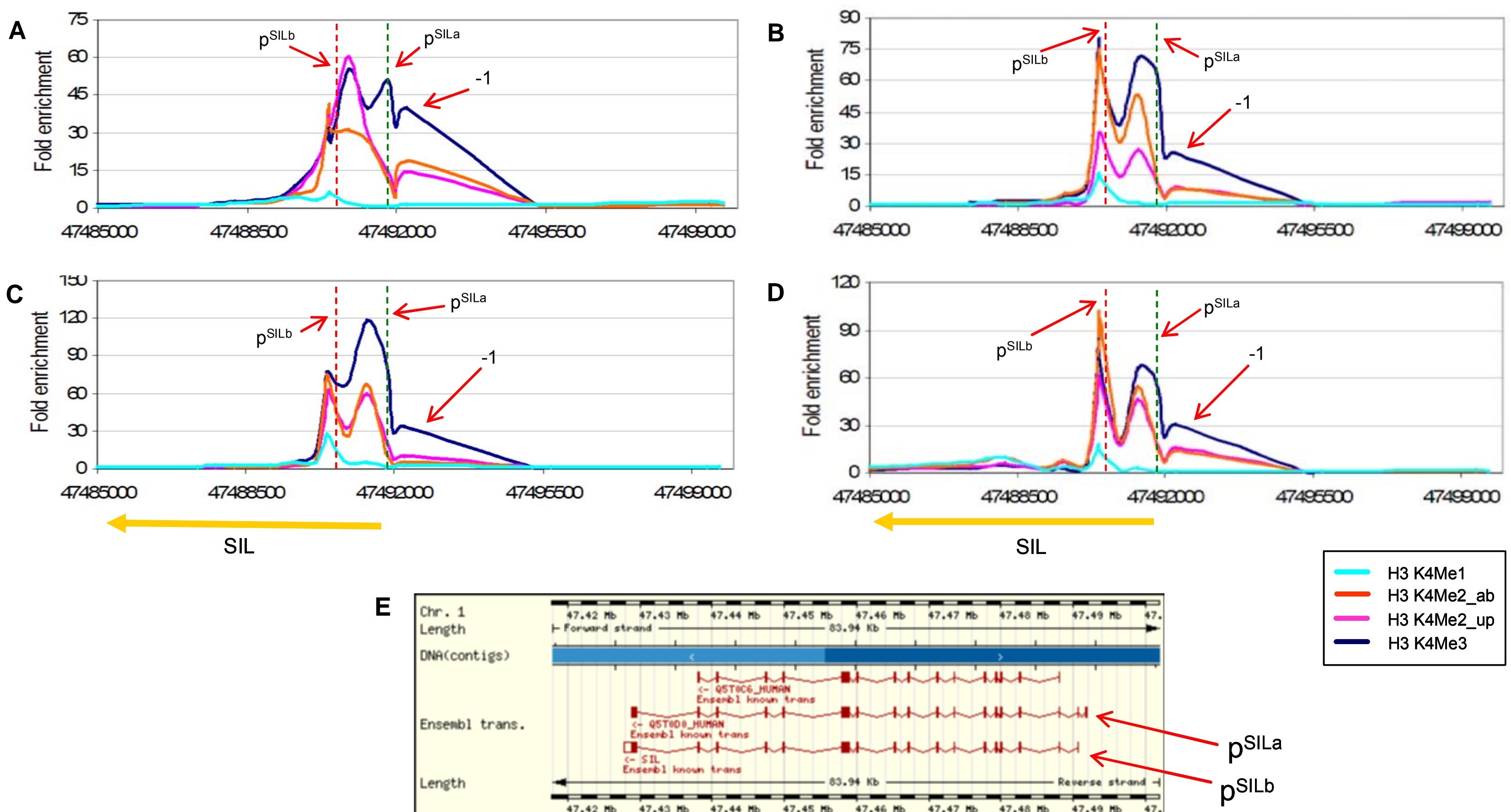

Figure 7.2: Composite ChIP-chip profiles across the 5' end of the SIL gene showing three distinct peaks. Panel A:K562, panel B: Jurkat, panel C: HL60 and panel D: HPB-ALL. The list of histone modifications is shown in a box at the bottom right of the figure. These include mono-, di- and trimethylation of H3K4 and are colour coded as shown in the box. In the profiles, the genomic positions of p^{SILa} and p^{SILb} are marked by green and red dotted lines respectively. It can be clearly seen in the ChIP-chip profiles (panels A-D, turquoise profiles) that the relative enrichments for monomethylation of H3 K4 are lower at p^{SILb} than its immediate downstream region. The yellow coloured arrow at the bottom of profiles B and D represent the SIL gene. The y-axes in the profiles show the fold enrichments and the x-axes represent genomic coordinates along human chromosome 1. Panel E shows the annotated SIL transcripts in ENSEMBL database and are marked with red arrows to show the mapping of p^{SILa} and p^{SILb} as shown in Figures 6.1 and 6.2.

Additionally, in the K562 cell line, binding interactions for CBP, HDAC2 (cofactors with HAT and HDAC activity) and $Taf_{II}250$ (component of pre-initiation complex) were also seen at this region (Chapter 5, sections 5.13, 514). These data provide support for the idea that p^{SILb} may represent an alternative SIL promoter. The histone code data suggests that this promoter may be active in K562. Taken together, these data suggested that the 5' end of the SIL gene contains two promoters, p^{SILa} and p^{SILb}, both of which show hallmarks of being active promoters which are consistent with respect to the consensus histone code in all the four human cell lines.

In addition, a third distinct peak at the SIL -1 region was observed as shown in Figure 7.2. This region is known to bind the transcription factor Elf-1 in K562, Jurkat and HPB-ALL (see chapter 5, section 5.15). This region also showed features of an active regulatory element including hypo-acetylation of H4 K16, hyper di- and trimethylation of H3 K4. Whether this represents a part of the SIL p^{SILa} promoter or a distinct regulatory element is not known, although it does have a histone code of that of a promoter region.

A region approximately 9 to 10 kb upstream of the 5' end of the SIL gene (annotated as SIL -10 and previously identified in chapter 5), showed histone modification marks in K562, Jurkat and HL60 but not in HPB-ALL (Figure 7.1). In the hierarchical clustering analysis of histone modifications across the SCL locus in K562 (Chapter 6, sections 6.8, 6.9), the SIL -10 region clustered with the known enhancers of SCL. The key histone marks displayed by this region, in K562, were consistent with the consensus histone code of an active enhancer element. These marks included hypo-acetylation for H4 K16, and hyper mono-, di- and trimethylation for H3 K4. However, in Jurkat, HL60 and HPB-ALL, this region is likely to be inactive according to its pattern of histone modifications in these three cell lines. These data suggest that a putative novel enhancer region is located at the SIL -10 region.

7.5 The KCY promoter

Figure 7.3 shows Treeview profiles across the 5' end and downstream regions of the KCY gene in K562, Jurkat, HL60 and HPB-ALL. The KCY gene is an ubiquitously expressed gene and thus, the promoter is expected to be active in the cell lines being tested. The profile showed that the genomic region displaying enrichments for various histone modifications extended almost 4 kb into the transcribed sequences. The histone marks present at the promoter region suggested that the KCY promoter is active in all the four cell lines. These marks included hypo monomethylation of H3 K4, hypo-acetylation of H4 K16, hyper di- and trimethylation of H3 K4 and hyper-acetylation of H3 K9. In addition, hyper trimethylation of H3 K79 was present at the regions immediately

downstream of the KCY promoter region. These data for the KCY promoter are in complete agreement with the proposed consensus histone code for the SCL locus. In addition, binding interactions for $Taf_{II}250$ were detected at the KCY promoter region (Chapter 5, section 5.14) in K562, thus providing further data to support that the KCY promoter is active in this cell line.

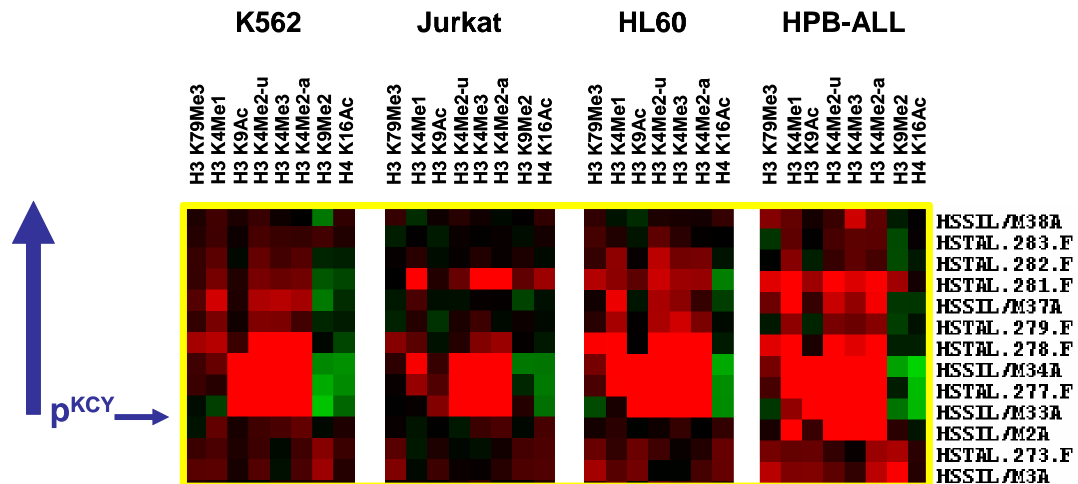

Figure 7.3: TreeView profiles at the 5' end of the KCY gene in four human cell lines. The cell line corresponding to each profile is shown at the top. The genomic position of the KCY promoter p^{KCY} is marked by the blue thin horizontal arrow and the thick blue vertical arrow represents the direction of transcription of the gene. The profiles were generated using only the key seven histone modifications which are listed along the x-axis at the top of the figure. Two datasets were included for H3 K4 dimethylation in each cell line. The array elements representing the shown genomic regions are listed along the y-axis at the extreme right of the figure. The names of the array elements representing the genomic regions are listed along the y-axis (extreme right). Each box in the profiles represents an array element for the corresponding histone modification and the red/black/green colour indicates enrichment/baseline levels/depletion for that modification respectively.

7.6 The SCL promoters and their immediate 3' flanking regions

Figure 7.4 shows the TreeView profiles across the 5' end of SCL gene encompassing the three known promoters (p^{1a}, p^{1b}, and p^{EXON4}) and the regulatory +1 and +3 regions. It is known that in K562, transcription is initiated at both p^{1a} and p^{1b}, whereas in Jurkat, transcription is initiated mostly at p^{1b} with some transcripts originating at p^{EXON4} (Aplan et al. 1990a; Begley et al. 1994; Bernard et al. 1992). Similarly, +1 and +3 are known enhancers which were first identified owing to their hypersensitivity to DNase I and activity in transient and stable reporter assays (Leroy-Viard et al. 1994; Gottgens et al. 1997; Fordham et al. 1999; Sinclair et al. 1999).

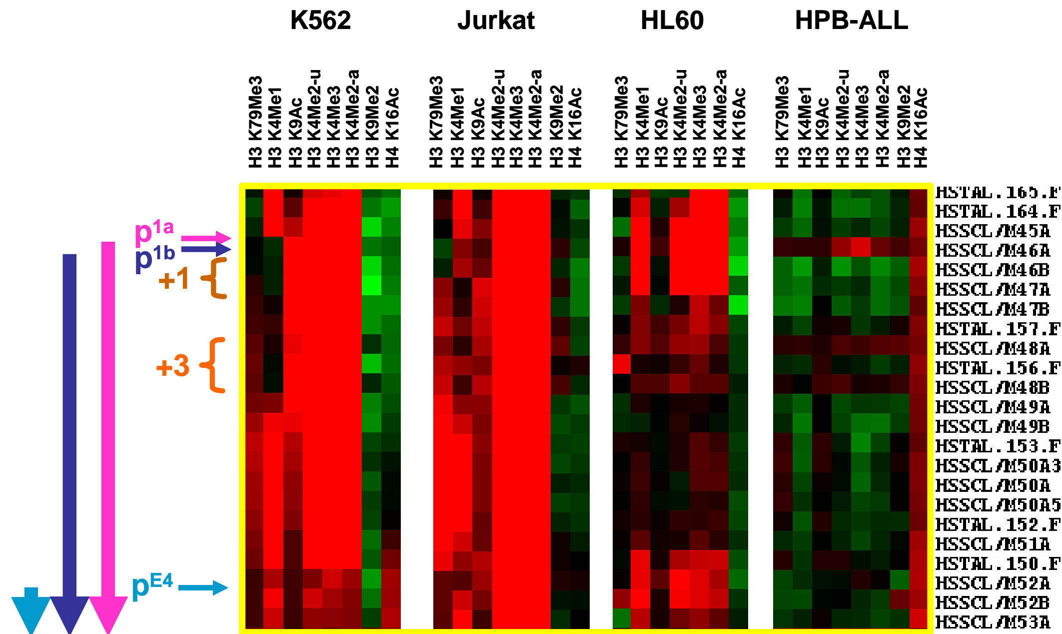

Figure 7.4: TreeView profiles at the 5' end of the SCL gene in four human cell lines. The cell line corresponding to each profile is shown at the top. The promoters, p^{1a}, p^{1b} and p^{EXON4} are marked with thin horizontal coloured arrows (colour coded with respect to the promoters) and the thick coloured vertical arrows represent the transcripts originating at these promoters and the direction of transcription. The +1 and +3 regions are also marked. The profiles were generated using only the key seven histone modifications which are listed along the x-axis at the top of the figure. Two datasets for H3 K9 dimethylation were included for each cell line. The names of the array elements representing the genomic regions are listed along the y-axis (extreme right). Each box in the profiles represents an array element for the corresponding histone modification and the red/black/green colour indicates enrichment/baseline levels/depletion for that modification respectively.

In K562, both p^{1a} and p^{1b} displayed hallmarks of active promoters, whereas the p^{EXON4} displayed histone marks associated with inactive promoters. The histone marks at p^{1a} and p^{1b} include hypo monomethylation of H3 K4, hypo-acetylation of H4 K16, hyper-acetylation of H3 K9 and hyper di- and trimethylation of H3 K4. The genomic region displaying these features extended up to the +3 region. Hyper trimethylation of H3 K79 was seen at the regions immediately downstream of p^{1a} and p^{1b} which is also a feature of active promoters. However, p^{EXON4} region showed hyper monomethylation of H3 K4, hypo-acetylation of H3 K9 and hyper-acetylation of H4 K16. In addition, the relative levels of di- and trimethylation of H3 K4 were very low as compared to that of p^{1a} and p^{1b} (Figure 7.5). It has already been demonstrated in chapter 6, that the tri- to monomethylation ratios for H3 K4 which depict the activity of a promoter more accurately, were found to be very low at p^{EXON4} in K562 (Chapter 6, section 6.5). Thus, the pattern of histone marks present at p^{EXON4} suggests that this promoter is inactive in K562. Binding interactions for GATA-1 were detected at the promoter and +3 regions of SCL in the K562 cell line (Chapter 4, section 4.6.2). GATA-1 binding at the SCL

promoter agrees with the previously known data (Aplan et al. 1990; Bockamp et al. 1995). The cofactors p300, CBP and HDAC2 also showed binding interactions at the SCL promoter in K562 along with $Taf_{II}250$ (Chapter 5, section 5.13, 5.14). All of these data support that SCL p^{1a} and p^{1b} are active but p^{EXON4} is inactive in the K562 cell line.

In Jurkat, p^{1b} and p^{EXON4} displayed hallmarks of active promoters including hypo monomethylation of H3 K4 (as compared to their immediate upstream and downstream regions) hyper di- and trimethylation of H3 K4, and hypo- or basal levels of H4 K16 acetylation (Figure 7.5). In addition, hyper trimethylation of H3 K79 was also present at the immediate downstream region of p^{1b}. Subtle enrichments of H3 K9 acetylation were also apparent (the reasons for this are discussed in section 7.3).

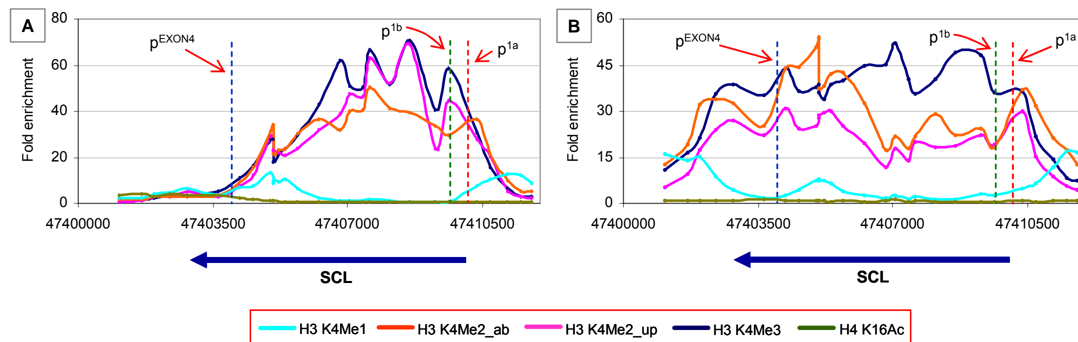

Figure 7.5: Composite ChIP-chip profiles across the 5' end of the SCL gene. Panel A: K562 and panel B: Jurkat. The histone modification shown in the above profiles include mono-, di- and trimethylation of H3 K4 and H4 K16 acetylation (colour coded as shown in the red box at the bottom of the figure). SCL p^{1a}, p^{1b} and p^{EXON4} are marked by red, green and blue dotted lines respectively. The thick blue arrows at the bottom of both panels represent the SCL gene and the direction of transcription. The y-axes represent fold enrichments and the x-axes represent genomic coordinates along the human chromosome 1.

In SCL non-expressing cell lines, the SCL p^{1a}, p^{1b}, and p^{EXON4} displayed hyper di- and trimethylation of H3 K4 and hypo-acetylation of H4 K16 in HL60 – these marks are linked to active regions (Figure 7.4). However, two features associated with inactive promoters were also present - hyper monomethylation of H3 K4 and hypo-acetylation of H3 K9. As shown in chapter 6 (section 6.5), despite the presence of all the methylation marks of H3 K4 at the SCL promoters in HL60, tri- to mono- ratios obtained were very low and reflect the inactivity of the SCL promoters (see chapter 6, section 6.5). Furthermore, the H3 K4 methylated regions in HL60 did not extend beyond the +1 region, whereas in Jurkat and K562 they extended across this region and included the +3 region - suggesting that the +3 region may be important in SCL activation. In HPB-ALL, the inactivity of the promoter was evident by the apparent lack of active histone marks and the presence of hyper-acetylation of H4 K16 – the histone mark associated with inactive regulatory elements.

7.7 Regulatory elements upstream of the SCL promoter 1a

The known SCL regulatory regions located upstream of the 5' end of the gene include the endothelial enhancer at -3/-4 (Gottgens et al. 2004), the -7 and -9/-10 regions (Leroy-Viard et al. 1994; Gottgens et al. 1997) were further analysed and compared for the presence and/or absence of histone marks in the four cell lines (Figure 7.6). A novel region identified in the present study at -12 (see chapter 4, section 4.5.2) was also characterized further.

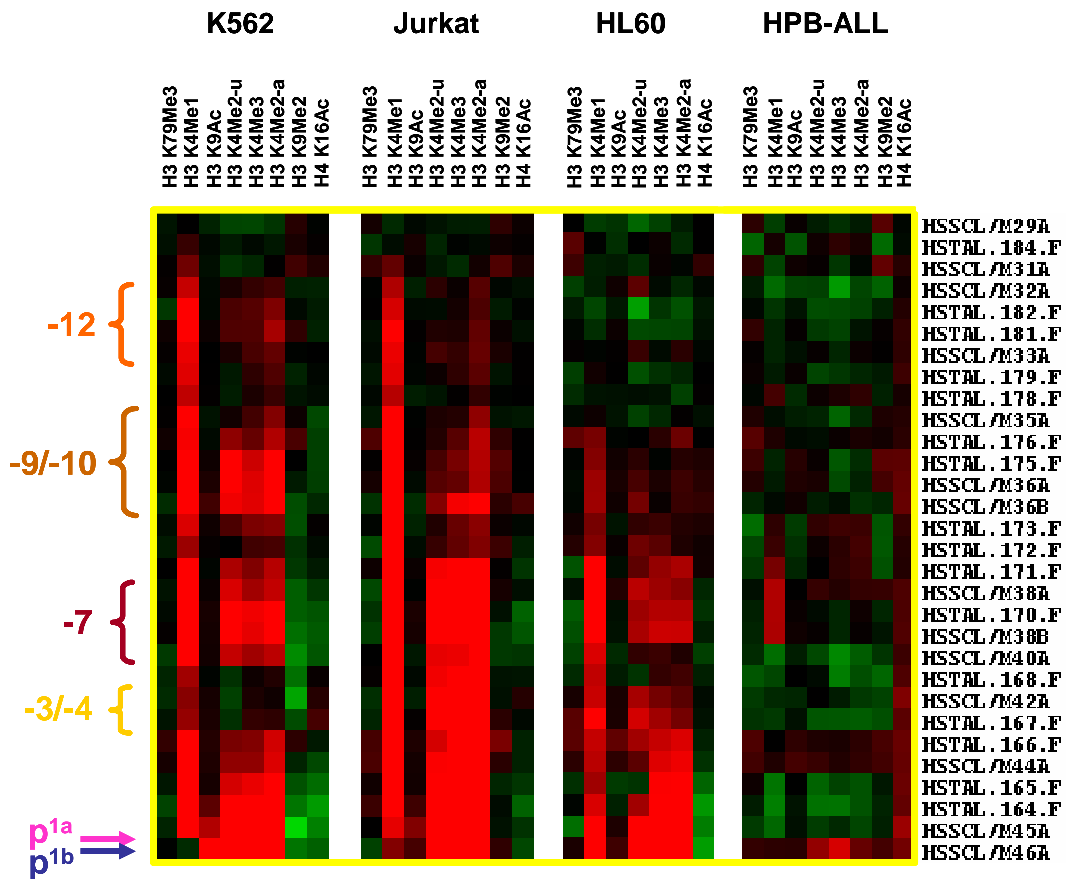

Figure 7.6: TreeView profiles across the genomic regions upstream of the SCL promoter in the four cell lines. The cell line corresponding to each profile is shown at the top. The SCL promoters, p^{1a}, p^{1b} are marked with small coloured horizontal (blue and pink) arrows. The SCL regulatory regions located upstream of the SCL promoter are marked, including the -3/-4 endothelial enhancer, and the -7, -9/-10 and -12 regions. The characterization of the above genomic regions with respect to the histone marks, function and activity is described in the text. The profiles were generated using only the key seven histone modifications which are listed along the x-axis at the top of the figure. Two datasets were included for H3 K4 dimethylation in each cell line. The names of the array elements representing the genomic regions are listed along the y-axis (extreme right). Each box in the profiles represents an array element for the corresponding histone modification and the red/black/green colour indicates enrichment/baseline levels/depletion for that modification respectively.

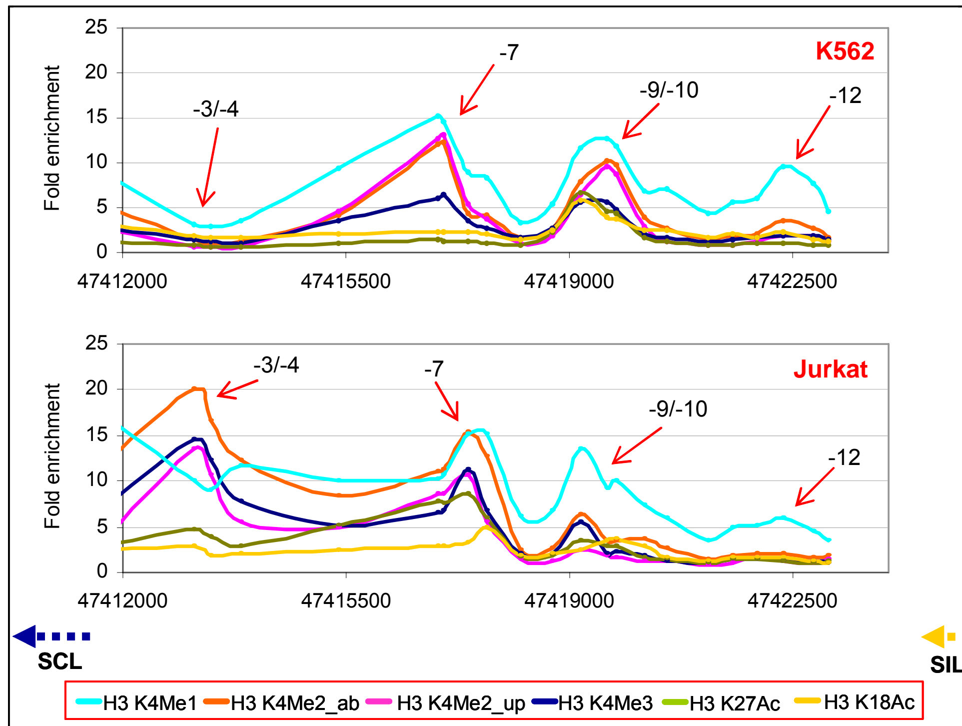

Figure 7.7: Composite ChIP-chip profiles across the SCL regulatory region located upstream of the SCL promoter. Top panel: K562 and bottom panel: Jurkat. The histone modifications shown in the above profiles include H3 acetylation at K18 and K27, mono-, di- and trimethylation of H3 K4 (colour coded as shown in the red box at the bottom of the figure). SCL regulatory regions located upstream of the SCL promoter at -12, -9/-10, -7, and -3/-4 are marked with red arrows. The thick, dotted arrows (yellow: SIL and blue: SCL) at the bottom of the figure indicate the genes on either side of the above shown genomic region and the direction of transcription of these genes. The y-axes represent fold enrichments and the x-axes represent genomic coordinates along the human chromosome 1.

7.7.1 The -9/-10 region

The -9/-10 (-8/-9 in mouse) has been shown to have enhancer activity (Gottgens at el. 1997), DNase I hypersensitivity (Leroy-Viard et al. 1994) and GATA-1 binding (Chapter 4, section 4.6.2) in K562. This region exhibited hyper mono-, di- and trimethylation of H3 K4 and hypo-acetylation of H4 K16 (Figure 7.6) in K562, further suggesting that this region may be active. This was also supported by the presence of histone H3 acetylation marks (Figure 7.7, top panel). In Jurkat, hyper mono-, di- and trimethylation of H3 K4 were present but the relative enrichments for di- and trimethylation H3 K4 and histone H3 acetylation marks were low as compared to K562 (Figure 7.6; Figure 7.7, bottom panel); increased levels of H3 K9 dimethylation were also observed and depletion of H4 K16 acetylation was seen on only one of the array elements at the -9/-10 region. These data suggest that this region may be less active or inactive in Jurkat. The absence of DNase I hypersensitivity at this region in Jurkat (Leroy-Viard et al. 1994) further supports these interpretations.

In HL60 and HPB-ALL, the -9/-10 region showed histone marks associated with inactive regions, such as, hyper-acetylation of H4 K16, hypo di- and trimethylation of H3 K4. These results are consistent with this region being inactive in SCL non-expressing cell lines.

7.7.2 The -7 region: evidence for an active enhancer in K562

The -7 region had previously been shown to exhibit DNAse I hypersensitivity in K562 (Leroy-Viard et al. 1994) and binding of GATA-1 in K562 (Chapter 4, section 4.6.2). However, its function in gene regulation is not known. The histone marks present at this region included hyper mono-, di- and trimethylation of H3 K4, and hypo acetylation of H4 K16 (Figure 7.6) which are all linked to active enhancer regions. Furthermore, the -7 region clustered along with the other known SCL enhancers in the functional clustering analysis (see Figure 6.5, chapter 6).

This region was further tested in transient luciferase reporter assays in K562 along with the -12 region (see section 7.7.4) and the +51 erythroid enhancer which acted as a positive control (Figure 7.8). The transient assays report the expression of the firefly luciferase gene under the control of the SV40 promoter. In the presence of an enhancer sequence cloned into the SV40 promoter-luciferase construct, increased levels of luciferase expression are observed. Both the -7 and +51 regions showed increased luciferase expression, suggesting that the -7 region had enhancer activity in K562. This provides evidence that the -7 region is a novel enhancer which is likely to regulate SCL expression in K562.

This novel enhancer displayed the same histone marks in Jurkat as was seen in K562. Noticeable nucleosome depletion was also seen at the -7 region in Jurkat (Chapter 5, section 5.11) as well as extreme hypersensitivity to DNase I (Leroy-Viard et al. 1994). It was further noticed that although significant enrichments were reported at the -7 region for H3 acetylation in both K562 and Jurkat, the relative enrichments were much higher in Jurkat as compared to K562 (see Figure 7.7). All of this, taken together, suggests that the -7 enhancer is active in Jurkat. In addition, it has been previously reported that the -7 region was found near one of the SCL breakpoints reported in SIL-SCL deletions (Breit et al 1993). This further alludes to the -7 enhancer being involved in the *cis*-activation of SCL in Jurkat.

This region displayed the presence of similar histone marks in HL60 as seen in K562, however, the relative enrichment levels for di- and trimethylation of H3 K4 were found to be very low in HL60 as compared to K562. Hyper-acetylation of H4 K16 and hypo di- and

trimethylation of H3 K4 was seen at the -7 region in HPB-ALL. These results are consistent with this region being inactive in SCL non-expressing cell lines

7.7.3 The endothelial enhancer at the -3/-4 region

The -3/-4 region, also known as the endothelial enhancer, directed expression to haematopoietic progenitors and endothelium and is expressed in blood progenitors and endothelial cells. This enhancer was found to be regulated by the transcription factors Elf-1 and Fli-1 (Gottgens et al. 2004).

In K562, this region exhibited hypo di- and trimethylation of H3 K4 and hyper-acetylation of H4 K16, all of which are histone marks for inactive regulatory region (Figure 7.6). Hypo- K18 and K27 acetylation was also a hallmark of this region in K562 (Figure 7.7). Conversely, in Jurkat, substantial hyper mono-, di- and trimethylation of H3 K4 were present at this region, and H4 K16 acetylation levels were lower than in K562. Significant enrichments for acetylation of histone H3 at K18 and K27 were also seen at this region in Jurkat (Figure 7.7). In a previous study, a DNase I hypersensitive site was identified at -3/-4 region in Jurkat only and not in K562 (Leroy-Viard et al. 1994). These data suggest that the -3/-4 region may not be active in K562 but is possibly active in Jurkat and may also represent a region which could be involved in *cis*-activation of SCL in Jurkat.

In HL60, significant but relatively low level enrichments for active histone marks such as di- and tri-methylation of H3 K4 were observed, but this was not accompanied by hypo-acetylation of H4 K16. In HPB-ALL hyper-acetylation of H4 K16 was observed. These results are consistent with this region being inactive in SCL non-expressing cell lines.

7.7.4 The -12 region: a putative enhancer of SCL regulation

The -12 region was identified through its association with significant levels of H3 K9/14 acetylation in K562 (Chapter 4, section 4.5.2). In K562, it also showed significant enrichments for H3 K18 acetylation (Figure 7.7), along with mono-, di- and trimethylation of H3 K4 and hypo-acetylation of H4 K16. This suggests that this region could represent a putative regulatory element which may be active in K562. In Jurkat, although hyper monomethylation was present, the enrichments for di- and trimethylation of H3 K4 were very low (Figures 7.6 and 7.7) suggesting that this region was inactive. In HL60 and HPB-ALL, the presence of hypo mono-, di- and trimethylation of H3 K4 and hyper-acetylation of H4 K16 were consistent with this region being inactive in SCL non-expressing cell lines (Figure 7.6).

To elucidate the function of the -12 region as a regulatory element, genomic sequences from this region were tested in transient luciferase reporter assays in the K562 cell line

(Figure 7.8). Two different constructs across the -12 region (denoted -12b and -12s in Figure 7.8) were used in these assays along with the -7 region (discussed in section 7.7.2) and the +51 erythroid enhancer which acted as a positive control. The -12 constructs decreased expression of the reporter suggesting that this sequence acted as a silencer/repressor of transcriptional activity (see Appendix 13 for the sequences of the primer pairs used to generate the constructs).

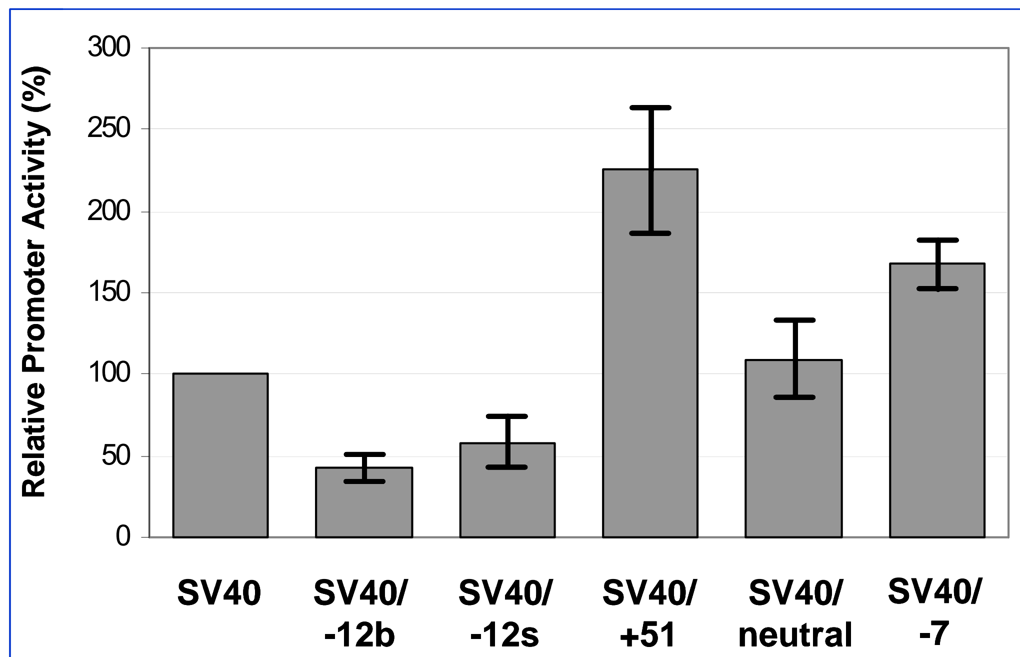

Figure 7.8: Transient luciferase reporter assays in K562 cells. The histogram shows luciferase activities for five constructs containing genomic DNA inserts from the human SCL locus. The y-axis is the luciferase activity expressed as a percentage of the SV40 promoter activity. Standard deviation bars are shown for each construct apart from SV40, to which all data was normalised. Labels on the x-axis are as follows: SV40 (promoter only), SV40/-12b (2.2 kb from the human -12 region), SV40/-12s (453 bp from the human -12 region), SV40/+51 (936 bp from the SCL erythroid enhancer at +51), SV40/neutral (951 bp from the CYP4Z1 gene used as a human DNA negative control), SV40/ -7 (952 bp from the human SCL -7 region).

The -12 region showed stretches of high non-coding sequence conservation in a number of mammalian species which suggests that this putative silencer/repressor may have an evolutionarily conserved function. In order to further elucidate the regulatory interactions which could occur at the -12 region, transcription factor binding sites which were conserved for the orthologous sequences in human, chimp, mouse, rat and dog were identified including an 11 bp sequence for the ETV6/7 (TEL1/2) proteins (Figure 7.9).

Figure 7.9: Conserved transcription factor binding sites found at the SCL -12 region. Sequence alignments are shown for human, chimp, mouse, rat and dog. Genomic sequence co-ordinates for each region of homology are shown in brackets at the beginning and the end of each sequence. Bases of sequence identity are denoted with an asterisk (*). The transcription factor binding site for ETV6/7 (TEL1/2) is boxed in red. Sites are shown for a variety of other transcription factors including Sp1, PEA3, ETS-1, GR (glucocorticoid receptors), RAR-x (retinoic acid receptors), AP-1,2,4 (activator proteins), and NFAT-x (nuclear factors of activated T cells), all boxed in black.

ETV6 (TEL1) and ETV7 (TEL2) are members of the Ets family of transcription factors, are expressed during haematopoiesis and have been shown to exhibit strong repressor activity *in vivo* (Chakrabarti and Nucifora 1999; Lopez et al. 1999; Poirel et al. 2000; Gu et al. 2001). Therefore, these proteins represent good candidates which may mediate the repressor activity of the -12 element.

7.8 Regulatory regions downstream of the SCL gene

7.8.1 The stem cell enhancer at the +20/+21 region

It has been shown that the SCL stem cell enhancer at +20/+21 (+18/+19 in mouse) is regulated by GATA-2, Elf-1, Fli-1 and targets a vast majority of haematopoietic progenitors (Gottgens et al. 2002). Thus, this enhancer is active at an early stage of blood development. However, SCL expression is not dependent on the activity of this enhancer as initiation of SCL transcription or for formation of haematopoietic cells can occur without the +20/+21 region being functional (Gottgens et al. 2004).

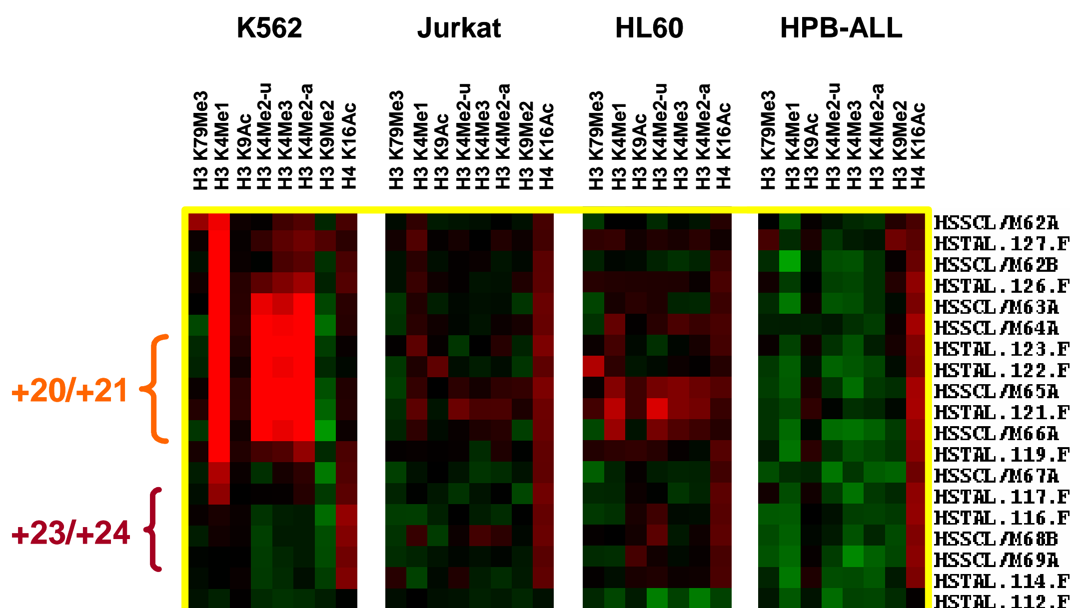

Figure 7.10: TreeView profiles across the stem cell enhancer and the neural enhancer in the four human cell lines. The cell line corresponding to each profile is shown at the top. The genomic regions encompassing the stem cell enhancer at +20/+21 (homologous to +18/+19 in mouse) and the neural enhancer at +23/+24 (homologous to +23/+24 in mouse) are marked on the left. The characterization of the above genomic regions with respect to the histone marks, function and activity is described in the text. The profiles were generated using only the key seven histone modifications which are listed along the x-axis at the top of the figure. Two datasets were included for H3 K4 dimethylation in each cell line. The names of the array elements representing the genomic regions are listed along the y-axis (extreme right). Each box in the profiles represents an array element for the corresponding histone modification and the red/black/green colour indicates enrichment/baseline levels/depletion for that modification respectively.

In K562, the +20/+21 region displayed hyper mono-, di- and trimethylation of H3 K4 which are all active histone marks (Figure 7.10). The presence of hypo/hyper acetylation of H4 K16 at this region is not pronounced and levels are near background levels. However, in comparison with H4 K16 levels at other SCL regulatory regions, the levels at the +20/+21 region are markedly higher (see chapter 5, Figure 5.9), suggesting a reduced activity for this region in K562. However, no binding interactions for Elf-1 and Fli-1 were detected at this region in K562, suggesting that although the activating histone marks may be in place, the region is not active (or other transcription factors may be recruited in K562). Further characterization of this region would be required to deduce its activity and function in the K562 cell line.

In Jurkat, HL60 and HPB-ALL, the interpretation was far easier for the +20/+21 region. It exhibited histone marks that are linked with inactive regions such as hypo mono-, di- and

trimethylation of H3 K4 and hyper-acetylation of H4 K16 suggesting that the +20/+21 region was inactive in these cell lines.

7.8.2 The neural enhancer at the +23/+24 region

The +23/+24 region was originally identified using comparative sequence analysis with the syntenic SCL regions in human, mouse and chicken. Transgenic reporter assays established that this enhancer directed reporter expression to hindbrain and spinal cord (Gottgens et al. 2000). In K562, Jurkat, HL60 and HPB-ALL, the +23/+24 region exhibited hypo mono-, di- and trimethylation of H3 K4 and hyper-acetylation of H4 K16, all of which suggest that this region is inactive in these cell lines (Figure 7.10). This is consistent with the known biology of this enhancer having activity in neural development but not during haematopoiesis.

7.8.3 The +39 region

The +39 region was identified in this study by virtue of its association with various histone modification assays in four human cell lines (See chapter 5). This region, exhibited all the features of an enhancer element based on the hierarchical clustering analysis presented in chapter 6 (section 6.8; 6.9).

This region displayed hyper mono-, di- and trimethylation of H3 K4, and hypo-acetylation of H4 K16 in all the cell lines (see Figure 7.11, top panel), although the relative levels of methylation were quite low in HPB-ALL as compared to other cell lines. In addition, decreased levels of nucleosomes were also observed at this region (Figure 7.11, bottom panel) in all four human cell lines. This data, taken together suggests that +39 is active in all of the cell lines analysed here. In K562, binding interactions for the transcription factor CTCF were detected at the +39 region implicating it in enhancer-blocking activities of insulators, or possibly suggesting that +39 was an insulator itself (Chapter 5, section 5.17.6.2). As the +39 region lies close to and upstream of MAP17 gene (3 kb upstream), it is not known whether this region is associated with regulation of MAP17 or whether it represents a 3’ distant element associated with expression of SCL. Therefore, the function of +39 needs further characterization.

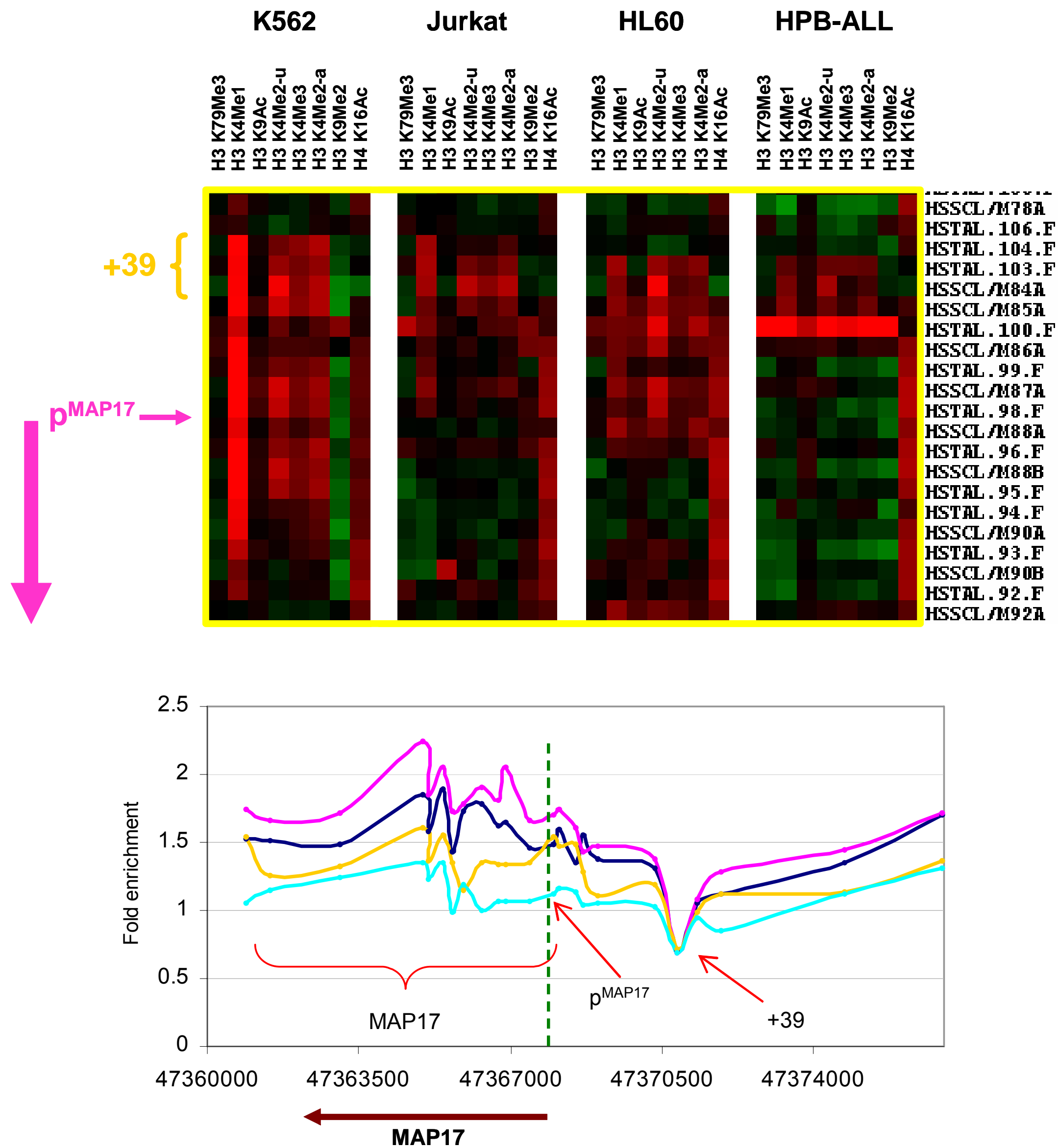

Figure 7.11: TreeView profiles and composite ChIP-chip profiles across the +39 region and MAP17 gene in four human cell lines. The top panel shows the TreeView profiles. The cell line corresponding to each profile is shown at the top. The +39 region is marked on the left. The MAP17 promoter, p^{MAP17} is shown with the thinner pink coloured horizontal arrow and the thick vertical arrow represents the direction of transcription of the MAP17 gene. The characterization of the above genomic regions with respect to the histone marks, function and activity is described in the text. The profiles were generated using only the key seven histone modifications which are listed along the x-axis at the top of the figure. Two datasets were included for H3 K4 dimethylation in each cell line. The names of the array elements representing the genomic regions are listed along the y-axis (extreme right). Each box in the profiles represents an array element for the corresponding histone modification and the red/black/green colour indicates enrichment/baseline levels/depletion for that modification respectively. The bottom panel shows the composite ChIP-chip profiles for histone H3 in the four human cell lines. The regions are marked by red arrows and the genomic position of the p^{MAP17} is shown by the dotted green line. K562: turquoise, Jurkat: blue, HL60: pink and HPB-ALL: yellow profiles. The thick arrow at the bottom represents the MAP17 gene and the direction of transcription. The y-axis in the bottom panel represents fold enrichments and the x-axis represents genomic coordinates along human chromosome 1.

7.8.4 The MAP 17 promoter

In a previously reported study, it was demonstrated that the cell lines which express SCL also co-express MAP17, albeit at low levels (Delabesse et al. 2005). Thus, it was proposed that the co-expression of these genes, was due to their co-regulation by shared regulatory sequences (Delabesse et al. 2005). Thus, the MAP 17 promoter would be expected to represent an active region in K562 and Jurkat, and would be inactive in HL60 and HPB-ALL.

In K562, the promoter region of MAP17 exhibited hyper monomethylation of H3 K4, hyper-acetylation of H4 K16, hypo trimethylation of H3 K4 and hypo-acetylation of H3 K9 - all these histone marks are associated with inactive promoter regions. This suggested that the promoter of MAP17 gene is inactive in K562 (Figure 7.11, top panel). Similarly, in Jurkat, the presence of hypo-acetylation of H3 K9, hypo trimethylation of H3 K4 and hyper dimethylation of H3 K9 suggested that, the MAP17 promoter was also inactive in Jurkat. Additionally, hyper trimethylation of H3 K79 downstream of active promoters was not seen at the region downstream of the MAP17 promoter in K562 or Jurkat.

Additional sources of evidence also support the idea that MAP17 is not active in the SCL-expressing cell lines studied here:

i) It was shown in chapter 6, that the tri- to monomethylation of H3 K4 were very low for the MAP17 promoter as compared to all the other active promoters across the SCL locus in K562 and Jurkat (section 6.5).

ii) The MAP17 promoter region did not display relatively low levels of nucleosomes (Figure 7.11, bottom panel).

iii) In ChIP-chip assay for $Taf_{II}250$ (which binds to active promoters), no significant enrichments were reported at the MAP17 promoter (Chapter 5, section 5.14).

iv) Binding interactions for CTCF transcription factor, which is known for its function as an enhancer-blocker (Bell et al. 1999), were also detected at the MAP17 promoter in the K562 cell line (Chapter 5, section 5.16). This interaction suggests that the promoter of MAP17 is possibly being blocked to modulate its expression (Chapter 5, section 5.17.6.2).

v) Histone modification profiles in HL60 and HPB-ALL, looked similar to those found in K562 and particularly to those in Jurkat; histone modification profiles linked with inactive regions were present at the MAP17 promoter which was in complete agreement with the known inactivity of this promoter in SCL non-expressing cell lines (Figure 7.11).

7.8.5 The erythroid enhancer at the +51/+53 region: evidence for sites of transcription downstream of the +51 region.

The +51 region (homologous to +40 in mouse) was identified as an erythroid enhancer which directed reporter expression to primitive erythroblasts (Delabesse et al. 2005). In the present study, binding interactions for GATA-1, SCL and Ldb-1, p300, CBP and HDAC2 were detected at the +51 region in K562 (see chapters 4 and 5). The highest ChIP enrichments in all the above mentioned ChIP assays were observed at the same array element – named HSSCL/M96B in Figure 7.12.

In K562, the +51 erythroid enhancer (centered around the HSSCL/96B array element) exhibited hyper mono-, di- and trimethylation of H3 K4 and hypo-acetylation of H4 K16 - the presence of these marks is in complete agreement with the known activity of this enhancer in K562 (Figure 7.12). In Jurkat and HL60, lower levels of mono-, di- and trimethylation of H3 K4 and hyper-acetylation of H4 K16 histone marks suggested this enhancer may not be active in these cell lines. Similarly, the presence of hyper-acetylation of H4 K16 in HPB-ALL suggested that this enhancer may not be active in this cell line.

However, there was a noticeable pattern of histone modifications across the genomic region spanning from +50 to +53 in K562 - covering the +51 enhancer and extending in either direction from it. This whole 4 kb region exhibited significant enrichments for histone modifications including acetylation of histone H3 at K9/14, K18, K27 and mono-, di- trimethylation of H3K4 (see Figures 7.12 and 7.13). The relative levels of enrichments for H3 K18 acetylation were higher upstream of +51 (towards the MAP17 gene), whereas enrichments for H3 K27 acetylation were higher at the downstream region of the +51 enhancer at the +53 region (Figure 7.13). This suggests that a different set of modifications marks are present on either side of the +51 enhancer. This was further supported by the presence of a peak of enrichment for H3 K4 di- and trimethylation at +53, and a smaller peak upstream of +51 (Figure 7.13).

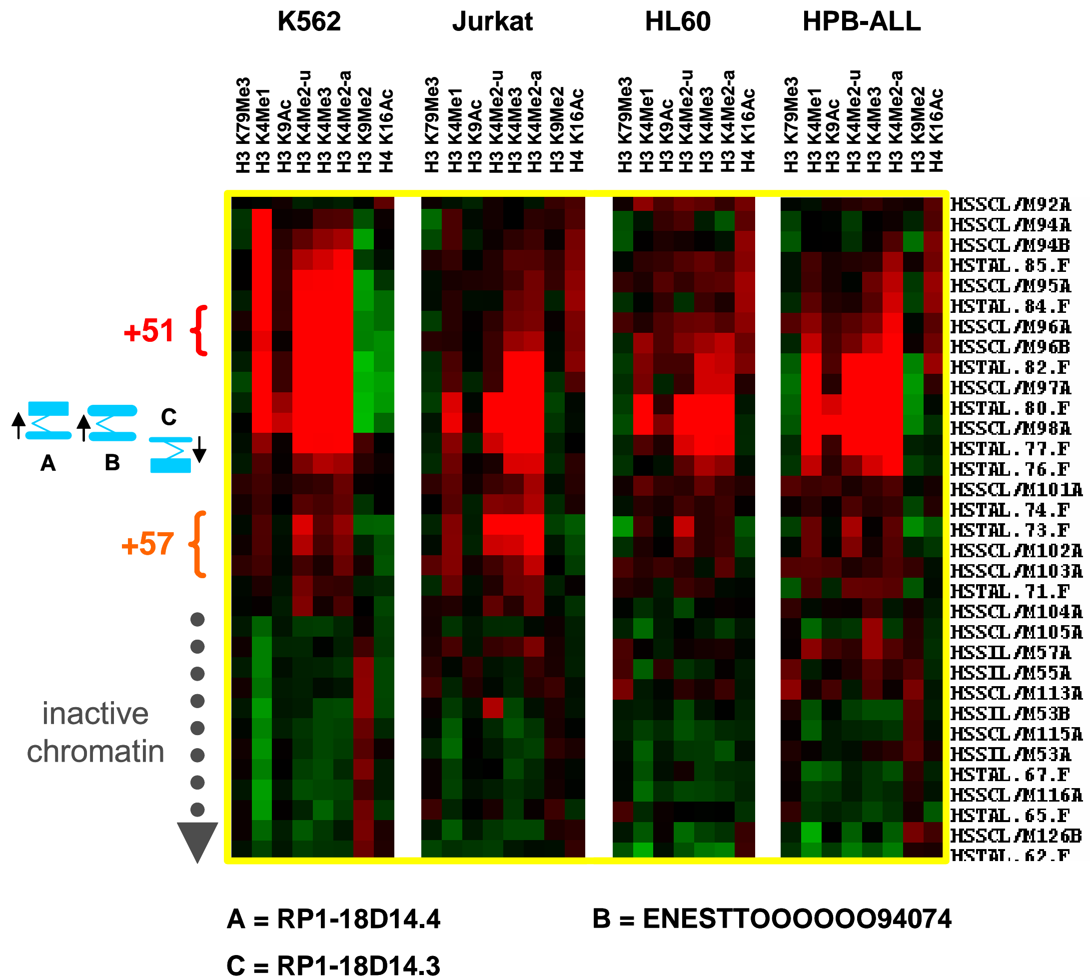

Figure 7.12: TreeView profiles across the erythroid enhancer and the +57 region in four human cell lines. The cell line corresponding to each profile is shown at the top. The genomic regions encompassing the erythroid enhancer at +51 region and the +57 regions are marked on the left. The characterization of the above genomic regions with respect to the histone marks, function and activity is described in the text. The three blue small transcripts (shown on the left) are annotated in the ENSEMBL database and named A, B and C (shown with their ENSEMBL IDs at the bottom of the figure). The small black arrows next to the blue transcripts represent the direction of transcription for these transcripts according to ENSEMBL. The dotted grey vertical arrow shown at the downstream region of +57 represents the beginning of the silent chromatin encompassing the CYP4A22 and CYP4Z1 genes. The above profiles were generated using only the key seven histone modifications which are listed along the x-axis at the top of the figure. Two datasets were included for H3 K4 dimethylation in each cell line. The names of the array elements representing the genomic regions are listed along the y-axis (extreme right). Each box in the profiles represents an array element for the corresponding histone modification and the red/black/green colour indicates enrichment/baseline levels/depletion for that modification respectively.

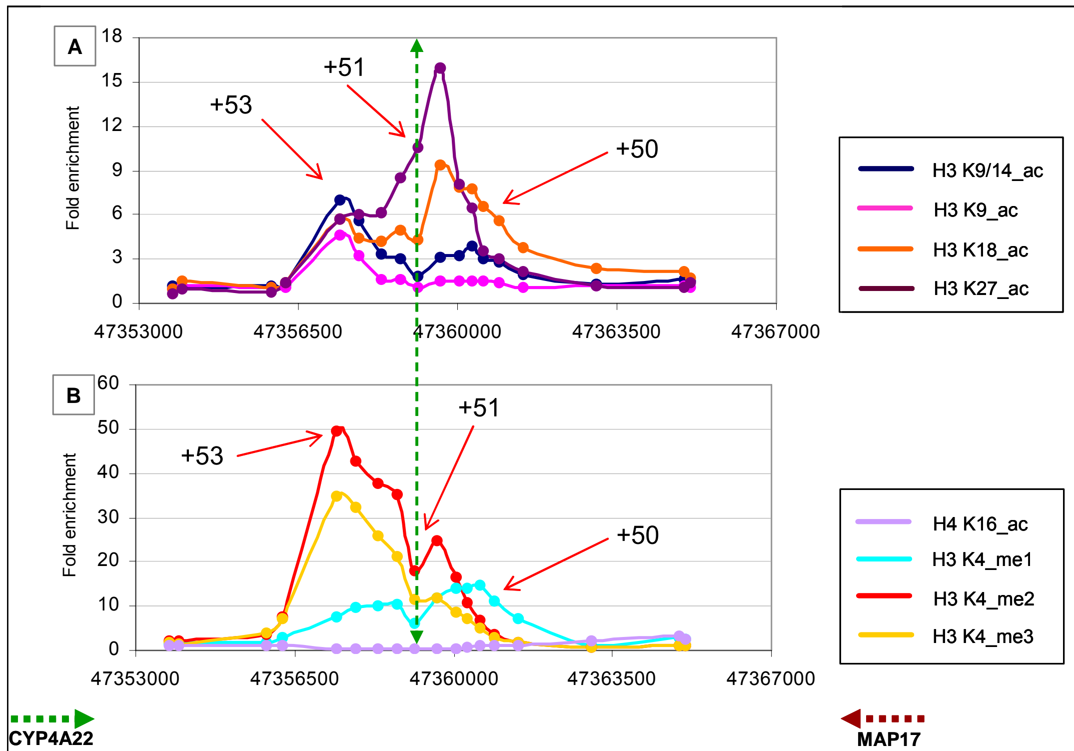

Figure 7.13: Composite ChIP-chip profiles for histone modifications at the +51/+53 regions in K562. The histone modification used in generating the above profiles included H3 K9/14 diacetylation, acetylation of histone H3 at K9, K18 and K27 (panel A) and H4 K16 acetylation, mono-, di- and trimethylation of H3 K4 (panel B). The colour codes of the profiles are colour coded as shown in the boxes on the right. The location of the array element (HSCL/96B) which showed transcription factor binding is represented by the green dotted arrow in panels A and B of the figure. The +50, +51 and +53 regions are marked with red arrows. The dotted thick arrows at the bottom of the figure represent the genes (and the direction of transcription) located on either side of the +50 to +53 region. The y-axes represent fold enrichments and the x-axes represent the genomic coordinates along human chromosome 1.

Another striking observation in the TreeView profiles shown in Figure 7.12 was that the cell lines Jurkat, HL60 and HBP-ALL only showed marked histone modifications on the downstream (+53 side) (Figure 7.12) part of this region distal to the location of array element HSCL/96B. This suggested that the region distal to +51 may have a regulatory function which is distinct from that of the erythroid enhancer, and which was active in all of the cell lines. In the hierarchical clustering analysis performed for histone modifications in K562, the +51/+53 region clustered together with the SCL p^{1a}, whereas the +50/+51 region clustered with the known SCL enhancers (Chapter 6, sections 6.8 and 6.9). These data suggest that the +50 to +53 region may exhibit features of both an enhancer (i.e., the known erythroid enhancer) and an active promoter, delineated on either side of the +51 region. Evidence to further support the presence of a promoter near +53 is as follows:

i) The +53 region showed hyper-acetylation of H3 K9, hyper di- and trimethylation of H3 K4, hypo dimethylation of H3 K9 and hypo-acetylation of H4 K16 in all four cell lines - all these are hallmarks of active promoters (Figure 7.12 and Figure 7.13).

ii) The ratios for tri- to monomethylation for H3 K4 across the +53 region were high (in K562, Jurkat and HPB-ALL), which is indicative of the transcriptional activity of a promoter region (Chapter 6, section 6.5).

iii) Three transcripts annotated in the ENSEMBL database (www.ensembl.org) were found to map to the +53 region. The three transcripts - A, B and C, are shown in Figure 7.12. All three transcripts are only two exons long; two of them are transcribed in one orientation and the third one in the opposite direction. All three, have their first exon beginning within the +53 genomic sequence.

iv) The ENSEMBL database shows a FirstEF promoter prediction (Davuluri et al. 2001) at the +52 region (data not shown).

Taken together, the above presented results suggest that the genomic region encompassing the +51 erythroid enhancer region seems to be associated with two distinct functions – that of an enhancer (+50/+51 region) and of a putative promoter (+53 region). More importantly, the histone code for this region predicts its enhancer and promoter activity, suggesting that the consensus histone code established in chapter 6 is not only consistent for the known regulatory regions across the SCL locus, but it could also be used to identify novel regulatory regions.

7.8.6 The putative insulator element at the +57 region

In K562 and Jurkat, the +57 region showed binding of CTCF in ChIP-chip, a protein known to bind to mammalian insulators (Bell et al. 1999). In K562 and Jurkat, the histone marks displayed by this region included hyper di- and trimethylation of H3 K4, hypo-acetylation of H4 K16 (Figure 7.12). The relative enrichments for di- and trimethylation of H3 K4 in Jurkat were higher as compared to K562 (Figure 7.12). The presence of these histone marks suggests that this region is possibly active in both the cell lines, and this is supported by CTCF binding. In Jurkat, significant enrichments were also seen for H3 K9/14 diacetylation (Chapter 5, section 5.6.1.1). In HL60 and HPB-ALL, the enrichments for di- and trimethylation of H3 K4 histone marks were not significant, although hypo-acetylation of H4 K16 was present. CTCF ChIP-chip was not performed in HL60 and HPB-ALL, therefore the activity of this region in those cell lines is not certain.

7.9 Discussion

The work described in this chapter involved a detailed characterization of a number of regulatory regions, known and novel, across the SCL locus in SCL expressing and non-expressing cell lines. These regions were characterized with respect to the proposed consensus histone code to test the consistency of the code in predicting the function and activity of these regulatory regions in SCL expressing and non-expressing cell lines. Based on the various ChIP-chip analyses performed in the previous chapters of this thesis, regulatory elements were also characterized for their function, role in transcriptional regulation of SCL and their associated DNA-protein interactions.

7.9.1 Identification and characterization of putative novel regulatory elements

The various ChIP-chip analyses performed for this study in SCL expressing and non-expressing cell lines in human and mouse identified a number of DNA-protein interactions across the SCL locus – the majority of which were known previously but some of which were novel. The identification of putative novel regulatory elements at the SCL locus further increases the complexity of regulatory elements which may control expression of SCL or its flanking genes. Table 7.1 summarizes all the known regulatory regions and the novel regions identified in K562 in the present study (data for the other human and mouse cell lines are in Tables 5.2, 5.3 and 5.4 in chapter 5). The binding interactions of transcription factors or other regulatory proteins, either known or identified in the present study, at various regulatory regions across the SCL locus are also listed in the table.

Known regulatory regions	Known *trans*-factor interactions	Novel regions	Novel *trans*-factor interactions	Potential function
KCY promoter			$Taf_{II}250$	
		KCY -4/-5	p300, CBP, HDAC2, $Taf_{II}250$	
SIL promoter, p^{SILa}			p300, CBP, HDAC2, $Taf_{II}250$	
		SIL promoter, p^{SILb}	CBP, HDAC2, $Taf_{II}250$	alternative SIL promoter
		SIL -1	Elf-1	
		SIL -10, +10, +34		
SCL p^{1a}, p^{1b}, +1, +3	GATA-1*		GATA-1*, p300, CBP, HDAC2, $Taf_{II}250$	
SCL p^{EXON4} at +7			GATA-1 at +7/+8 (downstream of p^{EXON4})	
SCL endothelial enhancer at -3/-4	Elf-1, Fli-1			
		SCL -23/-26, -12		SCL -12, repressor element
SCL DNase I HSs at -7, -9/-10			GATA-1, SCL -7	SCL -7, enhancer element
SCL stem cell enhancer at +20/+21	GATA-2, Elf-1, Fli-1			
SCL neural enhancer at +23/+24				
		SCL +39	CTCF	putative enhancer
MAP17 promoter			CTCF	
SCL +51 erythroid enhancer			GATA-1, SCL, Ldb-1, p300, CBP, HDAC2, $Taf_{II}250$ (not the same array element as the other TFs)	
		SCL +53		putative promoter
		SCL +57	CTCF	putative insulator element

Table 7.1: A summary of the known regulatory regions and novel putative regulatory regions at the SCL locus in K562. The list of novel regions and *trans*-factor binding interactions at the SCL locus, shown in

the 3rd and 4th column, has been compiled mostly based on the data generated in the present study in the human K562 cell line. The 5th column shows the proposed functions of a number of the regulatory region at the SCL locus. The data shown in the 1st and 2nd column has been compiled from previously published studies across the SCL loci in human and mouse. Please note: this list is by no means an exhaustive list of all the novel regions identified across the SCL locus. For other novel regions that were identified in the other human and mouse cell lines, see Tables 5.2, 5.3 and 5.4 and the text in chapter 5. The Elf-1 interaction, coloured pink, was identified in K562, Jurkat and HPB-ALL. The CTCF interaction, coloured yellow, was identified in the K562 and Jurkat cell lines. The GATA-1 interaction, marked with a red asterisk (*), was already known from previous studies and was confirmed in the present study (detailed discussions with respect to each regulatory region have been provided in the relevant sections in this chapter).

Notably, a region was identified upstream of the SIL gene at SIL -10, which was enriched for various histone modifications and was detected most notably in only SCL expressing cell lines. Whether this region represents a novel regulatory element which could be involved in the regulation of SCL needs to be further investigated. Elf-1 binding at the SIL -1 region in K562, Jurkat and HPB-ALL may suggest a role for this region in the regulation of the SIL and/or SCL gene which also requires further studies (also see discussion in chapter 5). Similarly, a number of other regions within the SIL gene, for example SIL +10, SIL +34 and SCL -23/-26 region, or the other regions identified in the mouse cell lines (see Table 5.4 in chapter 5) may represent additional regulatory elements associated with the regulation of SCL gene. These regions need further characterization to understand their function and role at the SCL locus.

Based on the analyses of ChIP-chip results for histone modifications and CTCF obtained at the +57 region in the SCL expressing cell lines K562 and Jurkat, this region may represent a putative insulator element. The role of the +57 region in insulator activity is also discussed in chapter 5; section 5.17.6.2). Whether it performs the dual-roles proposed for insulators (West et al. 2002), which include enhancer blocking activity and acting as a barrier to separate silent chromatin from active chromatin, is not known. CTCF is known for its enhancer-blocking functions at the HS4 insulator element at the chicken β-globin locus (Bell et al. 1999), where a separate protein, USF, is responsible for the barrier function (West et al. 2004). For the chicken β-globin locus, it was shown that the histone modification patterns at the HS4 insulator element (with the barrier activity) and the 3’ HS element (with only enhancer-blocking activity) were different (Litt et al. 2001a; Litt et al. 2001b). The HS4 element was found to be hyper-acetylated for H3 K9/14 and hyper-methylated for H3 K4, whereas, these modifications were not seen at the 3’ HS element. The data obtained in the present study showed higher enrichments for di- and trimethylation of H3 K4 at the +57 region in K562 and Jurkat. In Jurkat, significant enrichments were also seen for H3 K9/14 diacetylation. This may suggest that

the +57 region may indeed function as a barrier element (as well as enhancer blocking functions discussed in chapter 5, section 5.17.6.2).

In addition to the +57 region, two putative regulatory regions were characterised further in the present study. At the SCL gene, the -7 region identified in K562 showed significant levels of acetylation and GATA-1 binding and showed enhancer activity in reporter assays. The -12 region, showed H3 acetylation and repressor activity in reporter assays, but no transcription factors were found to bind to this region, although TEL1 or TEL2 make appealing candidates based on sequence conservation of their binding sites and their known role as repressors in haematopoiesis.

Another intriguing region at the SCL locus is the +51 erythroid enhancer and its adjacent regions at +50 and +53. The data obtained from various analyses in the present study suggest that, this region possibly has two separate functions – enhancer function at the +50/+51 region and a potential promoter function at the +52/+53 region. The erythroid enhancer at the +51 region (homologous to the mouse +40 region) which targets expression to primitive erythroblasts is already known (Delabesse et al. 2005). The proposed histone code at the SCL locus suggests that the +53 region represents a putative promoter, and this idea is further supported by (i) three annotated small transcripts mapping to this region, and (ii) a FirstEF promoter prediction at this region. These transcripts, however, have not been shown to have significant open reading frames which suggest they encode protein products. Recently published studies have reported that a huge number of transcripts in the human and mouse genomes – about half of total – do not appear to encode proteins (Kapronov et al. 2002; Bertone et al. 2004; Okazaki et al. 2002). It is being suggested that these non-coding RNAs (or ncRNAs) may provide an added layer of complexity to the regulation of complex genomes. Many of the ncRNAs appear to be developmentally regulated (Cawley et al. 2004). Functional analysis revealed the involvement of ncRNAs in pathways such as Hedgehog signalling and nuclear trafficking of transcription factor, NFAT (Willingham et al. 2005). It has also been suggested that several of the ncRNAs might also be antisense to known genes (Cawley et al. 2004) and antisense transcription plays important functions in eukaryotic cells (Yelin et al. 2003). Many of the ncRNAs are associated with nearby transcription factor binding sites (TFBS) (Cawley et al. 2004).

Based on the above mentioned studies, and the important role played by the ncRNAs in the regulation of protein-coding genes, suggests a potential role for the +53 region at the SCL locus. It is possible that the novel transcripts mapping to the +53 region represent ncRNAs which may have a functional role in the regulation of SCL itself or other genes.

The previously suggested association of ncRNAs with nearby transcription factor binding sites (TFBS) (Cawley et al. 2004) may mean that the transcription factors binding at the +51 erythroid enhancer may be involved in regulating the expression of these transcripts. However, all these hypotheses need to be further studied in detail. Thus, the +51/+53 region makes an interesting candidate region to be characterized further to understand its role at the SCL locus.

7.9.2 Transcriptional activation of SCL in Jurkat

Jurkat (a leukaemic T-cell line) exhibits inappropriate expression of SCL but the mechanism of its activation in this cell line is not known. Given the fact that no genomic rearrangements at the SCL locus have been found and SCL expression in Jurkat is mono-allelic (Leroy-Viard et al. 1994), this suggests that a *cis*-acting mechanism is involved in activating the single allele. The presence of significant acetylation at the -3/-4 endothelial enhancer (Gottgens et al. 2004) and at the -7 region in Jurkat suggests an open chromatin structure for these regions. This means that the increased positive charges due to acetylation weakens the histone-DNA contacts and promotes nucleosome mobility (Hansen 1998) which would make the underlying chromatin susceptible to nuclease digestion. Both these regions encompass known DNase I HSs in Jurkat (Leroy-Viard et al. 1994); the data obtained here was in complete agreement with these previous observations. The presence of histone marks, considered the hallmarks of active promoters, at SCL p^{1b} and p^{EXON4} were also consistent with the known activity of these two promoters in Jurkat (Bernard et al. 1992). Based on the data obtained, the -3/-4 region and the -7 region are likely candidates to be characterized further for their involvement in SCL activation in Jurkat as both these elements exhibit hallmarks of active regulatory regions. It is also possible that a genomic region more distant to the SCL gene (some other, yet, unknown region) is also involved in mediating the transcriptional activity of SCL, a mechanism by which Myc expression is known to be activated (Lazo et al. 1990). Interestingly, the SCL locus in Jurkat did not appear to be significantly enriched for H3 K9 modifications, acetylation as well as dimethylation. Whether this is specific cell-type preference for certain histone marks or represents a particular feature of the SCL locus in Jurkat is not known.

7.9.3 MAP17 expression in K562 and Jurkat

It has been previously reported that MAP17 is expressed, albeit at low levels, in all the cell types that express SCL, suggesting that the two genes share regulatory elements which coordinate their expression (Delabesse et al. 2005). Several lines of evidence from the present study suggest that this hypothesis is not true. The data obtained, for the

presence and/or absence of characteristic histone marks for active promoters suggest that MAP17 promoter may not be active in K562 or Jurkat. In addition, other data including the average levels of nucleosomes at the MAP17 gene in both these cell lines supports this idea. Another interesting observation in K562 is the binding interactions of CTCF at the MAP17 promoter region, which indicated that activity of the promoter of this gene may be blocked through interactions with an insulator or that CTCF binds directly to the promoter to control its activity. In Jurkat, CTCF interactions were not seen at the MAP17 promoter although both K562 and Jurkat express SCL as well as MAP17. This could be due to the reason that SCL is expressed inappropriately in Jurkat, and thus the regulation of SCL and therefore MAP17 may be different in K562 and Jurkat. The +39 region which is located upstream of the MAP17 gene, may also represent a putative novel regulatory element. The location of this element with respect to the MAP17 gene suggests that this element may be associated with the regulation of this gene. However, whether this element is associated with MAP17 or SCL, whether SCL and MAP17 expressions are linked or the mechanism of MAP17 regulation in haematopoietic cells is not known and requires further investigation.

7.9.4 Conclusions

The detailed characterization of regulatory regions with respect to the proposed consensus histone code suggests that the code is consistent at the regulatory regions across the SCL locus in SCL expressing and non-expressing cell lines. It was possible to infer the function and activity of novel regions using the code which was also confirmed by *in vitro* assays in some cases. The data presented suggests that the complexity of regulatory elements and interactions at the SCL locus represents an excellent model by which to study transcriptional regulation and identify features which may be applicable to mammalian genomes in general.