Chapter 5

Application of ChIP-chip to elucidate a Comprehensive Set of Regulatory Interactions at the SCL locus

5.1 Introduction

In order to fully understand how SCL is regulated in its biological contexts, it is important to not only identify all the key regulatory elements, but also the full complement of regulatory interactions that control their activity *in vivo*. Although a number of key regulatory elements at the SCL locus are already known, it is still far from clear how these elements interact with each other or with other *cis*- and *trans*-acting factors that play roles in regulating their activity. The location, function and known regulators of these elements have been discussed in chapters 1 and 4 of this thesis. To summarize:

i) **Promoters:** SCL has three promoters p^{1a}, p^{1b} (Aplan et al. 1990a) and p^{EXON4} (Bernard et al. 1992) all of which are sites for histone H3 and H4 acetylation (Delabesse et al. 2005; Chapter 4); p^{1a} is active only in erythroid and megakaryocytic lineages; p^{1b} is active in erythroid and megakaryocytic lineages and in leukaemic T-cells (Bernard et al. 1992; Aplan et al. 1990a); p^{EXON4} is active only in leukaemic T cells (Bernard et al. 1992). It has been shown that GATA-1 is involved in the regulation of p^{1a} in erythroid cells in both human and mouse (Bockamp et al. 1995; Lecointe et al. 1994; Chapter 4).

ii) **Endothelial enhancer:** The endothelial enhancer (at -3/-4 in both human and mouse) is functional in vascular endothelium and haematopoietic progenitors, is associated with histone H3 and H4 acetylation (Delabesse et al. 2005; Chapter 4) and is regulated by Fli-1 and Elf-1 (Sinclair et al. 1999; Gottgens et al. 2004).

iii) **Stem cell enhancer:** The stem cell enhancer (at +18/+19 in mouse and at +20/+21 region in human) directs SCL expression to the vast majority of haematopoietic progenitors and endothelium. It is associated with histone H3 and H4 acetylation (Delabesse et al. 2005; Chapter 4) and is regulated by GATA-2, Fli-1 and Elf-1 (Sanchez et al. 1999; Sanchez et al. 2001; Gottgens et al. 2002).

iv) **Neural Regulators:** Regulatory elements located at SCL p^{1a}, +1, +3 and at +23 region in mouse (SCL p^{1a}, +1, +3 and +23/+24 regions in human) direct SCL expression to specific regions within the brain and spinal cord. Most, but not all, of these are known to be associated with histone H3 and H4 acetylation (Delabesse et

al. 2005; Chapter 4). GATA factors may play roles in regulating some of these elements (Sinclair et al. 1999; Gottgens et al. 2000; Chapter 4).

v) **Erythroid enhancer:** The +40 erythroid enhancer (+51 in human) targets primitive erythroblasts in transgenic mice (Delabesse et al. 2005). It is associated with histone H3 and H4 acetylation (Delabesse et al. 2005; Chapter 4) and the transcription factor GATA-1 binds to this enhancer (Chapter 4).

vi) **SCL p^{EXON4} repressor:** A novel silencer element, tal-RE, located at +14 in human (+13 in mouse) represses the activity of SCL p^{EXON4} to ensure its restricted usage in only T-cells (Courtes et al. 2000). The regulators which bind to tal-RE are not known.

vii) **-8/-9 enhancer:** A region at -8/-9 in mouse (-9/-10 in human) shows enhancer activity in reporter assays (Gottgens et al. 1997) and is a site for histone H3 acetylation (Delabesse et al. 2005; Chapter 4). GATA-1 binds to this region in K562 cells (Chapter 4).

viii) **-7 element:** In chapter 4 of this thesis, a putative regulatory element at -7 (-6 in mouse) was described which binds GATA-1 and is a site for histone H3 acetylation in K562 (Chapter 4) and other haematopoietic cell types (Delabesse et al 2005).

It is now becoming increasingly clear that posttranslational modifications of histones regulate chromatin structure which in turn determines how a particular gene is regulated by various regulatory proteins and the transcriptional machinery (also discussed in chapter 1). Histone modifications can be highly reversible, such as acetylation (on lysine residues) and phosphorylation (on serine and threonine residues), or more stable, such as methylation (on lysine and arginine residues) (Fischle et al. 2003; Lachner et al. 2002) (Figure 5.1). The mechanisms by which histone acetylation and deacetylation affect transcription and other DNA-based processes are thought to involve two major pathways: first, histone acetylation may alter the folding properties of the chromatin fibre, thereby modulating the accessibility of DNA through structural changes (Tse et al. 1998); and second, the lysine residues and their modifications also provide specific binding surfaces for the recruitment of repressors and activators of gene activity.

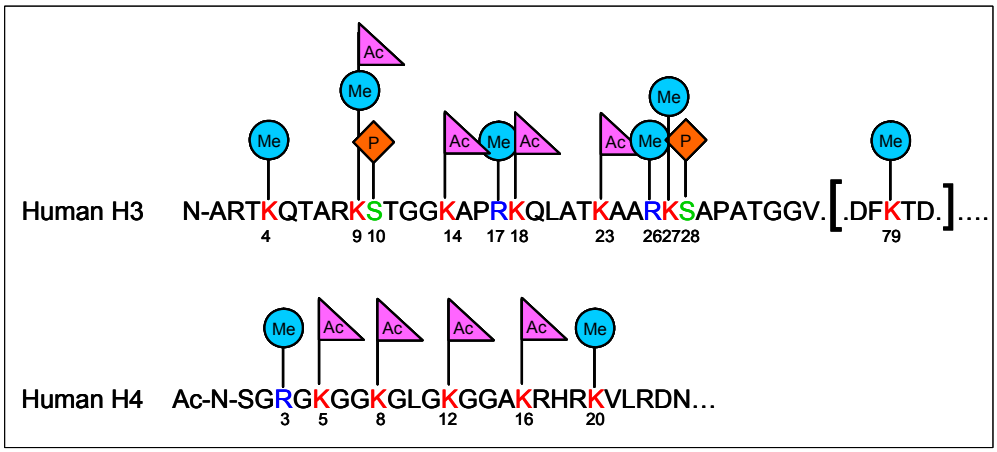


Figure 5.1: The N-terminal tails of histones H3 and H4. The N-terminal tails of H3 and H4 contain highly conserved lysines; K9, K14, K18, K23, K27 and K5, K8, K12 K16 respectively. The blue circles represent methylation, pink triangles represent acetylation and orange diamonds represent phosphorylation marks on various residues of H3 and H4 tails. The lysine residues in the N-terminal tail of H3 that can be methylated include K4, K9, K27 and K36. Lysine K20 can be methylated in H4. In addition, several arginine residues in H3 and H4 can be methylated. Another site of methylation occurs in the histone-fold domain of histone H3 at K79 (Feng et al. 2002; Ng et al. 2002; van Leeuwen et al. 2002). H3 K9 can be acetylated as well as methylated. Lysine residues can be mono-, di- or trimethylated. In addition serine residues at positions 10 and 28 in histone H3 can be phosphorylated.

A number of studies have provided insight into the functions of specific histone modifications. For instance, acetylation at lysine residues is generally associated with transcriptional activity (Wu and Grunstein 2000; Roh et al. 2004). In contrast, methylation of histone tail lysines and arginines has been linked to activation and repression, depending on the residue that is modified (Kouzarides 2002). A number of histone modification enzymes such as acetyltransferases (HATs), deacetylases (HDACs), methyltransferases (HMTs), demethylases and kinases have been identified that exhibit their activity on specific residues of histone H3 or H4. However, the information regarding the patterns of histone modifications at a given gene or entire genome have only begun to be elucidated (Agalioti et al. 2002; Schübeler et al. 2004; Bernstein et al. 2005; Liu et al. 2005). In order to fully comprehend the functional consequences of histone modifications on the transcriptional activity of a gene, it is important to identify the distribution of all of these modifications at a given gene. More specifically, given the development of a highly sensitive ChIP-chip platform for the SCL locus (Chapter 4), a detailed survey of the SCL genomic region to determine the distribution of histone modifications, and binding of transcription factors and other regulatory proteins would be very powerful in answering these questions.

5.2 Aims of this chapter

The aims of the work reported in this chapter were:

1. To identify and test antibodies, raised against various histone H3 and H4 modifications (i.e., acetylation, methylation and phosphorylation) using ChIP-chip assays across the SCL locus in the K562 cell line. Working assays were then used to map regulatory interactions across the SCL locus in SCL expressing and SCL non-expressing cell lines.
2. To further analyse the SCL locus with respect to histone H3 levels to determine the underlying nucleosome occupancy and relationships to gene activity in SCL expressing and SCL non-expressing cell lines.
3. To identify interactions of various transcription factors and other regulatory proteins across the SCL locus in SCL expressing and SCL non-expressing cell lines.

5.3 Overall strategy

This chapter describes the use of ChIP-chip assays to survey DNA-protein interactions across the SCL locus in SCL expressing (K562, Jurkat) and non-expressing (HL60, HPB-ALL) cell lines. The strategy used for these cell lines is outlined in Figure 5.2. The first step was to identify a range of antibodies raised to various histone modifications, transcription factors and other regulatory proteins (Figure 5.2). These were then tested in ChIP-chip experiments to distinguish working assays from non-working ones in K562. The antibodies which reported significant ChIP enrichments in K562 were then further used in the other three cell lines to construct maps of DNA-protein interactions across the SCL locus. Based on this work, it was possible to compare these cell lines with respect to regulatory interactions across the SCL locus, and correlate the findings with the expression of SCL and its flanking genes.

Although, the focus of the study presented in this thesis was to elucidate regulatory interactions at the SCL locus in human haematopoietic cell lines, the availability of the mouse SCL genomic tiling path array and relevant cell lines made it possible to perform some of the assays in mouse. These experiments are also described in this chapter.

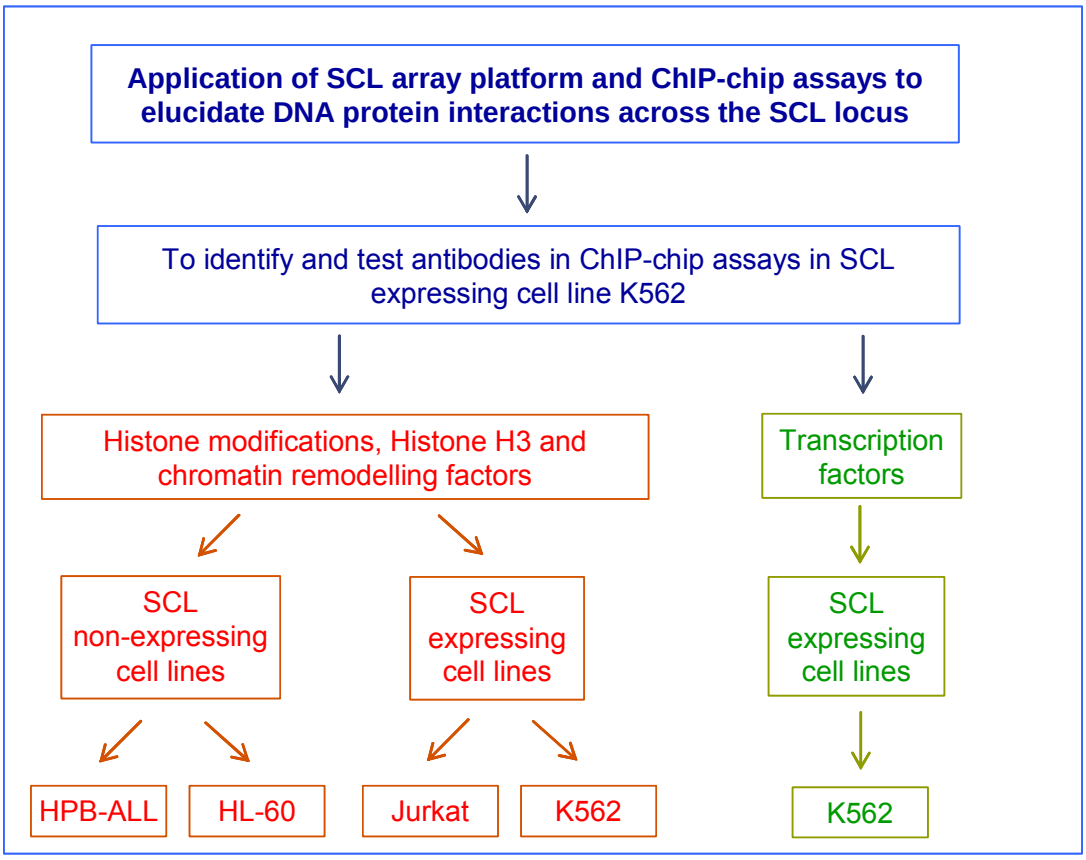


Figure 5.2: Overall strategy employed to map regulatory interactions at the SCL locus using ChIP-chip. This flow diagram shows an overview of the strategy employed to study interactions in four human haematopoietic cell lines. In addition, a sub-set of experiments were performed in mouse cell lines.

Results

5.4 Identifying and testing antibodies to use in ChIP-chip assays

In order to provide a detailed characterization of regulatory interactions at the SCL locus, it was important to identify a host of antibodies which work well in ChIP-chip, the assays for which reflect a range of biological activities involved in transcriptional regulation. The human haematopoietic cell line, K562, was selected to test all the antibodies in the ChIP-chip assays for several reasons. First, K562 had already been used successfully in the ChIP-chip assays described in chapter 4. Second, most of the human SCL regulatory elements studied previously had been characterized using K562 and were shown to be active in this cell line (Aplan et al. 1990a; Leroy-Viard et al. 1994; Courtes et al. 2000). Lastly, it is a well established cell line for the study of haematopoietic development (Horak et al. 2002).

Table 5.1 lists all the antibodies which were identified and tested in ChIP-chip assays across the SCL locus. The antibodies were selected based on a number of criteria as follows:

i) the antibody was not raised to the DNA binding domain of the protein; if possible the antibody was raised to either the N- or C-terminus of the protein

ii) the specificity of the antibody was supported by western blot evidence

iii) the antibody was certified as ChIP grade antibodies by the vendor

iv) the antibody had been shown to perform well in previously published studies

All the antibodies used in the work presented for this thesis had to fulfill at least one of the above-listed criteria. In many instances, antibodies fulfilled all or most of the criteria. To determine whether the antibodies performed well in ChIP-chip experiments in K562, significant ChIP enrichments were calculated for each assay using the statistical analysis described in chapter 4 (section 4.4). Antibodies used in ChIP-chip which gave enrichments which were three standard deviations above background levels were considered as working assays. However, datasets from ChIP-chip assays which measured histone levels were not analysed in this way, given that it was not possible to determine what constituted background levels (i.e., all genomic regions enrich for histones). In these instances, profiles were examined to determine if they differed from profiles obtained from control IgG experiments (using antisera from the appropriate host species in ChIP assays).

Table 5.1. List of antibodies used in ChIP-chip assays. The selected antibodies were raised to various histone modifications, transcription factors and other regulator proteins used in the ChIP-chip experiments. The human and mouse cell lines used in the ChIP-chip experiments are listed in the last 6 columns. All the listed antibodies were first tested in K562 and are colour coded for working (bright pink) and non-working (blue) assays. The criteria for working assays was based on visual assessment of profiles and/or statistical analysis as described in chapter 4 (also see section 5.4). The antibodies that worked successfully in K562 were then further used in other cell lines and are colour coded by light pink.

Classification	Histone	Modification	Function	Company	Cat. No.	Reference	Cell Line					
							HUMAN				MOUSE	
							K562	Jurkat	HL60	HPB-ALL	416B	E14_ES
Acetylation	H3	Lys9/14	gene activation	upstate	06-599	Litt et al. 2001; Roh et al. 2004						
		Lys9	gene activation	upstate	07-352	Litt et al. 2001; Roh et al. 2004						
		Lys14	gene activation	upstate	07-353	Suka et al. 2001; Kurdistani et al. 2004						
		Lys14	gene activation	abcam	ab2381	Suka et al. 2001; Kurdistani et al. 2004						
		Lys18	gene activation	upstate	07-354	Kurdistani et al. 2004						
		Lys23	gene activation	upstate	07-355	Suka et al. 2001; Kurdistani et al. 2004						
		Lys27	gene activation	upstate	07-360	Suka et al. 2001; Kurdistani et al. 2004						
	H4	Lys5/8/12/16	transcriptional regulation	upstate	06-866	Litt et al. 2001; Roh et al. 2004						
		Lys5/8/12/16	transcriptional regulation	abcam	ab76	Litt et al. 2001; Roh et al. 2004						
		Lys5	gene activation	upstate	07-327	Suka et al. 2001; Litt et al. 2001						
		Lys5	gene activation	abcam	ab1758	Suka et al. 2001; Litt et al. 2001						
		Lys8	transcriptional regulation	upstate	07-328	Agalioti et al. 2002; Suka et al.2001						
		Lys8	transcriptional regulation	abcam	ab1760	Agalioti et al. 2002; Suka et al.2001						
		Lys12	transcriptional regulation	upstate	07-323	Agalioti et al. 2002; Suka et al.2001						
		Lys12	transcriptional regulation	abcam	ab1761	Agalioti et al. 2002; Suka et al.2001						
		Lys16	transcriptional regulation	upstate	07-329	Dion et al. 2005; Suka et al. 2001						
		Lys16	transcriptional regulation	abcam	ab1762	Dion et al. 2005; Suka et al. 2001						
	H2A	Lys5/9/13/15	gene activation	upstate	07-376	Suka et al. 2001						
	H2B	Lys5/12/15/20	gene activation	upstate	07-373	Myers et al. 2003						
Methylation	H3	Lys4 mono	gene silencing	abcam	ab8895	Pokholok et al. 2005; Liu et al. 2005						
		Lys4 di	poised chromatin	upstate	07-030	Santos-Rosa et al. 2002; Liu et al. 2005						
		Lys4 di	poised chromatin	abcam	ab7766	Santos-Rosa et al. 2002; Liu et al. 2005						
		Lys4 tri	gene activation	abcam	ab8580	Santos-Rosa et al. 2002; Liu et al. 2005						
		Lys9 di	gene silencing	upstate	07-212	Lachner et al. 2003; Litt et al. 2001						
		Lys27 di	gene silencing	upstate	07-421	Lachner et al. 2003; Cao et al. 2002						
		Lys79 tri	gene activation	abcam	ab2621	Pokholok et al. 2005						
		Arg17 di	transcriptional activation	upstate	07-214	Chen et al. 1999; Daujat et al. 2002						
		Arg26 di	transcriptional regulation	upstate	07-215	Chen et al. 1999						
	H4	Lys20 di	cell-cycle regulation	upstate-a	07-367	Fang et al. 2002.						
		Lys20 di	cell-cycle regulation	upstate-b	07-031	Fang et al. 2002.						
		Arg3 di	gene activation	upstate	07-213	Strahl et al. 2001						
Phosporylation	H3	Ser 10	gene activation/mitotic condensation	upstate	05-598	Mahadevan et al. 1991						
		Ser 10	gene activation/mitotic condensation	abcam	ab4442	Mahadevan et al. 1991						
		Ser 28	cell division/mitotic condensation	upstate	07-145	Loury et al. 2003						

Continued in next page

Classification	Histone	Modification	Function			Reference	Cell Line					
							HUMAN				MOUSE	
							K562	Jurkat	HL60	HPB-ALL	416B	E14_ES
Histones	H3	H3	core histone	santa cruz	sc-8654	Pokholok et al. 2005						
		H3	core histone	abcam	ab1791	Pokholok et al. 2005						
		H3.3	histone H3 variant	abcam	ab4263	Ahmad and Henikoff 2002						
	H4	H4 (C-20)	core histone	santa cruz	sc-8658	Pokholok et al. 2005						
		H4	core histone	abcam	ab2423	Pokholok et al. 2005						
	H2A	H2A (C-19)	core histone	santa cruz	sc8648	Vicent et al. 2004						
	H2B	H2B (C-19)	core histone	santa cruz	sc8651	Vicent et al. 2004						
Chromatin remodellers	p300 (C-20)		HAT activity	santa cruz	sc585	Agalioti et al. 2002						
	p300		HAT activity	abcam	ab2830	Agalioti et al. 2002						
	CBP (A-22)		HAT activity	santa cruz	sc369	Agalioti et al. 2002; Bannister et al. 1996						
	CBP		HAT activity	abcam	ab2832	Agalioti et al. 2002; Bannister et al. 1996						
	GCN5 (C-16)		HAT activity	santa cruz	sc6302	Agalioti et al. 2002						
	PCAF		HAT activity	upstate	07-141	Blanco et al. 1998						
	CGBP (N-21)		Early mammalian development	santa cruz	sc12336	Carlone et al. 2001						
	HDAC1 (C-19)		HDAC activitycell growth/proliferation	santa cruz	sc6298	Lagger et al. 2002						
	HDAC2 (C-19)		HDAC activity	santa cruz	sc6296	De Ruijter et al. 2003						
	MeCP2		gene repression	upstate	07-013	Fan et al. 2005						
	CARM1		arginine methylation	upstate	07-080	Chen et al. 1999						
Preinitiation complex	TAF II p250 (L-20)		Preinitiation complex/HAT activity	santa cruz	sc17134	Ruppert et al. 1995						
	TAF II p250 (6B3)		Preinitiation complex/HAT activity	santa cruz	sc735	Ruppert et al. 1995						
	RNA pol II		Trancription initiation/elongation	abcam	ab5408	Kim et al. 2005						
Trancription factors	Elf-1 (C-20)			santa cruz	sc631	Gottgens et al. 2002; Gottgens et al 2004						
	Fli-1 (C-19)			santa cruz	sc356	Gottgens et al. 2002; Gottgens et al 2004						
	GATA-1 (N-6)			santa cruz	sc265	Wadman et al.1997; Horak et al. 2002						
	GATA-1 (M-20)			santa cruz	sc1234X	Wadman et al.1997; Horak et al. 2002						
	GATA-2 (H116)			santa cruz	sc9008	Gottgens et al. 2002						
	GATA-2 (F-20)			santa cruz	sc16044	Gottgens et al. 2002						
	TAL1			Mathieu lab, France		Wadman et al.1997; Vitelli et al. 2000						
	E2A (YAE)			santa cruz	sc416X	Wadman et al. 1997; Vitelli et al. 2000						
	LMO2 (L-17)			santa cruz	sc10498	Wadman et al. 1997; Vitelli et al. 2000						
	LDB1 (N-18)			santa cruz	sc11198X	Wadman et al. 1997; Vitelli et al. 2000						
	Sp-1 (PEP-2)			santa cruz	sc59X	Lahlil et al. 2004						
	CTCF			abcam	ab10571	Torrano et al. 2005						

5.5 Constructing maps of *in vivo* DNA-protein interactions at the SCL locus

ChIP assays and microarray hybridisations were performed as described in chapter 2. The ChIP DNAs obtained in all the assays described in this chapter were not amplified prior to hybridisation. In order to be able to obtain data for a large number of interactions, only one biological ChIP replicate was performed for each working assay; from the resultant ChIP DNA only one hybridisation was performed. However, given the high reproducibility of the SCL genomic tiling path array in ChIP-chip assays (see chapter 4), it would still be possible to obtain biologically meaningful data using this approach. All the ChIP-chip profiles which were generated for histone modifications in human and mouse cell lines have been compiled in Appendices 6-11. The ChIP-chip profiles described in this chapter and shown in the appendices have been annotated showing the regions with significant enrichments; the numbering system is based on their distance upstream (-) or downstream (+) in kilobases from the SCL promoter 1a (p^{1a}) or from the promoter of the nearest gene.

5.6 Histone H3 acetylation patterns at the SCL locus in SCL expressing and non-expressing cell lines

5.6.1 Histone H3 K9/14 diacetylation correlates with active genes in human cell lines

5.6.1.1 H3 9/14 diacetylation in SCL expressing cell lines

In K562, enrichments for histone H3 K9/14 diacetylation was associated with regions of known regulatory activity across the SCL locus (Chapter 4 and shown in Figure 5.3). Similarly, in Jurkat, the most prominent enrichments were found at or near the 5' ends (i.e. promoters) of KCY, SIL and SCL genes, all of which are expressed in this cell line (Figure 5.3). Acetylation across the 5' end of SCL extended almost 10 kb into its coding regions encompassing SCL p^{1a}, p^{1b}, +1, +3 and SCL p^{EXON4} at the +7 region, the latter of which is known to be active in Jurkat (Bernard et al. 1992). Upstream of SCL, significant enrichments were obtained at -7 and -4 regions, both of which are known to exhibit DNase I hypersensitivity (Leroy-Viard 1994). In addition, significant enrichments were seen at +39 and at p^{MAP17} (+42). Enrichments were low over the +51 erythroid enhancer, but increased over the +53 region. Low level enrichments were also identified at +57. This region was not found to be enriched in K562 and the function of this region is not known. No significant enrichments were observed across CYP4A22 and CYP4Z1 genes.

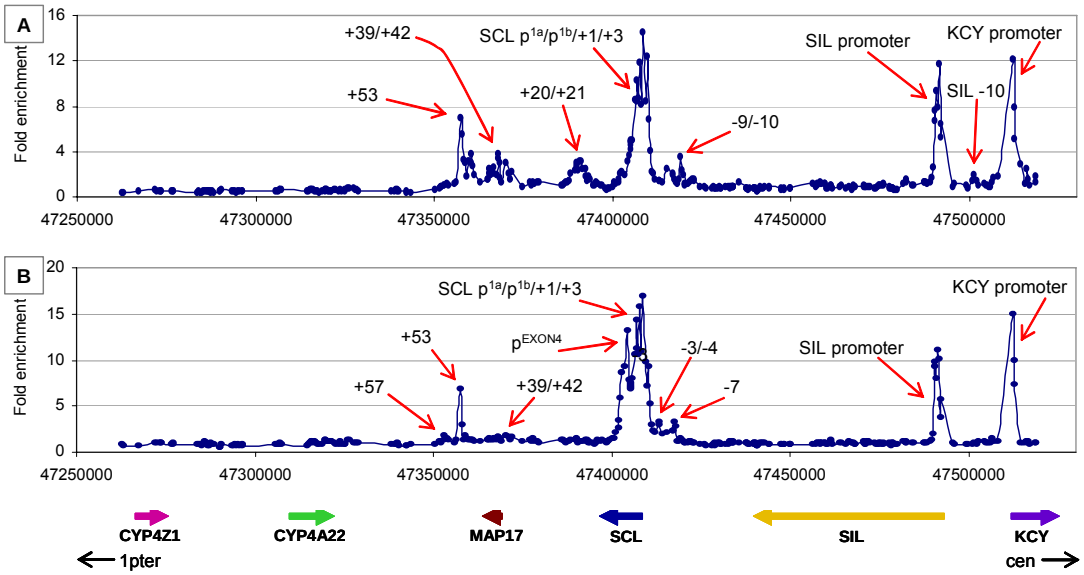


Figure 5.3: ChIP-chip profiles for H3 K9/14 diacetylation across the SCL locus in K562 and Jurkat. Panel A: K562 (as described in chapter 4), Panel B: Jurkat. The x-axes represent the genomic coordinates along human chromosome 1 and the y-axes represent the fold enrichments. Genomic regions showing significant enrichments for H3 acetylation are marked with red arrows. The genomic regions denoted by +1, -3 etc. are based on their distances upstream (-) or downstream (+) in kilobases from the human SCL p^{1a}. The thick, coloured arrows at the bottom of panel B represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and the telomere (ter) on human chromosome 1 is shown by black arrows at the bottom of the figure.

5.6.1.2 H3 K9/14 diacetylation in SCL non-expressing cell lines

Figure 5.4 shows the ChIP-chip profiles generated for H3 K9/14 diacetylation in HL60 and HPB-ALL. As observed in K562 and Jurkat, high levels of enrichments were seen at or near the 5' ends (promoters) of the KCY and SIL genes, both of which are expressed in these two cell lines. At the 5' end of the SCL gene, no significant enrichments were obtained in HPB-ALL, but in HL60, subtle but significant enrichments were present at the +1 region. These absent or greatly reduced levels of acetylation at the 5' end of SCL are in agreement with the absence of SCL expression in these cell lines. Enrichments were also observed at +39 in both cell lines, and at p^{MAP17} in HL60. Furthermore, although neither HPB-ALL nor HL60 express SCL, a region near the +51 erythroid enhancer, at +53, showed high levels of H3 K9/14 diacetylation. This suggests that there may be some genomic sequences in this region, which are active even in the absence of SCL expression. These sequences will be discussed in detail in chapter 7.

In HPB-ALL, a genomic region approximately 4-5 kb downstream of the SIL promoter (SIL +4/+5) also reported significant enrichments. No significant enrichments were observed across the CYP4A22 and CYP4Z1 genes.

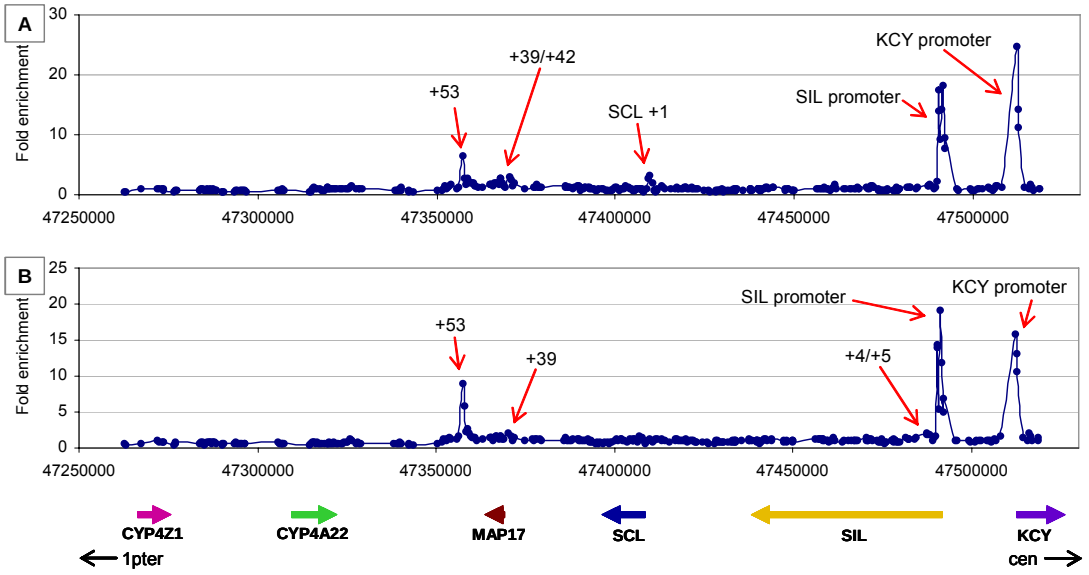


Figure 5.4: ChIP-chip profiles for H3 K9/14 diacetylation across the SCL locus in HL60 and HPB-ALL. Panel A: HL60, Panel B: HPB-ALL. Genomic regions showing significant enrichments for H3 K9/14 diacetylation are marked with red arrows. The x-axes represent the genomic coordinates along human chromosome 1 and the y-axes represent fold enrichments. The genomic regions denoted by +1, +39 etc. are based on their distances downstream (+) in kilobases from the human SCL p^{1a}. The thick, coloured arrows at the bottom of panel B represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown with black arrows at the bottom of the figure.

5.6.2 Histone H3 K9/14 diacetylation is present at SCL regulatory regions in mouse ES cells which do not express SCL

H3 K9/14 diacetylation profiles were determined for the mouse ES cell line E14 (Hooper et al. 1987) and compared to that obtained for the mouse cell line 416B (Chapter 4) as shown in Figure 5.5. 416B expresses SCL, while E14 does not – thus, analysis of these cell lines provides an interesting parallel to the analysis that was performed on the human SCL-expressing and non-expressing cell lines described in this chapter.

Both cell lines showed significant enrichments at or near the 5' end of SIL gene - which is expressed in both 416B and E14 ES cells. However, both cell lines also showed enrichments for H3 K9/14 diacetylation at SCL regulatory regions irrespective of the fact that the E14 cell line does not express SCL. These acetylated SCL regulatory regions encompassed sequences which are known to show DNase I hypersensitivity, enhancer or promoter activity (+1, +3 and SCL p^{EXON4} at +7). The presence of enrichments for H3 K9/14 acetylation at SCL regulatory regions in E14 ES cells suggests that substantial H3 acetylation can occur at genes in the absence of their expression in ES cell lines.

Although this result was incongruous with previous findings which suggest H3 acetylation is associated with gene activity (Schübeler et al. 2004; Bernstein et al. 2005; Liu et al. 2005), there were differences in the profiles obtained for these two mouse cell lines with respect to SCL regulatory regions. First, levels of enrichments at these regions were higher in the SCL-expressing cell line 416B. Second, in 416B a continuous block of acetylation at the 5' end of the SCL gene spanned the genomic regions starting at -1 and extended almost 12 kb into the coding region. In E14 ES cells, the enrichments over these regions were seen as discrete blocks.

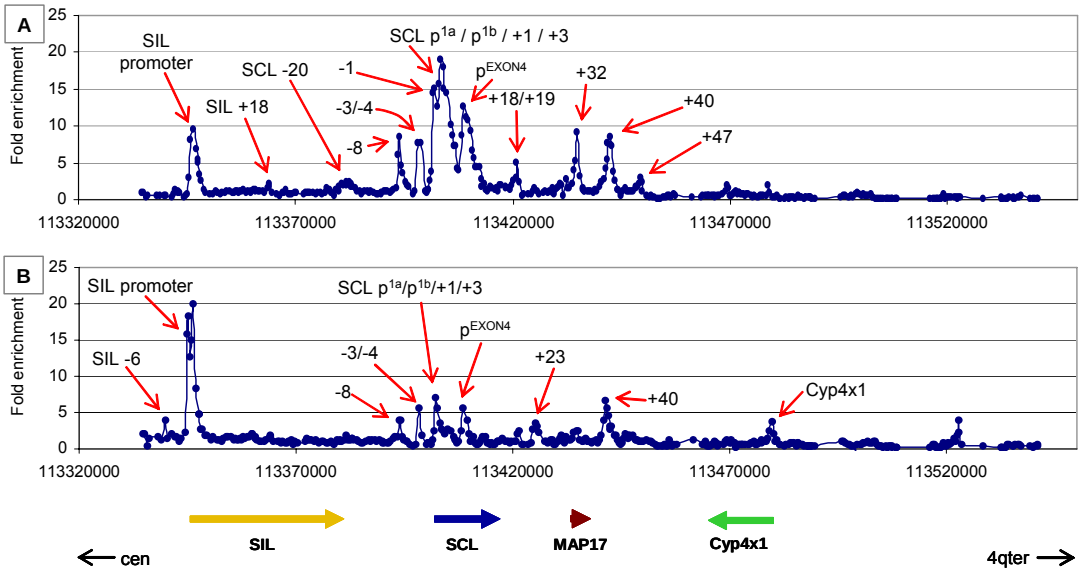


Figure 5.5: ChIP-chip profiles for histone H3 K9/14 diacetylation across SCL locus in 416B and mouse E14 ES cell lines. Panel A: 416B, panel B: mouse E14 ES cell line. Genomic regions showing significant enrichments are marked with red arrows. The x-axes represent genomic coordinates along mouse chromosome 4 and the y-axes represent fold enrichments. The genomic regions denoted by -1, +1 etc. are based on their distance upstream (-) or downstream (+) in kilobases from the mouse SCL p^{1a}. The thick, coloured arrows at the bottom of panel B represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on mouse chromosome 4 is shown with black arrows at the bottom of the figure.

Both cell lines also showed significant enrichments at additional known SCL regulatory elements including the -8/-9 region, the endothelial enhancer at -3/-4 region, the stem cell enhancer at +18/+19 region, the MAP17 promoter (p^{MAP17}) at +32 region and the erythroid enhancer at the +40 region. However, some regions showed significant enrichments in one cell line only. Significant enrichments were seen at the neural enhancer at +23/+24 (Gottgens et al. 2000), at a novel region at SIL -6, and in the cytochrome P450 gene cluster in the E14 ES cell line only. Conversely, 416B showed enrichments at SIL +18, SCL -20, SCL +15 and SCL+47 which were not seen in the E14 ES cell line.

5.6.3 Histone H3 modifications at lysines 9, 14, 18 and 27 correlate with transcriptional activity of genes

To gain a greater insight into the residue-specific acetylation events that modulate regulatory activity, ChIP-chip experiments were performed for acetylation at lysine residues 9, 14, 18, and 27 of histone H3 (Figure 5.6 and Table 5.2). A brief summary will now be described for the principle findings of these ChIP-chip experiments across the SCL locus with respect to each of these modifications in turn.

5.6.3.1 Acetylation of histone H3 at lysine 9 (H3 K9)

It has been shown that acetylation of histone H3 at lysine 9 is found at the predicted transcriptional start sites of active genes and correlated with transcriptional rates genome-wide (Pokholok et al. 2005). In K562, Jurkat, HL60 and HPB-ALL, significant enrichments were obtained at the 5' ends of genes that were active in the respective cell lines (Table 5.2 and Figure 5.6, blue profiles in panels A, B, C and D). In addition, significant enrichments were seen at or near the +51 erythroid enhancer in K562, HL60 and HPB-ALL. These results are in agreement with previous studies and demonstrate that H3 K9 acetylation represents a histone modification which delineates primarily the 5' ends of active genes. Interestingly, the relative levels of ChIP enrichments for H3 K9 acetylation were very low in Jurkat as compared to all the other cell lines (Figure 5.6). Although, the experiment was performed twice in Jurkat, similar results were obtained both times. This suggests that H3 K9 acetylation may not mark the 5' ends of active genes in Jurkat in the same way as it does in the other cell lines analyzed.

5.6.3.2 Acetylation of histone H3 at lysine 14 (H3 K14)

The relative levels of enrichments for this modification were very low in all the cell lines (Figure 5.6, pink profiles in panels A, B, C and D). In fact, no significant enrichments were observed for any region in HL60; and in Jurkat and HPB-ALL, no significant enrichments were obtained at the 5' end of KCY gene. These data suggest that this modification may be used in a cell-type specific manner. Furthermore, in K562, the H3 K14 profile was strikingly similar to that seen for histone H3 K9/14 diacetylation, although additional significant enrichments were also present at a novel region 4 to 5 kb upstream of the KCY gene (KCY -4/-5). The function of this region is not known and may represent a novel putative regulatory region.

	SCL expressing cell lines		SCL non-expressing cell lines	
	K562	Jurkat	HL60	HPB-ALL
H3 K9_Ac	5' end of KCY 5' end of SIL 5' end of SCL (p1a, p1b, +1, +3), p^{EXON4} at +7 p^{MAP17} at +42 at +51 region with two peaks on either side at +50 and +53 regions	5' end of KCY 5' end of SIL 5' end of SCL (p1a, p1b, +1, +3), p^{EXON4} at +7	5' end of KCY 5' end of SIL p^{MAP17} at +42 SCL +50, +51, +53 (three peaks with low levels of enrichment)	5' end of KCY 5' end of SIL SCL +53
H3 K14_Ac	5' end of KCY, KCY -4/-5 5' end of SIL 5' end of SCL (p1a, p1b, +1, +3), p^{EXON4} at +7 p^{MAP17} at +42 SCL -12, -9/-10, -5/-7, -3/-4 SCL +20/+21, +39, +50 and +53 regions (very low level enrichments at +51)	5' end of SCL (p^{1a}, p^{1b}, +1, +3), p^{EXON4} at +7 SCL -9/-10		5' end of SIL
H3 K18_Ac	5' end of KCY, KCY -4/-5 5' end of SIL, SIL -10 5' end of SCL (p1a, p1b, +1, +3) p^{MAP17} at +42 SCL -12, -9/-10, -5/-7 SCL +20/+21, +39, +51 region with two peaks at +51 and +53 regions	5' end of KCY, KCY -4/-5 5' end of SIL 5' end of SCL (p1a, p1b, +1, +3), p^{EXON4} at +7 SCL -9/-10, -7 (peak at -7) , -3/-4	5' end of KCY, KCY -4/-5 5' end of SIL SCL -9/-10 SCL +53	5' end of KCY, KCY -4/-5 5' end of SIL SCL +53
H3 K27_Ac	5' end of KCY, KCY -4/-5 5' end of SIL, SIL -10 5' end of SCL (p1a, p1b, +1, +3), p^{EXON4} at +7 p^{MAP17} at +42 SCL -9/-10 SCL +20/+21, +39, peak at +51 (significant enrichments at +50 and +53)	5' end of KCY, KCY -4/-5 5' end of SIL 5' end of SCL (p1a, p1b, +1, +3), p^{EXON4} at +7 SCL -9/-10, -7, -3/-4	5' end of KCY, KCY -4/-5 5' end of SIL	5' end of KCY, KCY -4/-5 5' end of SIL SCL +53

Table 5.2: Genomic regions across the SCL locus with significant ChIP enrichments for acetylation at lysines 9, 14, 18 and 27 of histone H3 in SCL expressing and non-expressing human cell lines. The specific histone modifications are listed in the first column and the cell lines are listed at the top in adjacent columns. The regions in blue showed significant but very low levels of enrichments and the regions in red represent novel putative regulatory regions which had not been identified previously across the SCL locus. The novel regions are named based on their location with respect to the promoter of the nearest gene. The

numbering is based on the distances upstream (-) or downstream (+), in kilobases, of the promoter of the respective gene. For instance, KCY -4/-5 region was located at four to five kb upstream of the KCY promoter. The majority of the novel regions are annotated on Figure 5.6. Note: regions annotated do not include those identified in the cytochrome P450 gene cluster (see Appendix 6).

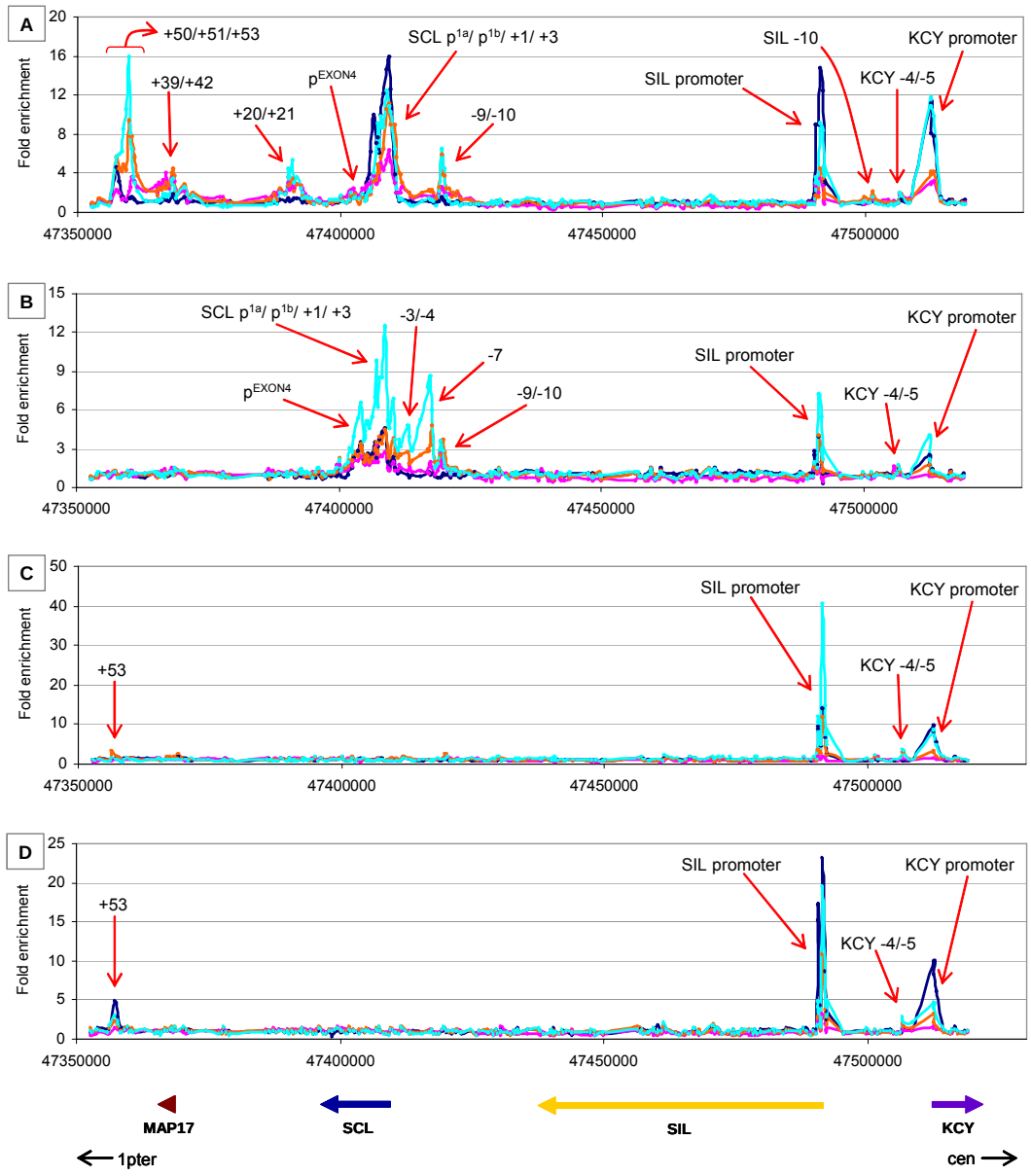


Figure 5.6: Composite ChIP-chip profiles for histone H3 acetylation at lysines 9, 14, 18 and 27 across the SCL locus in the four human cell lines. Panel A: K562, panel B: Jurkat, panel C: HL60 and panel D: HPB-ALL .The above profiles represent only the genomic region across the KCY, SIL, SCL, MAP17 genes and up to the genomic region just downstream of the SCL erythroid enhancer at +51. Blue profile represents H3 K9, pink profile represents H3 K14, orange profile represents H3 K18 and turquoise profile represents H3 K27 acetylation. The genomic regions showing significant enrichments in the ChIP assays for these modifications are marked with red arrows and also listed in Table 5.2. . The x-axes represent the genomic coordinates along human chromosome 1 and the y-axes represent fold enrichments. The thick, coloured arrows at the bottom of the panel D represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) is shown with black arrows at the bottom of the figure.

5.6.3.3 Acetylation of histone H3 at lysine 18 (H3 K18)

In all four human cell lines, significant enrichments for H3 K18 acetylation were observed primarily at the 5' ends of all the genes that were active in the respective cell lines (Table 5.2 and Figure 5.6, orange profiles in panels A, B, C and D). In K562 and Jurkat, the acetylation for H3 K18 at the 5' end of the SCL gene extended almost 9 kb into the coding region. In addition, both K562 and Jurkat showed significant enrichments at most of the regulatory regions which were also identified by H3 K9/14 diacetylation (Table 5.2) including the novel KCY -4/-5 region. No significant enrichments were observed across the CYP4A22 and CYP4Z1 genes in any of the cell lines apart from Jurkat where significant enrichment was present in intron 8 of the CYP4Z1 gene. This data suggests that H3 K18 acetylation marks virtually all the known regulatory regions across the SCL locus of expressed genes and also marks a few unknown regions which may represent novel putative regulatory regions.

5.6.3.4 Acetylation of histone H3 at lysine 27 (H3 K27)

Similarly, significant enrichments for acetylation at lysine 27 of histone H3 marked the 5' ends of the genes that were active in the respective cell lines (Figure 5.6, turquoise profiles in panels A, B, C and D). In addition, significant enrichments were also present at the novel KCY -4/-5 region in all of the cell lines. In K562 and Jurkat, significant enrichments were obtained at a number of the other SCL regulatory regions (Table 5.2).

5.7 Histones H2A and H2B acetylation patterns at the SCL locus in SCL expressing and non-expressing cell lines

5.7.1 Histone H2A acetylation mimics that of H3 9/14 diacetylation

ChIP-chip experiments for histone H2A tetra-acetylated at lysines 5/9/13/15 were performed in the K562 cells only. The pattern of histone H2A acetylation was very similar to that of histone H3 K9/14 diacetylation (Figure 6.21, Appendix 6). Significant enrichments were obtained at the 5' ends of KCY, SIL and SCL genes. At the SCL gene, significant enrichments at the upstream regulatory regions were obtained at -9/-10 (Gottgens et al. 1997), at -5/-7, and over the 5' end at p^{1a}, p^{1b}, +1, +3 and p^{EXON4}. In addition, significant enrichments were observed at the stem cell enhancer at +20/+21, the +39 region, p^{MAP17} at +42, and near the +51 erythroid enhancer with levels peaking at the +50 and +53 regions.

5.7.2 Acetylation of histone H2B marks tissue-specific genes

It has been shown that acetylation of histone H2B is a feature of actively transcribed tissue-specific genes and not of housekeeping genes (Myers et al. 2003). ChIP experiments for histone H2B tetra-acetylated at lysines 5/12/15/20 were performed only

in the K562 cell line. In agreement with these previous findings, very subtle but significant enrichments were observed at the 5' end of KCY gene (3.8 fold) but not at the 5' end of SIL gene (Figure 5.7).

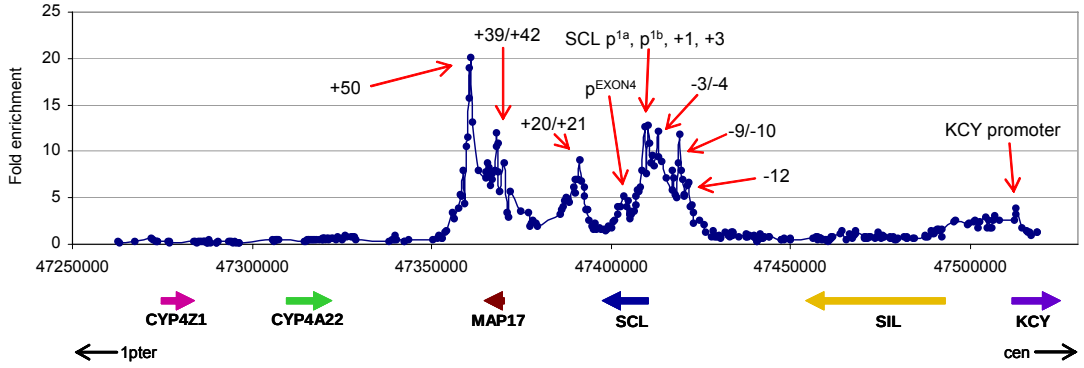


Figure 5.7: ChIP-chip profile of H2B tetra-acetylation at lysines 5/12/15/20 across the SCL locus in K562. Genomic regions with significant enrichments are marked with red arrows. The x-axis represents the genomic coordinates along human chromosome 1 and the y-axis represents fold enrichments. The genomic regions denoted by +1, +3 etc. are based on their distance upstream (-) or downstream (+) in kilobases from the human SCL p^{1a}. The thick, coloured arrows at the bottom of the figure represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) in human chromosome 1 is shown with black arrows at the bottom of the figure.

However, significant and much higher enrichments were obtained across the SCL gene at upstream regions -12, -9/-10, -5/-7, the endothelial enhancer at -3/-4 and the 5' end of SCL at p^{1a}, p^{1b} extending into the +1, +3 and p^{EXON4} (at +7) regions. In addition, significant enrichments were also observed at the stem cell enhancer at +20/+21, +39 region, p^{MAP17} at +42, and the +51 erythroid enhancer. At the +51 region, the region showing significant enrichments extended from +50 to +53 regions with the highest peak at the +50 region.

5.8 Histone H4 acetylation patterns at the SCL locus in SCL expressing and non-expressing cell lines

5.8.1 H4 acetylation coincides with H3 acetylation across the SCL Locus

It has been shown that, like histone H3 acetylation, histone H4 acetylation correlates with the transcriptional activity and transcriptional rates of genes (Pokholok et al. 2005). The results of this thesis are in agreement with these previous observations. ChIP experiments were performed in human cell lines to histone H4 tetra-acetylated at lysines 5/8/12/16 (Figure 5.8 and Table 5.3). In all the cell lines, significant enrichments were seen at or near the 5' ends of the KCY and SIL genes, and at the 5' end of SCL and its regulatory elements in the SCL expressing cell lines K562 and Jurkat (Table 5.3). Although significant enrichments at the p^{EXON4} were found both in K562 and Jurkat, the relative levels of ChIP enrichments at the p^{EXON4} were much higher in Jurkat as

compared to K562. This is in agreement with previous studies which show that SCL p^{EXON4} is active in T-cell lines (Bernard et al. 1992). In Jurkat and HL60, significant enrichments were observed at the SIL -10 and KCY -4/-5 regions, at SIL -10 in K562 and at KCY -4/-5 in HPB-ALL. These regions were also identified previously with a number of histone H3 acetylation modifications (see section 5.6).

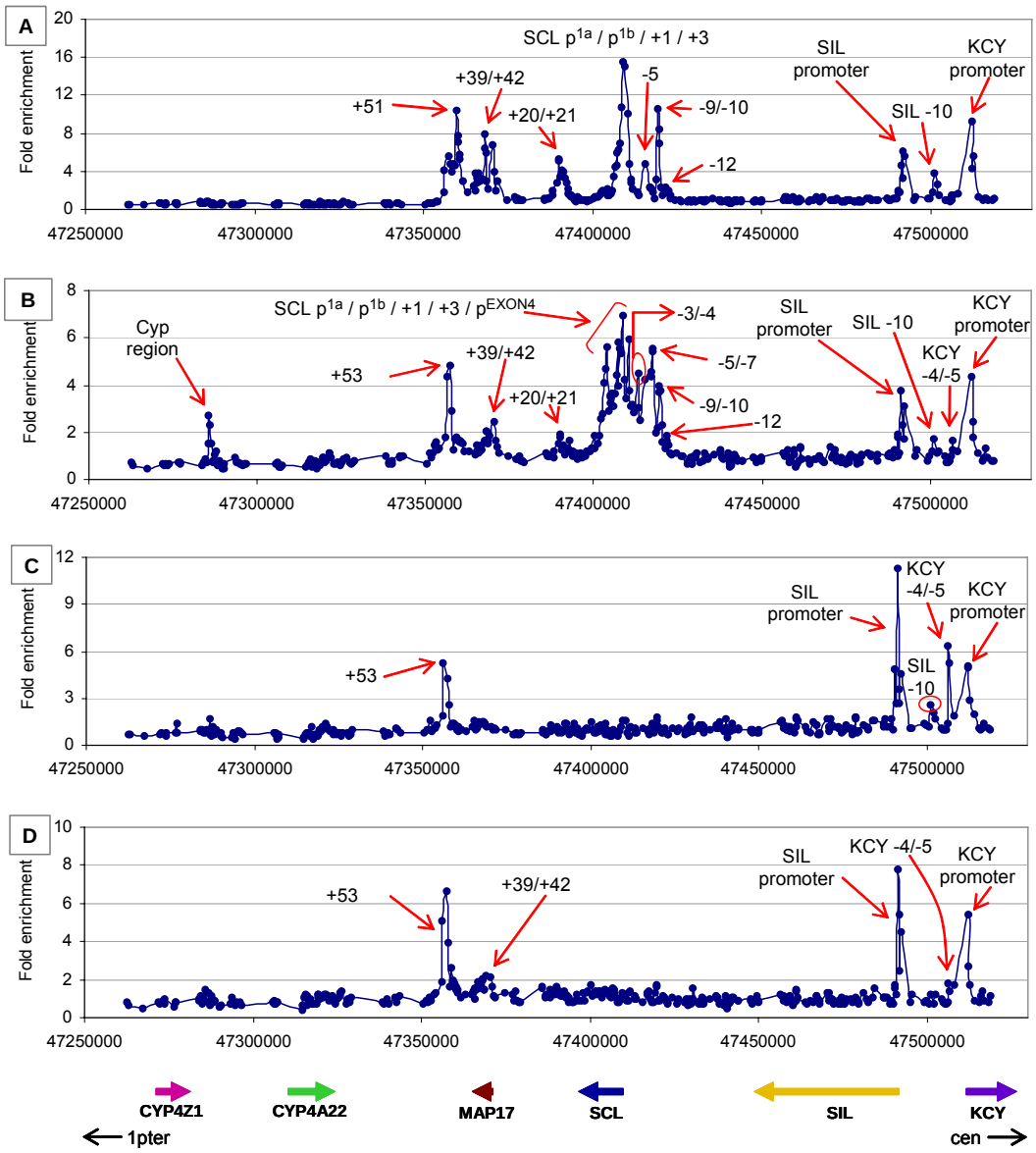


Figure 5.8: ChIP-chip profiles for H4 K5/8/12/16 tetra-acetylation across the SCL locus in the four human cell lines. Panel A: K562, panel B: Jurkat, panel C: HL60 and panel D: HPB-ALL. The genomic regions showing significant enrichments are marked with red arrows. The x-axes represent genomic coordinates along human chromosome 1 and the y-axes represent fold enrichments. The regions denoted by -12, -9/-10 etc. are based on their distance upstream (-) or downstream (+), in kilobases, from SCL p^{1a}. The thick, coloured arrows at the bottom of panel B represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown by the black arrows at the bottom of the figure.

	SCL expressing cell lines		SCL non-expressing cell lines	
	K562	**Jurkat**	**HL60**	**HPB-ALL**
H4 K5/8/12/16 tetra-acetylated	5' end of KCY 5' end of SIL, SIL -10 5' end of SCL (p^{1a}, p^{1b}, +1, +3) p^{EXON4} at +7 p^{MAP17} at +42 SCL -12, -9/-10, -5/-7 SCL +20/+21, +39 SCL +51 region (from +50 to +53 regions)	5' end of KCY, KCY -4/-5 5' end of SIL, SIL -10 5' end of SCL (p^{1a}, p^{1b}, +1, +3) p^{EXON4} at +7 p^{MAP17} at +42 SCL -12, -9/-10, -5/-7, -3/-4 SCL +20/+21, +39 SCL +51 region with highest peak at +53	5' end of KCY, KCY -4/-5 5' end of SIL, SIL -10 SCL +53	5' end of KCY, KCY -4/-5 5' end of SIL p^{EXON4} at +7 SCL +39 SCL +51 region with highest peak at +53
H4 K5_Ac	KCY -4/-5 SIL -10 5' end of SCL (p^{1a}, p^{1b}, +1) p^{EXON4} at +7 p^{MAP17} at +42 SCL -12, -9/-10, -5/-7, -3/-4 SCL +20/+21, +39, two peaks at +50 and +53 (no significant enrichments at +51)	SIL -10 5' end of SCL (p^{1a}, +1, +3) upstream of p^{EXON4} at +7 p^{MAP17} at +42 SCL -9/-10, -3/-4 SCL +20/+21, +39, +51 (three peaks at +50, +51, +53)	 p^{MAP17} at +42 SCL +51 region with highest peak at +53	5' end of SIL (not over p^{SIL}) downstream region of p^{EXON4} (+8) p^{MAP17} at +42 SCL +39, +51 region with highest peak at +53
H4 K8_Ac	5' end of KCY 5' end of SIL 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL +51 SCL -12, -3/-4 p^{EXON4} at +7 SCL +20/+21, p^{MAP17} at +42 SIL -10, KCY -4/-5	5' end of KCY 5' end of SIL 5' end of SCL (p^{1b}, +1, +3) SCL -9/-10, -3/-4 SCL +20/+21, p^{MAP17} at +42 SCL +51 (with three peaks at +50, +51, +53) SIL -10, KCY -4/-5	 5' end of SIL 5' end of SCL (+3) SCL -9/-10 p^{MAP17} at +42 SCL +51 (on either sides at +50, +53 with a peak at +53 region) SIL -10, KCY -4/-5	5' end of KCY 5' end of SIL SCL -9/-10, -7 SCL +3, p^{EXON4} at +7 SCL +20/+21, p^{MAP17} at +42 SCL +51 (and on either sides at +50, +53)
H4 K16_Ac	5' end of KCY 5' end of SIL, SIL -10 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL -9/-10, -5/-7 SCL +39, +51 (extending from +50 to +53 region), +57 p^{EXON4} at +7 SCL +23/+24, p^{MAP17} at +42	5' end of KCY 5' end of SIL 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL -7 SCL +39, +57 SCL +20/+21, p^{MAP17} at +42, +51 with peak at +50	5' end of KCY 5' end of SIL 5' end of SCL (p^{1a}, p^{1b}, +1) SCL +39	5' end of KCY 5' end of SIL

Table 5.3: Genomic regions across the SCL locus showing significant ChIP enrichments for histone H4 acetylation. The results presented in the above table include tetra-acetylation of H4 at lysines 5/8/12/16 and acetylation at individual lysines at 5, 8 and 16. The histone modifications are listed in the first column and the cell lines are listed at the top of the adjacent columns. The ChIP-chip profiles for H4 K8 and H4 K16 showed significant enrichments as well as decreased levels over regulatory regions across the SCL locus. The regions showing decreased levels are listed in the pink coloured sections. The genomic regions listed in the rest of the table were significantly enriched for the respective histone modifications. The regions in blue showed significant but very low levels of enrichments and the regions in red represent novel putative regulatory regions which were not known previously. The novel regions are named based on their location with respect to the promoter of the nearest gene. The numbering is based on the distances upstream (-) or downstream (+), in kilobases, of the respective gene. The regions collectively identified with histone H4 acetylation included the SCL p^{1a}, p^{1b}, +1, +3 and SCL p^{EXON4} at the +7 region (Aplan et al. 1990a; Leroy-Viard et al. 1994; Bernard et al. 1992; Sinclair et al. 1999), -9/-10 (Gottgens et al. 1997), -7 region (Leroy-Viard et al. 1994), the endothelial enhancer at -3/-4 region (Gottgens et al. 2004), the stem cell enhancer at +20/+21 (Gottgens et al. 2002), the erythroid enhancer at +51 (Delabesse et al. 2005).

5.8.2 Acetylation of histone H4 at specific lysine residues correlates variably with transcriptional activity at the SCL locus

Histone H4 can be acetylated at lysines 5, 8, 12 and 16 (Figure 5.1) and it has been suggested that hyperacetylation of H4 at lysine 5 and 12 and hypoacetylation of lysines 8 and 16 are linked to transcriptional activity (Liu et al. 2005). The genomic regions showing significant enrichments for acetylation at lysines 5, 8 and 16 of histone H4 across the SCL locus in the four human cell lines are listed in Table 5.3. Antibodies assaying for H4 K12 acetylation did not produce any significant enrichment in K562 (Table 5.1). Complete profiles are shown in Appendix 7.

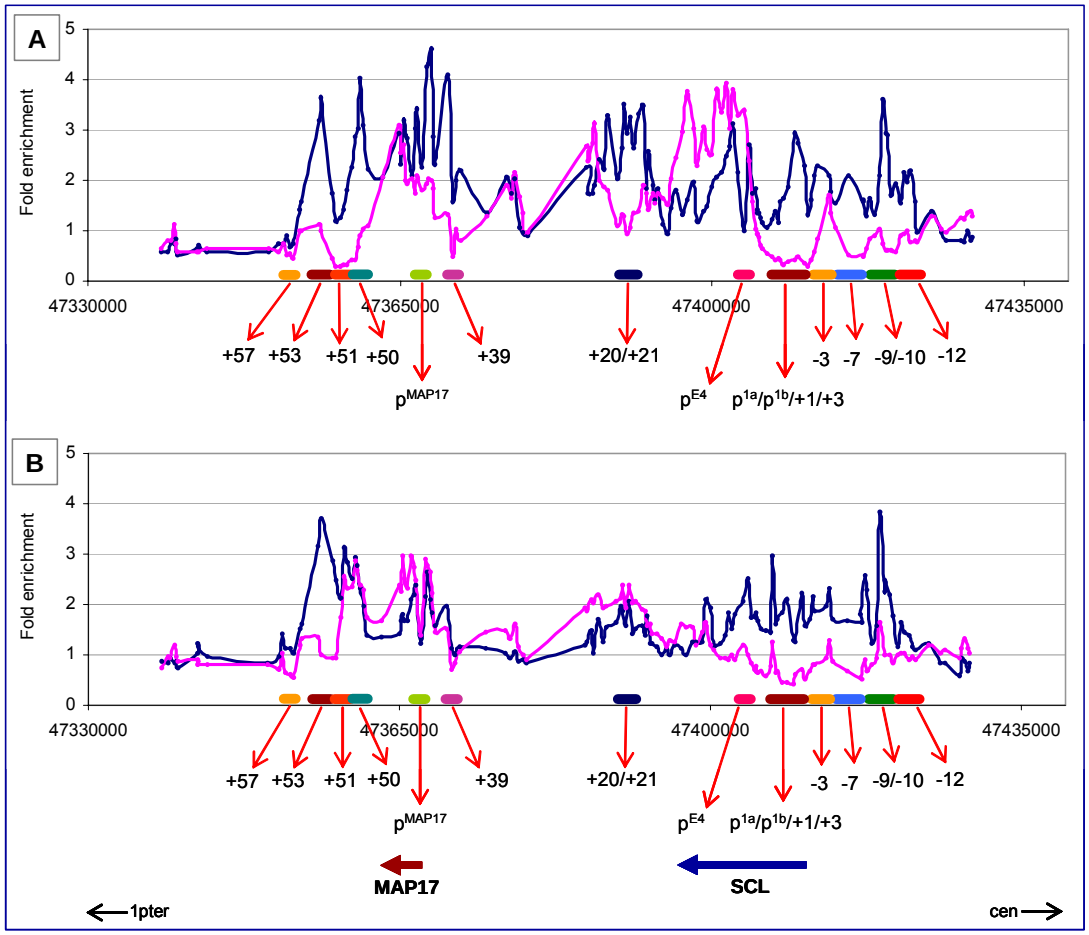


Figure 5.9: ChIP-chip profiles for acetylation at lysines 5 and 16 of histone H4 across the SCL and MAP17 genes. Panel A: K562 and panel B: Jurkat. ChIP-chip profiles for H4 K5 are represented in blue and for H4 K16 in pink across the genomic region containing the SCL and MAP17 genes. The x-axes represent the genomic coordinates along human chromosome 1 and the y-axes represent fold enrichments. The known regulatory regions across the region are represented with coloured bars at the bottom of the plot and annotated with red arrows. The thick, coloured arrows at the bottom of panel B represent the SCL and MAP17 genes and their direction of transcription with respect to the centromere (cen) and telomere (ter) on human chromosome 1. Complete ChIP-chip profiles across the entire SCL locus for all four cell lines are shown in Appendix 7.

5.8.2.1 Acetylation of histone H4 at lysine 5 (H4 K5)

ChIP-chip profiles for H4 K5 acetylation across the SCL and MAP17 genes in K562 and Jurkat are shown in Figure 5.9 (blue profiles in panels A and B). One of the striking observations of the ChIP-chip profiles for H4 K5 acetylation was that the 5' ends (i.e. promoters) of the KCY, SIL and SCL genes were not significantly enriched in any cell lines (Appendix 7). However, a number of the other SCL regulatory regions showed significant ChIP enrichments for H4 K5 acetylation (Table 5.3), although these enrichments were relatively low. This data suggests that H4 K5 acetylation was not linked appreciably to gene activity across the SCL locus and, in particular, is not associated with active promoter regions at the 5' ends of genes.

5.8.2.2 Acetylation of histone H4 at lysine 16 (H4 K16)

Figure 5.9 shows ChIP-chip profiles for H4 K16 acetylation across the SCL and MAP17 genes in K562 and Jurkat cell lines. In all of the cell lines, decreased relative levels were seen at or near the 5' ends (i.e. promoters) of active genes. However, in the SCL non-expressing cell lines, decreased levels were also observed at the 5' end of SCL gene in HL60, while HPB-ALL showed increased levels across this promoter (Appendix 7). The results obtained across all four cell lines suggest that hypo H4 K16 acetylation may be linked to gene activity at the SCL locus. In other words, increases in H4 K16 acetylation may be linked to inactive genes.

5.8.2.3 Acetylation of histone H4 at lysine 8 (H4 K8)

ChIP-chip profiles for H4 K8 acetylation also showed significant enrichments at various regulatory regions across the SCL locus in all of the cell lines (Table 5.3). However, as was the case for H4 K16 acetylation, decreased levels of H4 K8 acetylation were seen at the 5' end of genes that were active in the relevant cell lines. These results are in general agreement with those obtained for H4 K16, which would suggest that the presence of both H4 K8 and K16 acetylation is inversely related to gene activity. This agrees with observations previously reported (Liu et al. 2005).

5.9 Histone H3 lysine methylation patterns at the SCL locus in SCL expressing and non-expressing human and mouse cell-lines

Histone lysine methylation plays an important part in transcriptional regulation and can exist in three states on the lysine residue *in vivo*: mono, di and tri (Strahl et al. 1999). Methylation of lysine 4 of histone H3 has been associated to both the activation and repression status of a gene, depending largely on the degree of methylation of this residue. Trimethylation of lysine 4 of histone H3 has been associated with active genes

and it has been suggested that dimethylation may be linked to "poised" state of chromatin (Santos-Rosa et al. 2002; Ng et al. 2003; Schneider et al. 2004) whereas monomethylation of this residue has been considered to be a mark for repressed or inactive DNA regulatory sequences in chromatin (van Dijk et al. 2005). Similarly, in studies in *Saccharomyces cerevisiae*, the monomethylation of H3 lysine 79, which lies in the central globular domain of the histone H3, has been implicated in gene silencing (Ng et al. 2002; van Leeuwen et al. 2002). Dimethylation of H3 K79 is found to be correlated with active genes (Im et al. 2003) and trimethylation of H3 K79 is associated with coding regions (Pokholok et al. 2005). However, the function and distribution of trimethylation of H3 at lysine 79 in mammalian genomes is not yet clear. Thus, the distribution of these modifications across the SCL locus would provide an insight into how histone lysine methylation states are associated with the SCL regulatory elements and their regulation.

5.9.1 The 5' ends of active genes are hypomethylated for histone H3 lysine monomethylation (H3 K4Me1)

5.9.1.1 Monomethylation of histone H3 lysine 4 in SCL expressing cell lines in human and mouse

As stated earlier, SCL p^{1a} and p^{1b} are transcriptionally active in the erythroid and primitive myeloid cell lines K562 and 416B (Aplan et al. 1990a; Bockamp et al. 1997) whereas, p^{1b} and p^{EXON4} at the +7 region are known to be active in T-cell lines such as Jurkat (Bernard et al. 1992). ChIP-chip profiles for monomethylation at lysine 4 of histone H3 across the SIL, SCL and MAP17 genes in K562, Jurkat and 416B cell lines are shown in Figure 5.10 (profiles across the entire locus are shown in Appendices 9 and 11).

In the human cell lines K562 and Jurkat, lack of enrichments for H3 K4 monomethylation were obtained at or near the 5' ends (i.e. promoters) of the transcriptionally active KCY, SIL and SCL genes. This pattern of decreased levels was also observed for SIL and SCL in 416B cell line (KCY gene is not represented on the mouse SCL array). In the three cell lines, the region at the 5' end of the SCL gene with decreased levels of enrichments extended up to 4 kb into the coding region. The 5' end of the SCL gene showing lack of enrichments for H3 K4 monomethylation encompassed a number of known regulatory regions which had previously been reported to show enhancer activity or DNase I hypersensitivity (see Table 5.4). In Jurkat, decreased H3 K4 monomethylation was observed over the SCL promoter within exon 4 (p^{EXON4}) which is active in this cell line (Bernard et al. 1992).

However, sequences adjacent to active promoter or inactive promoters displayed significant enrichments for H3 K4 monomethylation in all three cell lines (see Table 5.4, yellow coloured section). In K562 and Jurkat, increased levels of enrichments were

observed downstream of both the KCY and SIL promoters and in 416B, enrichments were seen upstream as well downstream of the SIL promoter. At the SCL gene, increased levels were observed at p^{EXON4} (at +7) in K562 correlating with the transcriptional inactivity of this promoter in K562. In Jurkat and 416B, increased levels of enrichment were observed at genomic sequences either side of p^{EXON4}. These data support the idea that decreased levels of H3 K4 monomethylation, in combination with increased levels at either side, are hallmarks of active promoters.

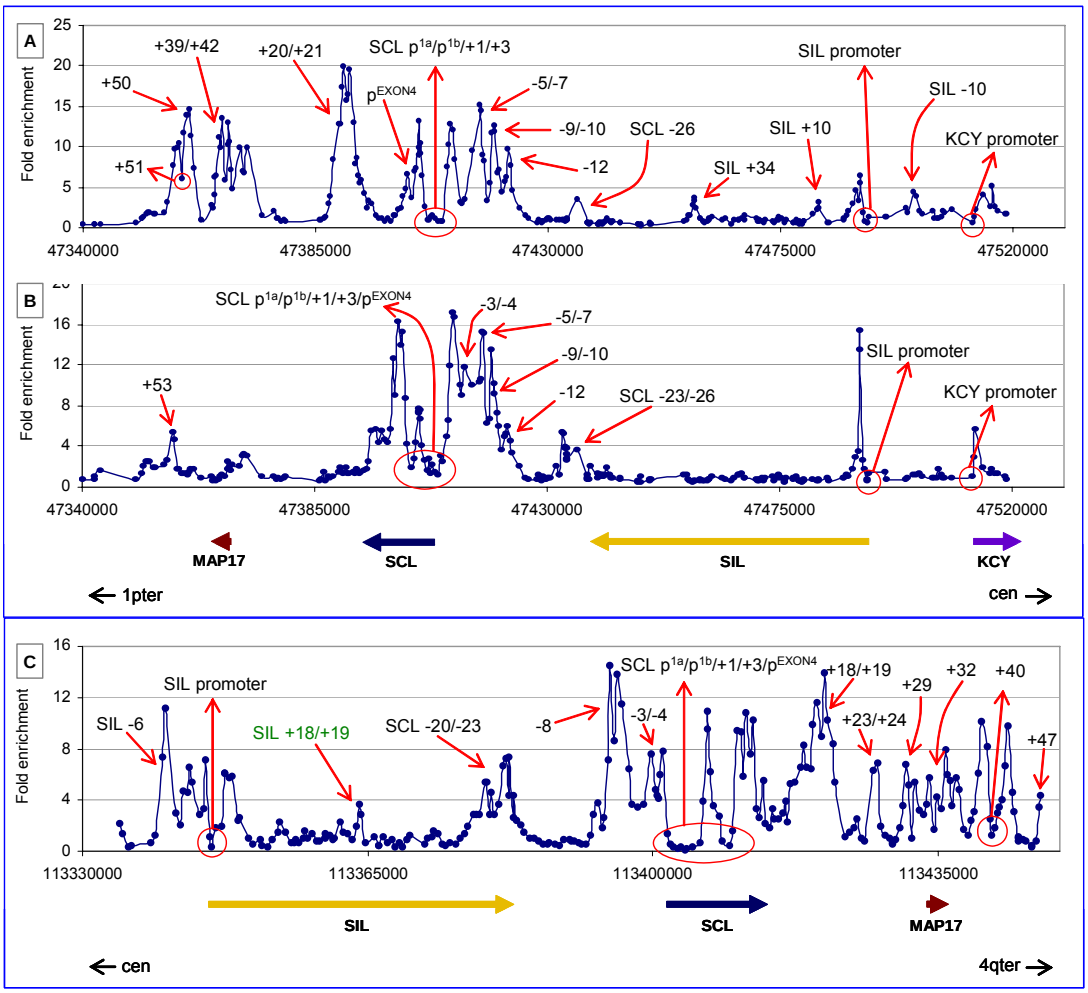


Figure 5.10: ChIP-chip profiles for monomethylation of H3 K4 across the SCL locus in SCL expressing human and mouse cell lines. Panels A and B show the profiles across KCY, SIL, SCL and MAP17 genes in K562 and Jurkat respectively. Panel C represent the profile across the SIL, SCL and MAP17 genes in 416B cell line. The genomic regions showing significant enrichments are marked by red arrows. The genomic regions showing lack of enrichments for monomethylation of H3 K4 are highlighted by red circles. The x-axes in panels A and B represent genomic coordinates for human chromosome 1 and for mouse chromosome 4 in panel C. The y-axes in all three panels represent fold enrichments. The genomic regions are annotated based on their distance (in kilobases) upstream (-) or downstream (+) to the promoter of the nearest gene. For instance, in panel A, SIL +10 and SIL +34 are located 10 and 34 kb downstream of p^{SIL} respectively. Similarly, the SCL -26 region in the same panel is located 26 kb upstream of the SCL p^{1a}. The regions annotated in green were identified within the SIL gene which showed ChIP enrichments but the enrichments were not found to be

significant (see text). The thick, coloured arrows at the bottom of panels B and C represent the gene order and the direction of transcription in human and mouse respectively. The orientation of the SCL loci in human and mouse with respect to the centromere (cen) and telomere (ter) on human chromosome 1 and mouse chromosome 4 are shown by black arrows at the bottom of panels B and C respectively.

Increased H3 K4 monomethylation was associated with the majority of all other known regulatory elements across the SCL loci in human and mouse (Table 5.4; Figure 5.10). In addition, novel genomic regions of unknown function were also identified with significant enrichments for monomethylation in all three cell lines - some of which had been identified previously with a number of histone H3 and H4 modifications (SIL -10 and SCL +39 in human and SCL +47 in mouse) (Table 5.4). Other novel regions identified specifically with monomethylation of H3 K4 are shown in Table 5.4 and Figure 5.10. Many of these H3 K4Me1 novel regions were located within the genomic region of the SIL gene which was used to calculate the background levels to determine significant ChIP enrichments for all assays (Chapter 4). Thus, some of them did not report significant enrichments, since the SDs of background levels in the SIL gene were very high. However, these peaks of enrichment were still highlighted as they were clearly visible by visual inspection and constitute sequences which may have, as yet, an undetermined function in the regulation of the genes across the SCL locus.

5.9.1.2 Monomethylation of histone H3 lysine 4 in SCL non-expressing cell lines in human and mouse

Figure 5.11 shows the ChIP-chip profiles obtained for monomethylation of histone H3 K4 across the SIL, SCL and MAP17 genes in HL60, HPB-ALL and the mouse ES cell line E14 (complete profiles are shown in Appendices 9 and 11). As was observed in SCL expressing cell lines, decreased levels of monomethylation were found at the 5' ends (i.e. promoters) of active genes. In HL60 and HPB-ALL, decreased levels were seen at the promoter regions of KCY and SIL genes, and at the promoter region of the SIL gene in the mouse E14 ES cell line. In E14 ES cells, the promoter region at the 5' end of the SIL gene was bordered by regions of increased levels of monomethylation on either side whereas, in HL60 and HPB-ALL, increased levels of enrichment were present at the downstream regions (Table 5.4). Over the inactive SCL promoters, high levels of monomethylation were observed in both HL60 and mouse E14 ES cells. These findings support the idea that monomethylation levels "drop-out" over active promoters as described in the previous section.

Significant enrichments were also obtained at a number of regulatory regions across the SCL locus (Table 5.4); these were particularly evident in E14 ES cells. Thus, significant monomethylation at H3 K4 can occur at other regulatory sequences across the SCL

gene in the absence of gene expression. Furthermore, E14 ES cells revealed a number of genomic regions within the SIL gene and the cytochrome P450 genes with ChIP enrichments (Figure 5.11 and Appendix 11). Some of these regions were not statistically significant for reasons discussed above.

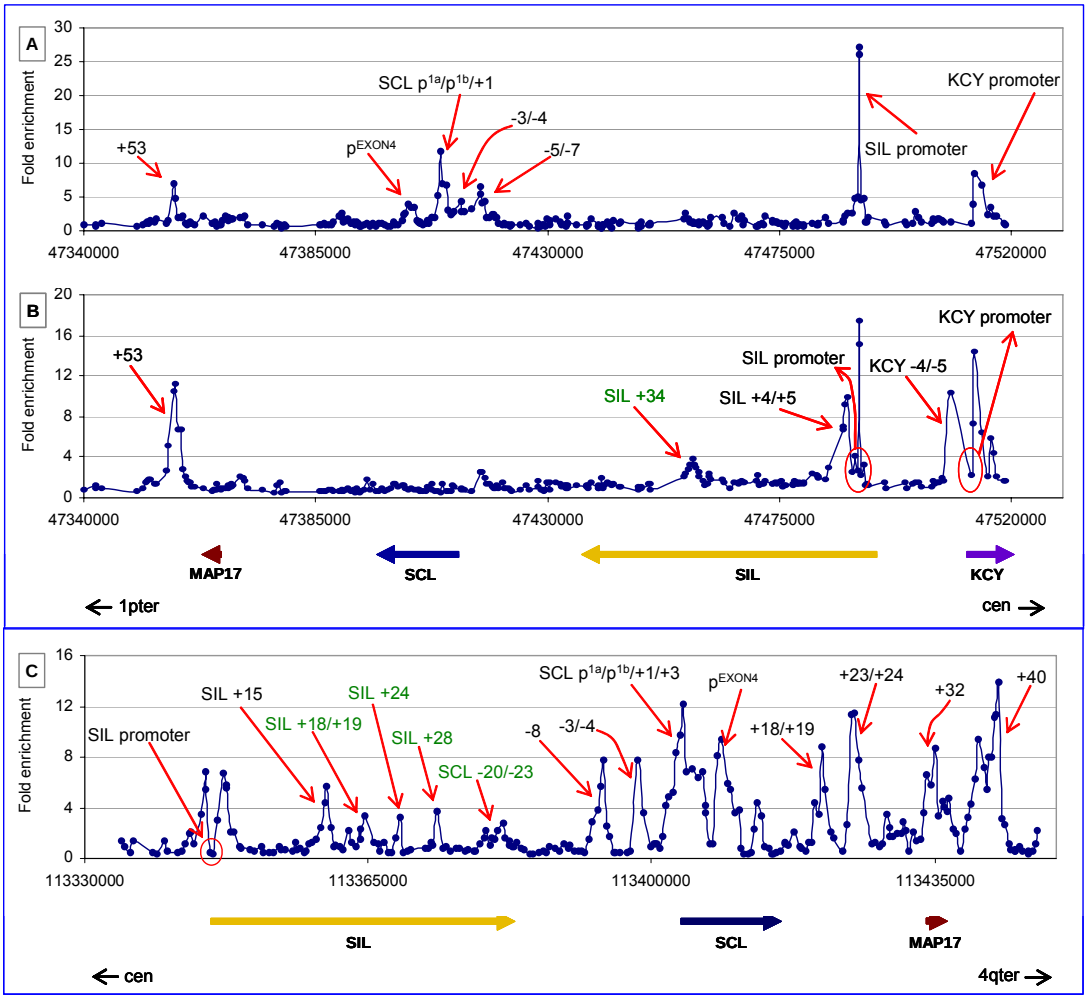


Figure 5.11: ChIP-chip profiles for monomethylation of H3 K4 across the SCL locus in SCL non-expressing human and mouse cell lines. Panels A and B show the profiles across KCY, SIL, SCL and MAP17 genes in HL60 and HPB-ALL respectively. Panel C represents the profile across the SIL, SCL and MAP17 genes in mouse E14 ES cell line. The genomic regions showing significant enrichments are marked by red arrows. The x-axes in panels A and B represent genomic coordinates for human chromosome 1 and in panel C for mouse chromosome 4. The y-axes represent fold enrichments. The genomic regions showing lack of enrichments for monomethylation of H3 K4 are highlighted by red circles. The regions are annotated based on their distance (in kilobases) upstream (-) or downstream (+) to the promoter of the nearest gene. The regions annotated in green text were identified within the SIL gene which showed ChIP enrichments but the enrichments were not found to be significant (for reasons discussed in the text). The thick, coloured arrows at the bottom of panels B and C represent the gene order and the direction of transcription in human and mouse respectively. The orientation of the SCL loci in human and mouse with respect to the centromere (cen) and telomere (ter) on human chromosome 1 and mouse chromosome 4 are shown by black arrows at the bottom of panels B and C respectively.

	SCL expressing cell lines			SCL non-expressing cell lines		
	K562 (Human cell line)	**Jurkat** (Human cell line)	**416B** (Mouse cell line)	**HL60** (Human cell line)	**HPB-ALL** (Human cell line)	**Mouse ES cells** (Mouse cell line)
H3 K4_Me1	5' end of KCY 5' end of SIL 5' end of SCL (p^{1a}, p^{1b}, +1, +3)	5' end of KCY 5' end of SIL 5' end of SCL (p^{1a}, p^{1b}, +1, +3) p^{EXON4} at +7	5' end of SIL 5' end of SCL (p^{1a}, p^{1b}, +1, +3) p^{EXON4} at +7 p^{MAP17} at +32, SCL +40	5' end of KCY 5' end of SIL	5' end of KCY 5' end of SIL	5' end of SIL
	downstream region of p^{KCY} downstream region of p^{SIL} SIL -10, +10, +34 SCL -26 SCL -12, -9/-10, -5/-7 SCL p^{EXON4} at +7 SCL +20/+21, +39, p^{MAP17} at +42 SCL +51 with peaks on either side, highest peak at +50	downstream region of p^{KCY} downstream region of p^{SIL} SCL -23/-26 SCL -12, -9/-10, -5/-7, -3/-4 either side of p^{EXON4} SCL +53	either side of p^{SIL} SIL -6, +18/+19 SCL -20/-23 SCL -8, -3/-4 either side of p^{EXON4} SCL +18/+19, +23/+24, +29, either side of p^{MAP17} SCL +47	downstream region of KCY downstream region of SIL 5' end of SCL (p^{1a}, p^{1b}, +1) SCL -5/-7, -3/-4 SCL p^{EXON4} at +7 SCL +53	downstream region of KCY downstream region of SIL KCY -4/-5, SIL +4/+5, +34 SCL +53	either side of p^{SIL} SIL +15, +18/+19, +24, +28 SCL -20/-23 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL -8, -3/-4 SCL p^{EXON4} at +7 SCL +18/+19, +23/+24, p^{MAP17} at +32 SCL +40
H3 K4_Me2	5' end of KCY 5' end of SIL, SIL -10 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL p^{EXON4} at +7 SCL -12, -9/-10, -5/-7 SCL +20/+21, +39, pMAP17 at +42 SCL +51 with highest peak at +53, +57	5' end of KCY 5' end of SIL, SIL -10 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL p^{EXON4} at +7 SCL -9/-10, -5/-7, -3/-4 SCL +39 SCL +51 with highest peak at +53, +57	5' end of SIL, SIL -6, SIL -2 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL p^{EXON4} at +7 SCL -20, -8, -3/-4 SCL +18/+19, +29, p^{MAP17} at +32 SCL +40, +47	5' end of KCY 5' end of SIL 5' end of SCL (p^{1a}, p^{1b}, +1) SCL p^{EXON4} at +7 SCL -5/-7 SCL +51 with highest peak at +53	5' end of KCY 5' end of SIL, SIL +4/+5 SCL +53	5' end of SIL, SIL +24, +28 5' end of SCL (p^{1a}, p^{1b}, +1) SCL p^{EXON4} at +7 SCL -8, -3/-4 SCL +23/+24 SCL +40
H3 K4_Me3	5' end of KCY 5' end of SIL, SIL -10 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL p^{EXON4} at +7 SCL -12, -9/-10, -5/-7 SCL +20/+21, +39, +42 (but not over ^{pMAP17}) SCL +51 with highest peak at +53, +57	5' end of KCY 5' end of SIL, SIL -10 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL p^{EXON4} at +7 SCL -9/-10, -5/-7, -3/-4 SCL +39 SCL +51 with highest peak at +53, +57	5' end of SIL, SIL -6, SIL -2 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL p^{EXON4} at +7 SCL -20, -8, -3/-4 SCL +18/+19, +29, p^{MAP17} at +32 SCL +40, +47	5' end of KCY 5' end of SIL 5' end of SCL (p^{1a}, p^{1b}, +1, +3) SCL p^{EXON4} at +7 SCL -5/-7 SCL +51 with highest peak at +53	5' end of KCY 5' end of SIL, SIL +4/+5 SCL +53	5' end of SIL, SIL +24, +28 5' end of SCL (p^{1a}, p^{1b}, +1) SCL p^{EXON4} at +7 SCL -8, -3/-4 SCL +23/+24 SCL +40

Table 5.4: Genomic regions identified with mono-, di-, trimethylation of histone H3 at lysine 4 in human and mouse cell lines. The histone modifications are listed in the first column and the cell lines are listed at the top in adjacent columns. ChIP-chip assays were performed in human and mouse cell lines. A number of genomic regions for monomethylation showed significant enrichments (shown in the yellow region) as well as lack of enrichments (shown in the pink region). The genomic regions listed for di- and trimethylation showed significant enrichments. The regions in blue text showed very low levels of enrichments, the regions in red text represent novel regions. The regions in green text represent novel regions which showed ChIP enrichments but the enrichments were not found to be significant (see text). The regions are named based on their location with respect to the promoter of the nearest gene. The numbering is based on the distances upstream (-) or downstream (+) in kilobases from the respective genes. The known SCL regulatory regions that were collectively identified with histone H3 methylation included the SCL p^{1a}, p^{1b}, +1, +3 and SCL p^{EXON4} at the +7 region (Aplan et al. 1990a; Leroy-Viard et al. 1994; Bernard et al. 1992; Sinclair et al. 1999), -9/-10 (Gottgens et al. 1997), -7 region (Leroy-Viard et al. 1994), the endothelial enhancer at -3/-4 region (Gottgens et al. 2004), the stem cell enhancer at +20/+21 (Gottgens et al. 2002), the erythroid enhancer at +51 (Delabesse et al. 2005).

5.9.2 Dimethylation and trimethylation at lysine 4 of histone H3 (H3 K4Me2, H3 K4Me3) occur at transcriptionally active genes across the SCL locus in human and mouse

5.9.2.1 Dimethylation and trimethylation at histone H3 lysine 4 in SCL expressing human cell lines

ChIP experiments were performed in human and mouse cell lines using antibodies raised to dimethylated and trimethylated lysine 4 of histone H3 (Table 5.1). Figure 5.12 shows composite ChIP-chip profiles of dimethylation and trimethylation of H3 K4 across the KCY, SIL, SCL and MAP17 genes in K562 and Jurkat cell lines respectively (complete profiles across the entire locus are shown in Appendix 9). The genomic regions which showed significant enrichments for di- and trimethylation of H3 K4 in K562 and Jurkat are listed in Table 5.4.

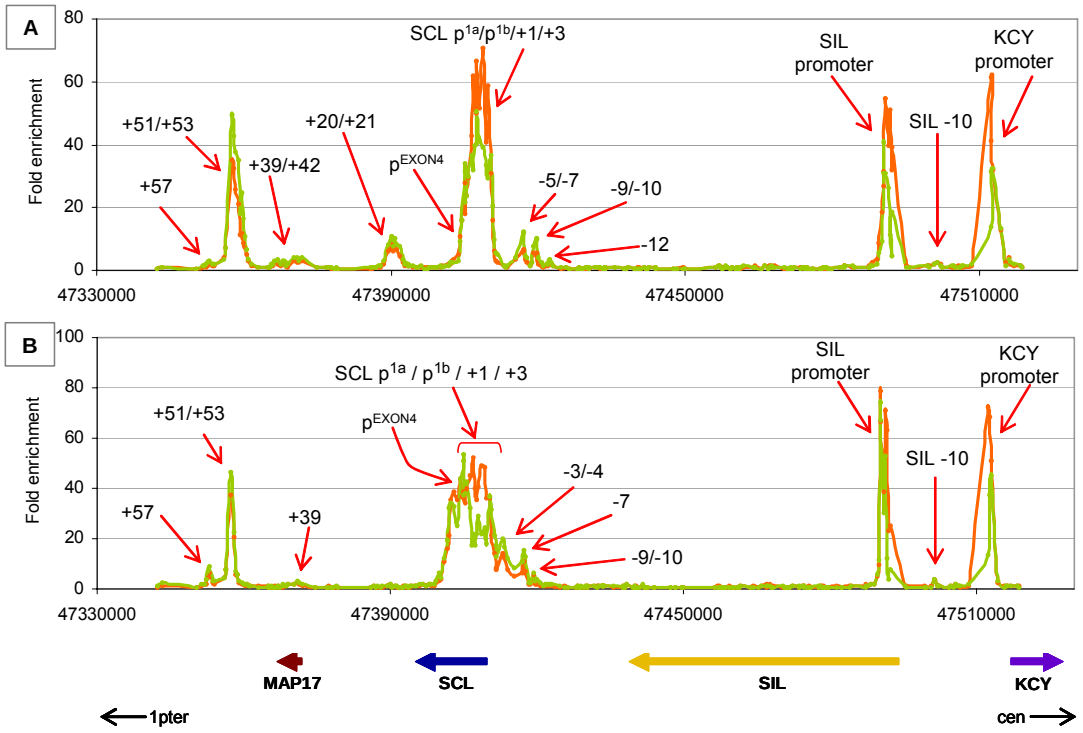


Figure 5.12: ChIP-chip profiles of di- and trimethylated lysine 4 of H3 across the SCL locus in K562 and Jurkat. Panel A: K562 and panel B: Jurkat. The orange profiles in panels A and B represent trimethylation of H3 K4 and the green profiles represent dimethylation of H3 K4. The genomic regions showing enrichments are marked with red arrows. The x-axes represent genomic coordinates along human chromosome 1 and the y-axes represent fold enrichments. The regions are denoted based on their distance upstream (-) or downstream (+) in kilobases from the promoter of the nearest gene. The thick, coloured arrows represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown with black arrows at the bottom of the figure.

In both cell lines, significant levels of enrichment for di- and trimethylation H3 K4 were obtained at or near the 5' ends of the KCY, SIL and SCL genes, all of which are expressed in these cell lines. The highest peaks of enrichments for trimethylation were observed at the promoter regions while the highest peaks for dimethylation were seen just downstream (3') of the promoter regions for all the active genes. In K562, at the 5' end of the SCL gene, both di- and trimethylation extended almost 9 kb into the coding region; in Jurkat, the significant enrichments extended almost 11 kb into the coding region. In both cell lines, significant enrichments were obtained at SCL p^{EXON4} which is known to be active only in Jurkat (Bernard et al. 1992). However, in K562, the relative levels of enrichments at p^{EXON4} were almost ten fold lower as compared to the active promoter p^{1a} and p^{1b}; in contrast, the relative levels of enrichments at p^{EXON4} were very similar to the other active SCL promoters in Jurkat (Figure 5.12, panel B). These observations suggest that levels of di- and trimethylation may reflect the activity of a promoter.

A number of other regulatory regions across the SCL locus showed significant enrichments for di- and trimethylation (Table 5.4). In K562, significant enrichments at the upstream (5') SCL regulatory regions were seen as discrete regions; in Jurkat, the enrichments were present over a continuous block encompassing all the known upstream SCL regulatory regions. This suggested that the regulation of SCL in K562 and Jurkat may be different with respect to these regulatory regions. Interestingly, significant enrichments for both di- and trimethylation were identified at the +57 region which had previously been shown to report significant enrichments with H3 K9/14 diacetylation (in Jurkat only) (Figure 5.3). The function of this region is not known.

5.9.2.2 Dimethylation and trimethylation at histone H3 lysine 4 in SCL non-expressing human cell lines

Figure 5.13 shows composite ChIP-chip profiles of dimethylation and trimethylation of histone H3 K4 across the KCY, SIL, SCL and MAP17 genes in HL60 and HPB-ALL. Significant enrichments were obtained at the 5' end of KCY, SIL genes, and were in agreement with data from K562 and Jurkat showing the association of transcriptional activity and these modifications. Interestingly, in HL60, relatively low level but significant enrichments were also obtained at the 5' end of the SCL gene although SCL is not expressed in HL60; in HPB-ALL, however, no significant enrichments were seen at the 5' end of SCL. Furthermore, although neither HPB-ALL nor HL60 express SCL, a region near the +51 erythroid enhancer, at +53, showed high levels of di- and trimethylation for H3 K4. This data further supports the findings obtained with H3 acetylation in these cell lines, suggesting that there may be some genomic sequences

in this region, which are active even in the absence of SCL expression. These sequences will be discussed in detail in chapter 7.

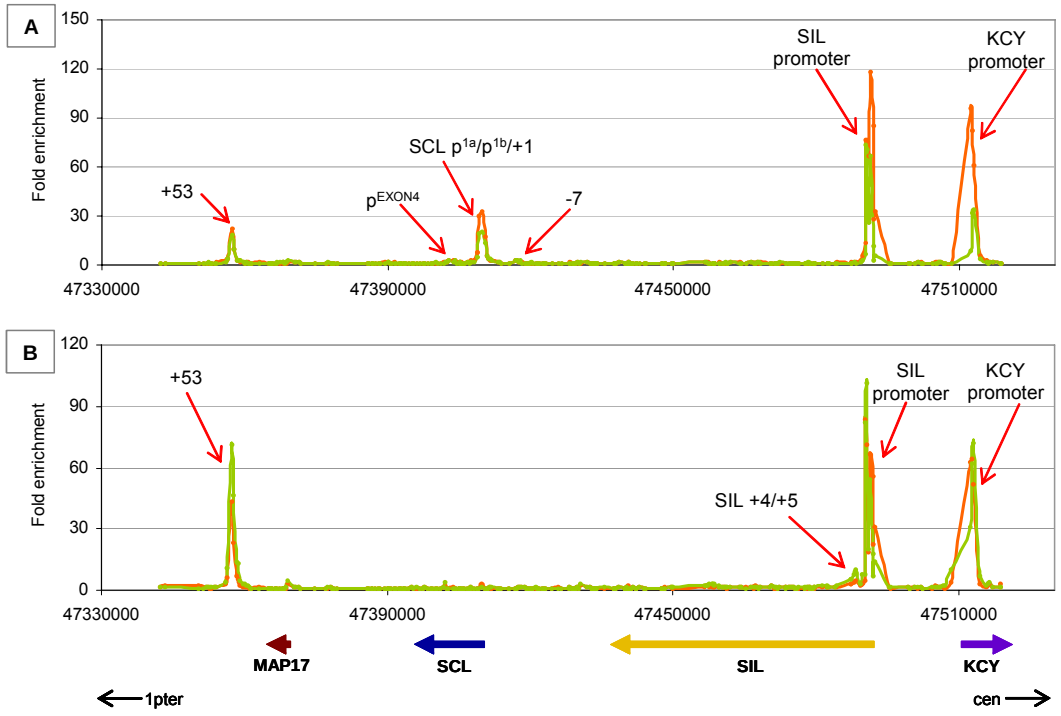


Figure 5.13: ChIP-chip profiles for di- and trimethylated lysine 4 of H3 across the SCL locus in HL60 and HPB-ALL. Panel A: HL60 and panel B: HPB-ALL. The orange profiles in panels A and B represent trimethylation of H3 K4 and the green profiles represent dimethylation of H3 K4. The genomic regions showing enrichments are marked with red arrows. The x-axes represent genomic coordinates along human chromosome 1 and the y-axes represent fold enrichments. The regions are denoted based on their distance upstream (-) or downstream (+), in kilobases, from the promoter of the nearest gene. The thick, coloured arrows represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown with black arrows at the bottom of the figure.

5.9.2.3 Dimethylation and trimethylation at histone H3 lysine 4 in SCL expressing and non-expressing mouse cell lines

Figure 5.14 shows the composite ChIP-chip profiles for dimethylation and trimethylation of H3 K4 across the SCL locus in 416B and mouse E14 ES cell line. As seen in human cell lines, 416B showed significant enrichments for di- and trimethylation of H3 K4 at the 5' ends of the active SIL and SCL genes - with highest relative levels of enrichments for trimethylation at the promoters, and the highest levels for dimethylation present immediately downstream of the promoters. Similarly, significant enrichments for di- and trimethylation were obtained at the 5' end of the SIL and SCL genes in E14 ES cells although SCL is not expressed in this cell line. This again supports previous results obtained for H3 acetylation, suggesting that modifications normally associated with gene activation (in this instance H3 K4 di- and

tri- methylation) can occur in ES cells, in the absence of gene expression (see section 5.6.2). However, unlike 416B and the human cell lines, the highest enrichments for di- and trimethylation in E14 ES cells were found to be over the same regions at the 5' end of the SCL gene (Figure 5.14, panel B).

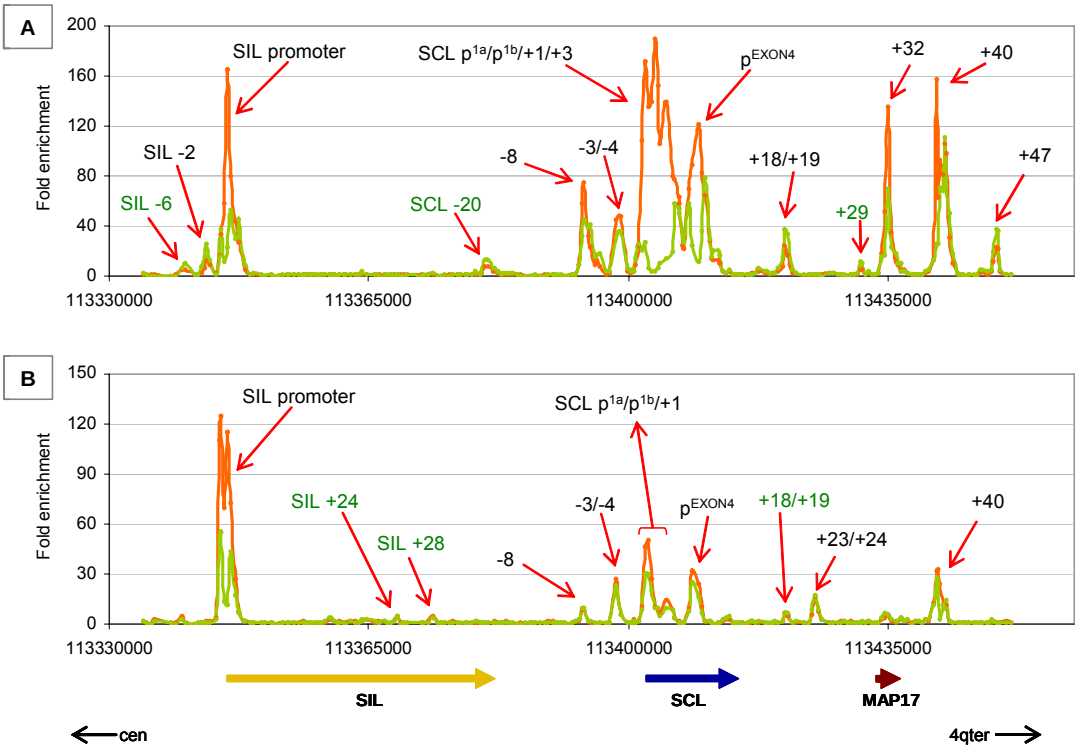


Figure 5.14: ChIP-chip profiles of di- and trimethylated lysine 4 of H3 across the SCL locus in 416B and mouse E14 ES cell line. Panel A: 416B. Panel B: E14 ES cell line. The orange profiles represent trimethylation of H3 K4 and the green profiles represent dimethylation of H3 K4. The genomic regions showing enrichments are marked with red arrows. The x-axes represent genomic coordinates along mouse chromosome 4 and the y-axes represent fold enrichments. The regions are denoted based on their distance upstream (-) or downstream (+), in kilobases, from the promoter of the nearest gene. The regions annotated in green text were identified by visual assessment of the ChIP-chip profiles. The thick, coloured arrows represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on mouse chromosome 4 is shown with black arrows at the bottom of the figure.

A number of other known SCL regulatory regions were found to be significantly enriched for H3 K4 di- and trimethylation in 416B and mouse E14 ES cell lines (Table 5.4). Some of these regions were found to be enriched in both cell lines, but others were specifically enriched for one or the other cell line. For instance, the +47 region was significantly enriched in 416B whereas the neural enhancer at +23/+24 showed significant enrichments in E14 ES cells (Figure 5.14). Both cell lines also identified regions within the cytochrome P450 genes (Appendix 11). As previously seen for monomethylation of H3 K4, a number of novel regions were identified with ChIP

enrichments, which could be seen by visual assessment of the profiles but were not statistically significant. These regions included SIL -6, SCL -20 and SCL +29 in 416B cell line and SIL +24, SIL +28 and the stem cell enhancer at +18/+19 (known) in mouse ES cell line. The functions of these regions (apart from the stem cell enhancer) are not known.

5.9.3 Trimethylation of histone H3 lysine 79 (H3 K79Me3) marks the immediate 3' downstream regions of promoters of active genes

Given that the role of histone H3 K79 trimethylation is not known in human cells, ChIP-chip experiments were performed in four human cell lines using an antibody raised to trimethylated lysine 79 of histone H3 (Table 5.1). Figure 5.15 shows ChIP-chip profiles generated from these experiments in all cell lines across SCL, SIL and KCY genes. The ChIP profiles generated across the entire locus from each cell line have been shown in Appendix 9.

In all of the cell lines, significant enrichments were found at the immediate 3' downstream regions of the promoters of active genes. In K562, the SCL gene showed enrichments over a genomic region which spanned from +3 to +6 (with respect to SCL p^{1a}). Similarly, in Jurkat, enrichments were seen at the genomic region spanning the +2 to +6 regions. However, additional enrichments spanning the SCL +10 region were seen in Jurkat, which were downstream of the SCL p^{EXON4} (located at +7). This result correlates with the known active and inactive states of this promoter in Jurkat and K562 respectively. Conversely, HL60 and HPB-ALL, both of which do not express SCL, showed no significant enrichments across the SCL gene. In all four cell lines, significant enrichments were also obtained in the regions downstream of the SIL promoter (from +1 to +3 region seen as two separate peaks in Figure 5.15, panel B) and also the KCY promoter (from +2 to +4). Taken together, these results suggest that the trimethylation at lysine 79 of histone H3 marks the regions immediately 3' downstream of the promoters of active genes.

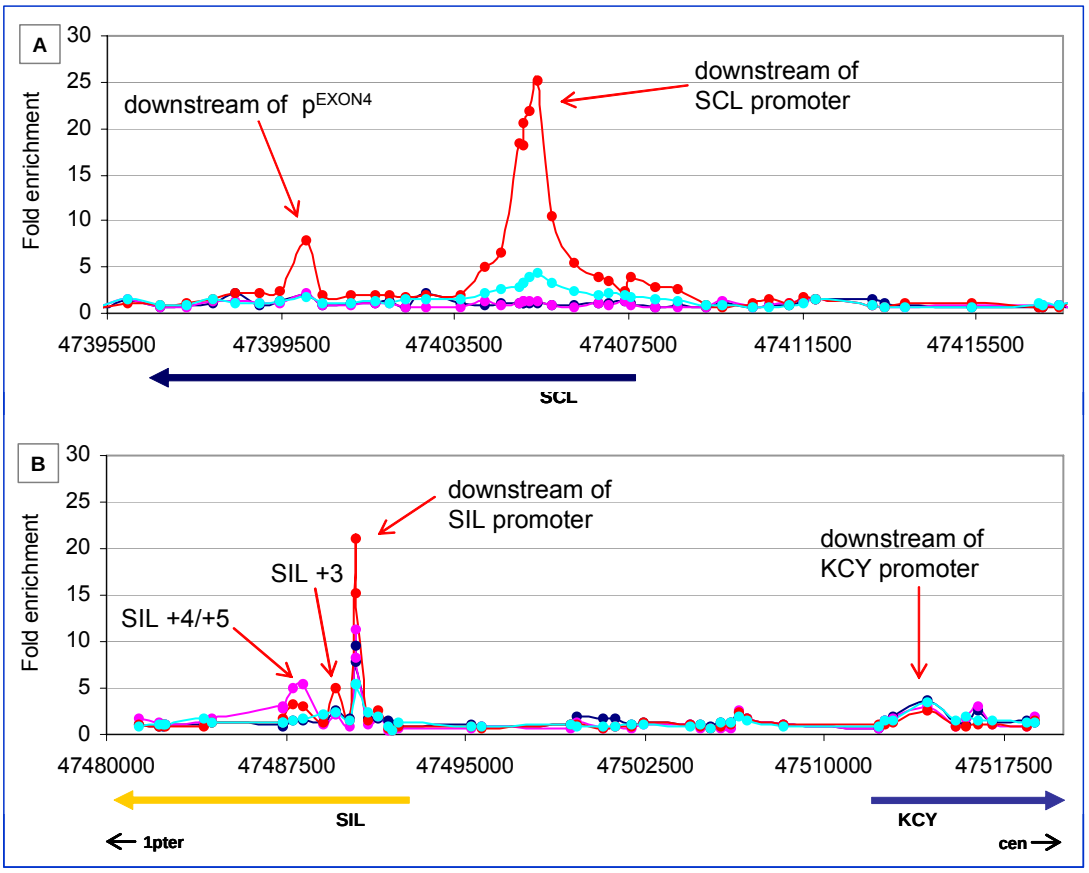


Figure 5.15: ChIP-chip profiles for histone H3 K79 trimethylation across the SCL locus in four human cell lines. The profile in panel A represents only the region at the 5' end of SCL gene and the panel B represents the genomic regions across the 5' ends of KCY and SIL genes. Four colours in the profiles shown above represent the four cell lines: turquoise (K562), red (Jurkat), blue (HL-60) and pink (HPB-ALL). The x-axes represent the genomic coordinates along human chromosome 1 and the y-axes represent fold enrichments. The thick, coloured arrows at the bottom of each panel represent the gene order and the direction of transcription with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown with black arrows at the bottom of panel B.

5.10 Histone H3 phosphorylation at the SCL locus in SCL expressing and non-expressing human cell lines

Histone H3 phosphorylated at serine 10 and 28 occurs during cell division and is also related to the regulation of transcription (Mahadevan et al. 1991; Sassone-Corsi et al. 1999). Figure 5.16 shows the ChIP-chip profiles for phosphorylation of H3 serine 10 across the 5' end of the SCL gene in all four human cell lines (complete ChIP-chip profiles across the SCL locus are shown in Appendix 10).

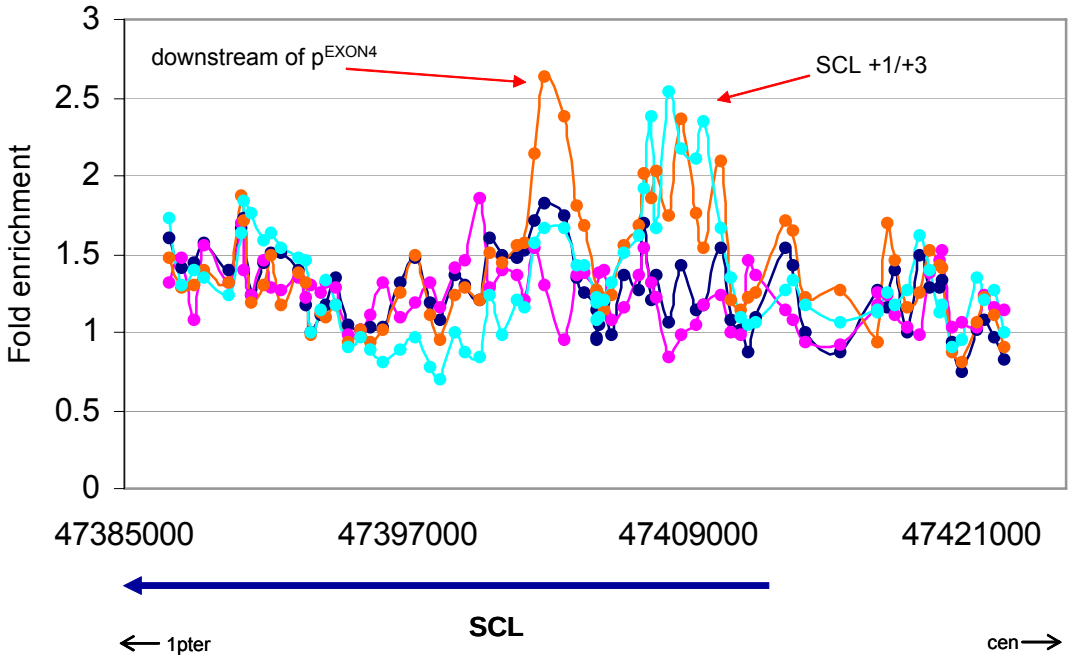


Figure 5.16: ChIP-chip profiles for H3 serine 10 phosphorylation across the 5' end of the SCL gene in four human cell lines. The figure shows ChIP-chip profiles across a small genomic region at the 5' end of the SCL gene only. The four coloured lines in the profile shown above represent the four cell lines: turquoise (K562), orange (Jurkat), blue (HL-60) and pink (HPB-ALL). The x-axis represents the genomic coordinates along human chromosome 1 and the y-axis represents fold enrichments. The thick, coloured arrow at the bottom of the figure represents the SCL gene and the direction of its transcription. The orientation of the gene with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown by black arrows at the bottom of the figure.

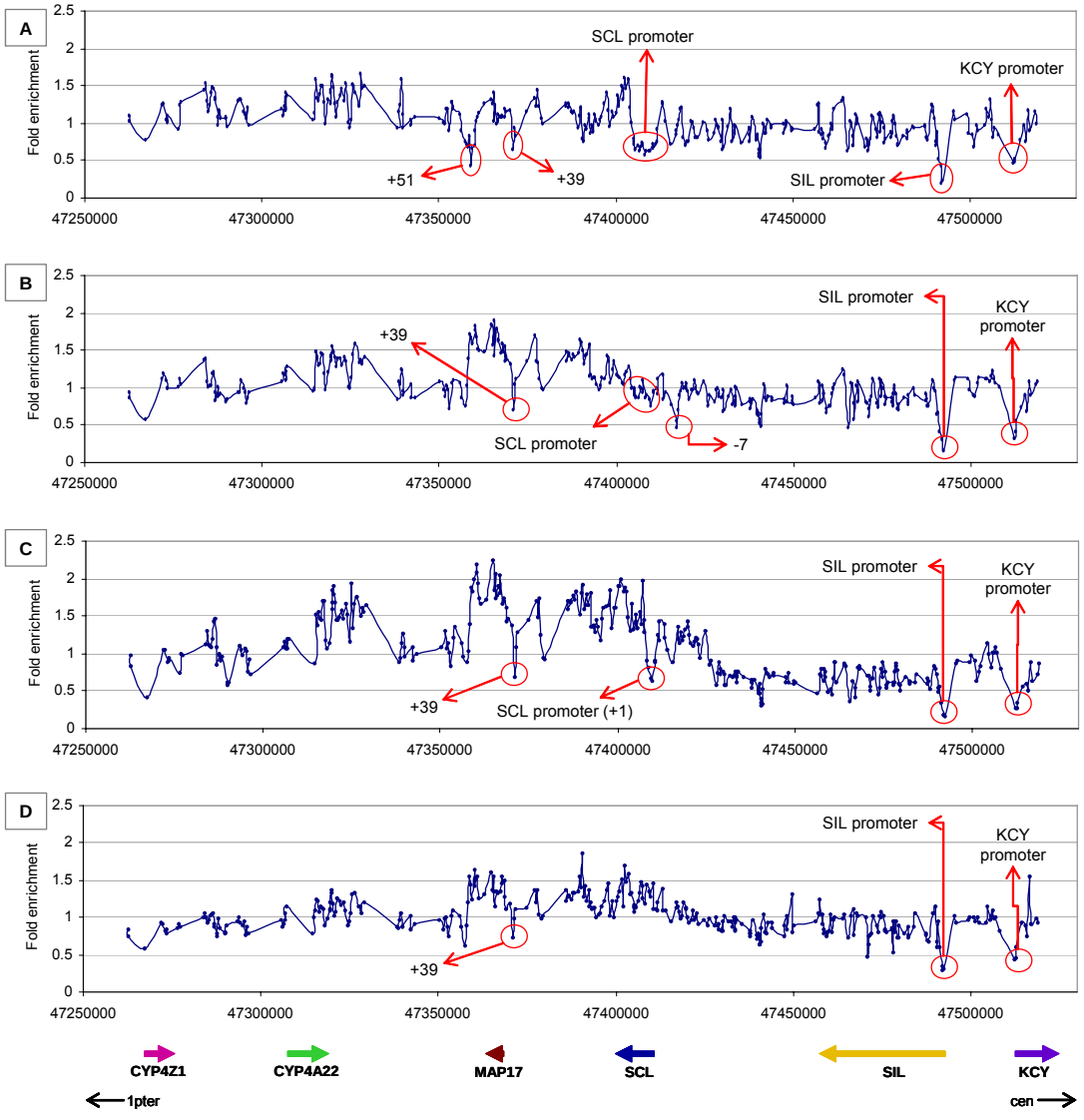
At the 5' end of the SCL gene in K562 and Jurkat, increased levels of enrichments for phosphorylation were observed at +1 and +3 regions located downstream of p^{1a} and p^{1b}, whereas such increased levels were not observed for HL60 and HPB-ALL. In addition, increased levels of enrichment were also observed at the downstream region of p^{EXON4} only in Jurkat and not in the other cell lines. Taken together, it could be interpreted that phosphorylation at serine 10 of H3 marked the downstream regions of active promoters in these cell lines. However, similar pattern of increased levels of enrichments were not observed at the 5' ends of SIL and KCY genes (Appendix 10) suggesting that phosphorylation of H3 S10 downstream of promoters may not be a widespread mechanism applicable to all genes.

5.11 Analysis of nucleosome density at the human SCL locus in Jurkat, HL60 and HPB-ALL cell lines

In chapter 4, the results obtained in ChIP-chip for histone H3 levels across the human SCL locus in K562 (see section 4.7.2, chapter 4) demonstrated that decreased levels of nucleosomes coincide with the regulatory and coding regions of active genes. Based on

these results, the levels of histone H3 were also analysed in the other three human cell lines to determine whether this relationship was present in other cell types. Figure 5.17 shows the ChIP-chip profiles across the SCL locus in K562, Jurkat, HL60 and HPB-ALL cell lines. Panel E of the figure represents the composite ChIP-chip profiles at the 5' end of SCL gene and the +51 erythroid enhancer in the four cell lines.

Decreased relative levels of histone H3 were detected at, and close to the promoter regions of the KCY and SIL genes in all the three cell lines (results for K562 have been discussed in chapter 4). At the 5' end of the SCL gene, decreased relative levels of histone H3 were observed in Jurkat at the -7 region, the promoter region p^{1b}, +1 and +3 regions and also at the +7 region encompassing p^{EXON4} (Figure 5.17, panel E). It has been shown that the -7 region exhibited strong DNase I sensitivity in Jurkat (Leroy-Viard et al. 1994) and that p^{EXON4} is active in Jurkat (Bernard et al. 1992).



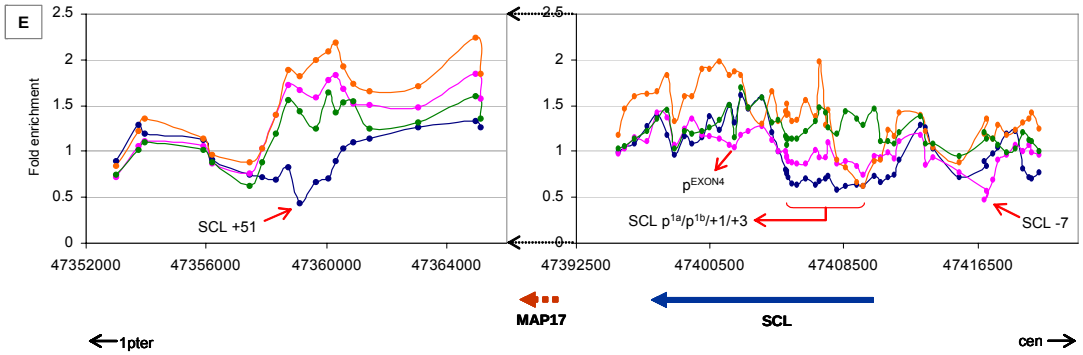


Figure 5.17: ChIP-chip profiles for histone H3 levels across the SCL locus in four human cell lines. Panels A: K562, panel B: Jurkat, panel C: HL60 and panel D: HPB-ALL. The genomic regions with decreased levels for histone H3 are marked with red circles. Panel E shows the ChIP-chip profiles of selective regions across the SCL locus; the panel shows a small genomic region at the 5' end of the SCL gene (right panel) and the genomic region across the erythroid enhancer at +51 (left panel). The four colours of the profiles in panel E represent the four cell lines: blue (K562), (pink) Jurkat, (orange) HL60, (green) HPB-ALL. The genomic regions with decreased levels of histone H3 in these regions are marked with red arrows. The x-axes represent genomic coordinates along the human chromosome 1 and the y-axes represent fold enrichment. The two plots of panel E are joined by dotted arrows in the middle to indicate the continuity between the two profiles. The thick, coloured arrows at the bottom of the panels D and E and the brown coloured dotted arrow in panel E represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown with black arrows at the bottom of panels D and E.

In HL60, decreased levels were detected only at the +1 region coincident with the region of H3 K9/14 diacetylation seen in this region (see section 5.6.1.2). In HPB-ALL, decreased relative levels were not detected at these regions. In addition, decreased relative levels for histone H3 were also detected at the +39 region in all cell lines which was also seen in K562 (Figure 5.17). In Jurkat, HL60 and HPB-ALL cell lines, decreased levels of histone H3 were not observed at the erythroid enhancer at +51 (Delabesse et al. 2005) which was previously seen in K562 (Figure 5.17, panel E). This observation correlates with the activity of this enhancer only in K562.

Interestingly, it was noted that there were noticeably elevated ratio on either side of the regulatory regions which reported a drop in histone H3 levels. Taken together, these results suggested that, as previously seen for K562, depletion of nucleosomes (reflected by decreased levels of histone H3) occurred at active regulatory regions across the SCL locus in Jurkat, HL60 and HPB-ALL cell lines. The results of these experiments are analyzed further in chapter 6.

5.12 SCL itself binds to the erythroid enhancer along with GATA-1 and Ldb-1 in K562: evidence for the involvement of the SCL multi-protein complex in SCL regulation.

As described in chapter 4, *in vivo* interactions of GATA-1 transcription factor across the SCL locus were identified at nine genomic regions, all of which had conserved and/or non-conserved GATA binding sites. In addition, one array element, at the +51 erythroid enhancer (homologous to mouse +40 region), which showed the highest ChIP enrichments for GATA-1, contained three highly conserved GATA sites and also binding sites for other transcription factors (chapter 4, section 4.6.4). One of these GATA sites was found to be contained within a 20 bp sequence containing a GATA/E-box composite site which showed the canonical hallmarks of the SCL erythroid DNA-binding complex which includes the SCL, E47, GATA-1, LMO2 and Ldb-1 transcription factors (Wadman et al. 1997). Based on this information, ChIP-chip experiments were performed in K562 using antibodies raised to all the proteins involved in the SCL multi-protein complex (Table 5.1). Although ChIP experiments were performed using antibodies raised to all of the above mentioned transcription factors, however, significant enrichments were obtained only with anti-SCL and anti-Ldb-1 antibodies.

Figure 5.18 shows the ChIP-chip profiles generated for GATA-1 (the results have already been discussed in chapter 4, section 4.6.2), SCL and Ldb-1 transcription factors across the SCL locus in K562 cell line. For both SCL and Ldb-1, significant and the highest enrichments were obtained at the +51 region and at the same array elements which were also enriched for GATA-1. These data suggested that at least three of the components of the SCL erythroid DNA binding complex (Wadman et al. 1997) bind to the +51 region. More importantly, this suggests that SCL, in a complex with other transcription factors, regulates the activity of its own expression.

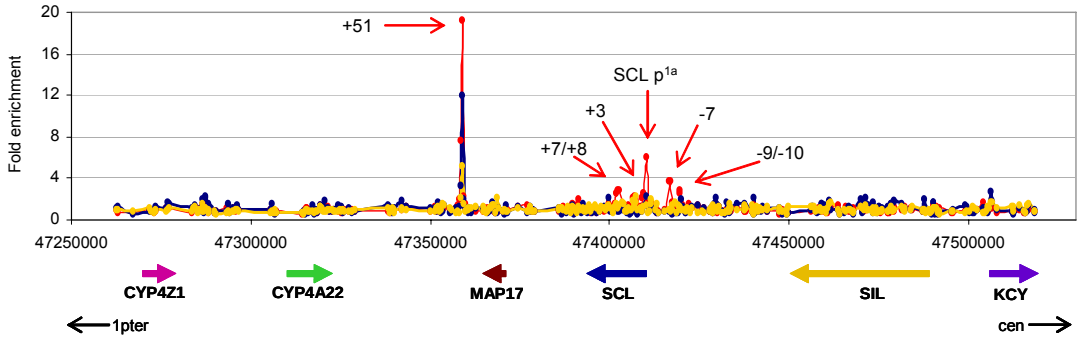


Figure 5.18: Composite ChIP-chip profiles for GATA-1, SCL and Ldb-1 across the SCL locus in K562. Enrichments for GATA-1, SCL and Ldb-1 are represented by red, yellow and blue profiles respectively. The genomic regions enriched for these transcription factors are marked by red arrows and discussed in the text. The x-axis represents the genomic coordinates along human chromosome 1 and the y-axis represents fold

enrichments. The thick, coloured arrows at the bottom of the figure represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown with black arrows at the bottom of the figure.

5.13 Mapping interactions of proteins with HAT and HDAC activities across the SCL locus in K562

Proteins that catalyse reversible acetylation of histones; histone acetyltransferases (HATs) and histone deacetylases (HDACs) exist as multi-protein complexes that have co-activator and co-repressor activities respectively. Some HATs are also able to acetylate non-histone transcription factors involved in haematopoiesis (Blobel et al. 2000; Huang et al. 1999; Huang et al. 2000). A number of ChIP-chip assays were tested to identify the binding of these proteins at the SCL locus (Table 5.1). Although, ChIP experiments using antibodies raised to these proteins were performed in all cell lines, many of these antibodies did not exhibit significant ChIP enrichments across the SCL locus. However, significant ChIP enrichments were obtained for some of these factors in K562 cell line and the results obtained with p300, CBP and HDAC2 are presented below (Figure 5.19). HDAC2 is a histone deacetylase whereas p300 and CBP possess intrinsic HAT activity. p300 and CBP are also known to be important cofactors for a number of transcription factors both within and outside the haematopoietic system (such as GATA-1 and SCL) (Blobel et al. 2000) and share high sequence homology which suggests that the antibodies raised to these proteins may be cross-reactive.

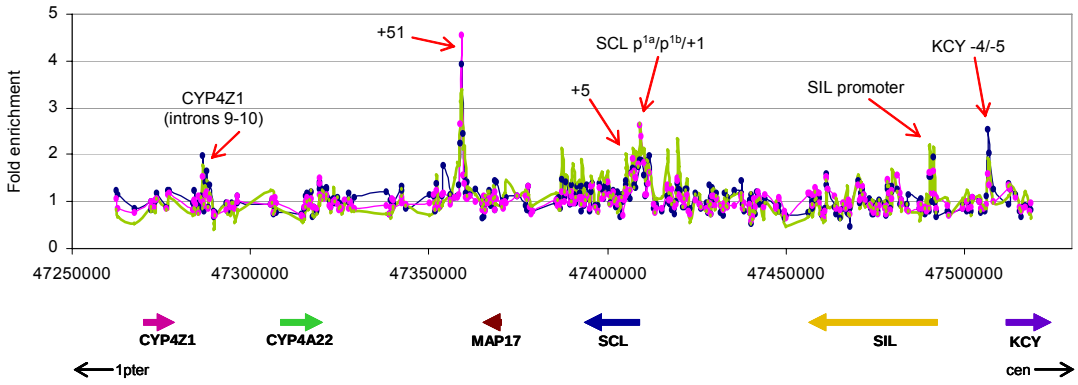


Figure 5.19: Composite ChIP-chip profiles for p300, CBP and HDAC2 across the SCL locus in K562. ChIP enrichments for p300, CBP and HDAC2 are represented by blue, pink and green profiles respectively. The genomic regions showing enrichments for these regulatory proteins are marked by red arrows. The x-axis represents genomic coordinates along human chromosome 1 and the y-axis represents fold enrichments. The thick, coloured arrows at the bottom of the figure represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown by the black arrows at the bottom of the figure.

Significant enrichments for all three factors were obtained at the +51 erythroid enhancer and at the 5' end of the SCL gene at -1, p^{1a}, p^{1b}, +1 and +3 regions. Additionally,

significant enrichments for all three factors were obtained at the 5' end of the SIL gene but not at the 5' end of the KCY gene. However, significant enrichments for all three factors were obtained at KCY -4/-5 region. The presence of factors with HAT and HDAC activity at the same regions is consistent with their role in maintaining a dynamic acetylation state (Kurdistani and Grunstein 2003).

For HDAC2, significant enrichments were also obtained at the +5 region which were not observed with p300 or CBP. Across the SCL gene, it was observed that the ChIP enrichments for HDAC2 (green profile in figure 5.19) showed several peaks at the downstream as well as upstream region of the 5' end of the gene and the enrichments obtained at these regions were above the significant threshold value. Most of these regions with significant enrichments mapped to known regulatory elements of SCL at -9/-10, -7, and +20/+21 (stem cell enhancer). A region with significant enrichments was also identified within the genomic region containing the CYP4Z1 gene (between introns 9-10) for all three factors but this region did not map to any region with known regulatory function.

5.14 Mapping interactions of factor involved in the pre-initiation complex across the SCL locus in K562

General transcription factor TFIID and RNA pol II are both involved in the assembly of the pre-initiation complex at the promoters (Burley and Roeder 1996) and, of all the components, TFIID is the only component that is capable of binding specifically to the core promoters (also see chapter 1). $Taf_{II}250$ which is a subunit of TFIID is an important cofactor also shows HAT activity *in vitro* (Mizzen et al. 1996). ChIP-chip experiments were performed in human cell lines to map the interactions of RNA pol II and $Taf_{II}250$ at the active promoters. However, no significant ChIP enrichments were obtained for RNA pol II in any cell line tested. For $Taf_{II}250$, significant enrichments were observed only in the K562 cell line.

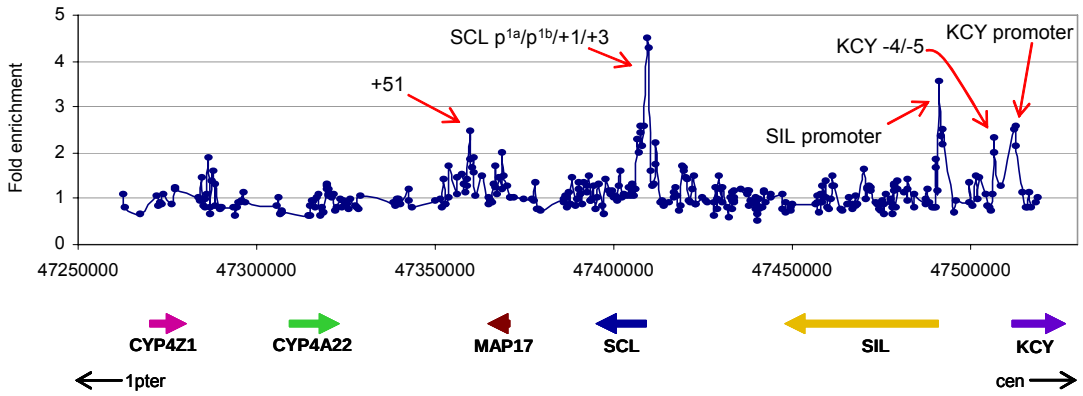


Figure 5.20: ChIP-chip profile for $Taf_{II}250$ across the SCL locus in K562. The genomic regions with significant enrichments are marked with red arrows. The x-axis represents genomic coordinates along human chromosome 1 and the y-axis represents fold enrichments. The thick, coloured arrows at the bottom of the figure represent gene the order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown by the black arrows at the bottom of the figure.

In K562, significant enrichments for $Taf_{II}250$ were obtained at or near the 5' ends, i.e. promoters, of KCY, SIL and SCL genes (Figure 5.20). The enrichments at the 5' end of SCL encompassed p^{1a}, p^{1b}, +1 and +3 regions. In addition, significant enrichments were obtained at the +51 erythroid enhancer. However, the highest enrichments were not observed on the array element which showed highest enrichments for GATA-1, SCL and Ldb-1 but on the adjacent array element (500 bp proximal to GATA-binding towards the SCL gene). At the 5' end of the SIL gene, significant enrichments for $Taf_{II}250$ were observed at two distinct regions, one at the promoter and another one approximately 2 kb downstream of the SIL promoter (SIL +2). In the ENSEMBL data base (http://www.ensembl.org), three transcripts for the SIL gene have been annotated and the 5' end of one of these transcripts maps to a region approximately 2 kb downstream of the known SIL promoter, raising the possibility of an alternate promoter for the SIL gene (see also chapter 7). In addition, significant enrichments were also obtained at the KCY -4/-5 region which had previously been identified with various histone modifications but no regulatory function for this region is known.

5.15 Elf-1 binds to the SIL -1 region in K562, Jurkat and HPB-ALL

Elf-1 is mainly a lymphoid specific Ets transcription factor (Wang et al. 1993) involved in the regulation of the SCL stem cell enhancer (+20/+21 region) (Gottgens et al. 2002) and the endothelial enhancer (-3/-4 region) (Gottgens et al. 2004). It was therefore tested in ChIP-chip assays in all cell lines. Figure 5.21 shows a composite ChIP-chip profile for Elf-1 in K562, Jurkat and HPB-ALL.

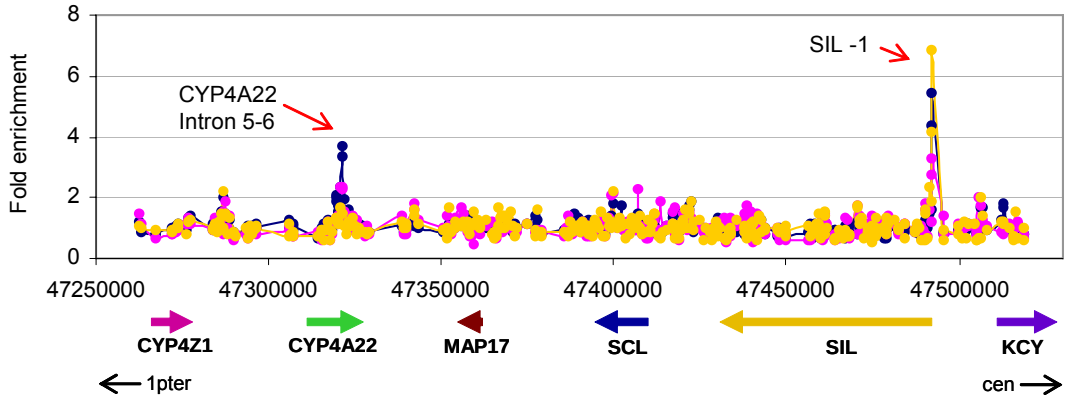


Figure 5.21: Composite ChIP-chip profile for Elf-1 across the SCL locus in K562, Jurkat and HPB-ALL. ChIP enrichments for Elf-1 in K562, Jurkat and HPB-ALL are represented by blue, pink and yellow profiles respectively. The genomic regions showing enrichments are marked with red arrows. The x-axis represents genomic coordinates along human chromosome 1 and the y-axis represents fold enrichments. The thick, coloured arrows at the bottom of the figure represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown by the black arrows at the bottom of the figure.

Significant enrichments for Elf-1 were obtained at the 5' end of the SIL gene with the highest enrichments at the SIL -1 region in all three cell lines, and additional low level enrichments were observed at SIL +1 region only in HPB-ALL. In addition, significant enrichments were also seen in a region mapping to the CYP4A22 gene (introns 5-6) in K562 and Jurkat, but not in HPB-ALL.

5.16 CTCF binding at the SCL locus in K562 and Jurkat

The MAP17 gene is located between the SCL erythroid enhancer at +51 and the SCL gene itself. Furthermore, the +51 region is closer to the CYP4A22 gene than it is to SCL. Thus, in order to ensure appropriate expression of the SCL gene by this distal enhancer, a mechanism must exist to ensure that the erythroid enhancer interacts with the SCL promoter and does not inappropriately enhance the expression of MAP17 or CYP4A22. Insulators are regulatory sequences that are known to possess dual roles: they form boundaries between regions of actively and inactively expressed genes, as well as modulating interactions between enhancers and promoters. The zinc-finger DNA-binding protein, CTCF, is known to bind to a number of mammalian insulators characterized to date (Bell et al. 1999). In addition to its enhancer blocking function (Bell et al. 1999), CTCF also plays roles in transcriptional regulation (Klenova et al. 1993), and gene imprinting (Bell and Felsenfeld 2000). Based on all of this information, ChIP-chip experiments were performed in two SCL expressing cell lines, K562 and Jurkat, to determine whether CTCF binding was involved in regulating the activities of the SCL erythroid enhancer.

In K562, significant and highest enrichments for CTCF were obtained at the +57 region (Figure 5.22, panel A). In addition, the +39 region and p^{MAP17} also showed significant enrichments for CTCF. A genomic region in the CYP4A22 gene (introns 5-6) was also significantly enriched. In Jurkat, CTCF interactions were identified only at the +57 region (Figure 5.22, panel B). In addition, the array elements exhibiting highest enrichments for CTCF in K562 and Jurkat were different. Another genomic region at the CYP4Z1 gene (introns 8-9) was also found to be significantly enriched in Jurkat, the function for which is not known. The +57 region had been identified previously with significant enrichments for di- and trimethylation at lysine 4 of histone H3 and decreased levels of H3 K16

acetylation, in K562 and Jurkat. Additionally, the +39 region was also found to be enriched for various histone modifications in both the cell lines. The +39 region also displayed nucleosome depletion. Given the facts that (i) the genomic region downstream of +57 contains the CYP4A22 and CYP4Z1 genes (both of which are silent), and (ii) the MAP17 promoter and +39 region is bound with CTCF, it is likely that CTCF binding may contribute to the regulation of the various regulatory sequences in this region as well as serve to demarcate an insulator element at +57.

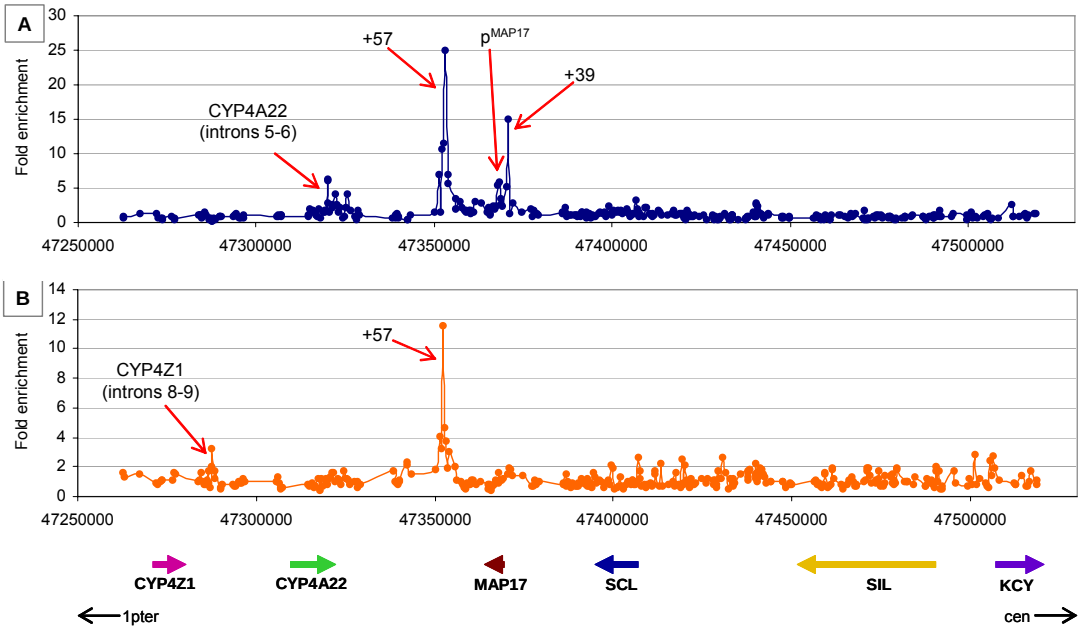


Figure 5.22: ChIP-chip profile for CTCF binding across the SCL locus in K562 and Jurkat. Panel A: K562, panel B: Jurkat. The genomic regions showing significant enrichments for CTCF are marked with red arrows. The x-axes represent the genomic coordinates along human chromosome 1 and the y-axes represent fold enrichments. The thick, coloured arrows at the bottom of panel B represent the gene order and the direction of transcription. The orientation of the locus with respect to the centromere (cen) and telomere (ter) on human chromosome 1 is shown at the bottom of the figure.

5.17 Discussion

5.17.1 Patterns of histone acetylation across the human and mouse SCL locus

The histone H3 and H4 acetylation patterns (using assays for H3 K9/14 diacetylation and H4 K5/8/12/16 tetra-acetylation) that were obtained for the four human haematopoietic cell lines and the mouse 416B line were in complete agreement with those described in a PCR-based analysis of a smaller genomic region across the SCL locus (Delabesse et al. 2005). Furthermore, the distribution of histone H3 and H4 acetylation were consistent with the location of known regulatory elements for SCL and its flanking genes in all of the respective cell lines. It has been reported that the acetylation patterns for histone H3 and H4 are very similar (Roh et al. 2004; Litt et al. 2001) and correlated with transcriptional

activity of a gene; the data obtained across the SCL locus were in agreement with these previous observations.

Hyperacetylation of all specific lysine residues of histone H3 (H3 K9, H3 K14, H3 K18 and H3 K27) was highly correlated with the transcriptional status of a gene which is consistent with similar observations in yeast (Kurdistani et al. 2004; Liu et al. 2005). These modifications mostly followed the pattern of H3 K9/14 diacetylation and marked virtually all the known and active regulatory regions in the human cell lines. In contrast, correlation of acetylation of specific lysine residues of H4 (H4 K5, H4 K8, H4 K12 and H4 K16) with the transcriptional status of the gene was variable. The data obtained in this study suggested that hyperacetylation at lysine 5 of H4, and hypoacetylation at lysines 8 and 16 of H4 were correlated with the transcriptional activity of a gene. Similar results have been observed in yeast with the existence of a hypoacetylated domain for these modifications (H4 K8 and H4 K16) adjacent to the transcription start sites (Liu et al. 2005). It has also been shown by mutating yeast H4 lysines 5, 8, 12 and 16 that only lysine 16 mutations had specific transcriptional consequences independent of the mutational state of the other lysines. This suggests that out of all the lysine modifications of histone H4 N-terminal tail, only the lysine 16 modification is functionally non-redundant (Dion et al. 2005).

Results obtained in the present study for tetra-acetylation of H2A at lysines 5/9/13/15 and H2B at lysines 5/12/15/20 in K562 were in agreement with the observations at the chicken β-globin locus (Myers et al. 2003), which suggested that H2B specifically marks tissue-specific genes. H2A tetra-acetylation pattern was similar to that of histone H3 K9/14 diacetylation for all genes in K562, whereas H2B tetra-acetylation specifically marked the SCL gene and not the SIL gene which is a ubiquitously expressed gene. However, with H2B tetra-acetylation, low level but significant enrichments were observed at the 5' end of KCY gene which is also a ubiquitously expressed gene. This suggests that H2B tetra-acetylation may also have a role in the regulation of housekeeping genes.

5.17.2 Patterns of histone H3 methylation across the human and mouse SCL locus

Patterns of histone H3 K4 methylation at the 5' ends of genes also correlated with the transcriptional activity of genes in the human and mouse cell lines studied here. Di- and trimethylation enrichment levels of H3 K4 were highest at the 5' ends of active genes, i.e., at or near the active promoters. In addition, enrichments for trimethylation were highest over the promoters, but for dimethylation, the enrichments were highest just downstream of the promoter region. Conversely, at active promoters, monomethylation enrichment levels were very low, but rose on either side (see below). These observations

are consistent with the idea that mono-methyl lysine 4 residues are converted to di- and then trimethyl residues to accompany gene activation at promoter regions.

Similar associations of trimethylation of H3 K4 with the promoter regions of active genes have been well documented (Santos-Rosa et al. 2002; Schneider et al. 2004; Liu et al. 2005; Schübeler et al. 2004; Pokholok et al. 2005). However, the presence of dimethylation has not been linked clearly to transcriptional activity. It has been reported that dimethylation of histone H3 K4 is enriched at active and inactive regions (Santos-Rosa et al. 2002; Schneider et al. 2004) with a preferential enrichment over the coding regions of active genes (Bernstein et al. 2002; Liu et al. 2005; Pokholok et al. 2005). The data obtained in this study demonstrated that both di- and trimethylation were enriched over similar regions, with higher levels of enrichment at the 5' ends and with both modifications extending into the coding region to the same extent.

While lack of enrichments for histone H3 K4 monomethylation was found to be associated with the active SCL promoters in K562, Jurkat and 416B (see above section), increased levels of enrichments were found at the SCL promoters in the HL60 and mouse ES E14 cell lines – neither of which express SCL. It has been suggested that when genes are repressed, dimethylation is present but trimethylation is absent and when genes become active, trimethylation appears but dimethylation persists (Santos-Rosa et al. 2002). Thus, relative increase of trimethylation versus dimethylation discriminates the active from the inactive genes (Schneider et al. 2004). However, based on the data obtained in the present study, this suggestion can be extended further by including the relative levels of enrichments for monomethylation which means that instead of di- and tri-, it is possibly the relative enrichments for monomethylation which discriminates active from inactive promoters (see chapter 6).

Increased enrichment levels of monomethylation were found to be associated with other types of regulatory regions including active enhancers. Furthermore, increased levels of enrichments for monomethylation have been associated with the 3' ends of the active genes in yeast (Liu et al. 2005; Pokholok et al. 2005), although this was not apparent across the SCL locus. Overall, the ChIP-chip data obtained for histone H3 lysine 4 methylation in the present study suggests that the distribution of these modifications at the SCL locus are more complex than that previously reported in yeast. Thus, extrapolations of H3 K4 methylation data from yeast to human may not be accurate, given the increased complexity of regulatory interactions found in mammalian genomes.

The data obtained with trimethylation of H3 K79 suggested that this histone mark is linked to transcriptional activity which was not observed in previous studies (Pokholok et

al. 2005). However, the presence of this histone mark at the downstream regions of active promoters suggested that this modification of H3 may be linked to transcriptional elongation. It has been suggested that the Paf1 complex which is involved in transcriptional elongation is also required for methylation at lysine 79 of H3 (Krogan et al. 2003). Taken together, these data provide support for trimethylation of H3 K79 to possibly play a role in transcriptional elongation in mammalian cells.

5.17.3 Histone H3 acetylation and methylation modifications can also be associated with inactive genes

Surprisingly, histone modifications associated with gene activation were found at the inactive SCL locus in HL60 and E14 ES cells. In HL60, subtle, but significant, ChIP enrichments for H3 K9/14 diacetylation were found at the 5' end of the SCL gene. Histone H3 K9/14 diacetylation patterns in E14 ES cells were found to have similarities with those of HL60, although in E14, almost all of the SCL regulatory elements were acetylated albeit at low levels. In addition to the acetylation marks, the SCL promoter also showed significant enrichments for mono-, di- and trimethylation at lysine 4 of histone H3 in HL60 and in E14 ES cells. These data are not consistent with the established relationships between H3 acetylation, H3 K4 methylation and gene activity. However, the relationship between histone H3 K4 methylation and gene activity will be further explored in chapter 6.

The presence of "activating" histone modifications (H3 acetylation and H3 K4 di- and trimethylation), at the SCL gene when it is not expressed, suggest two interesting hypotheses which both have supporting evidence from other studies found in the literature as described below.

HL60 is a promyelocytic cell line which lies downstream of the granulo/myelopoietic progenitors (GMP) in the hierarchy of blood development and SCL is known to be expressed in GMP cells. It is possible that HL60 cells may still be carrying histone acetylation and methylation marks although the cells have differentiated to monocyte precursors and SCL expression has been turned off. This suggests that the silencing of the SCL is not completely dependent on its modification status and other factors are involved in turning its expression off. Similar low levels of acetylation have also been reported over regulatory elements and hypersensitive sites across the β-globin locus in a cell line where the gene was not expressed (Litt et al. 2001).

These results also suggest that regulatory regions in mouse ES cells may already be primed to specify the activation of SCL gene when pluripotent cells become blood progenitors. Similarly, histone H3 acetylation and histone H3 lysine 4 methylation at a *cis*-acting element at the mouse λ5-VpreB1 locus (involved in B-cell development) has

been shown in mouse E14 ES cells prior to the gene becoming active (Szutorisz et al. 2005); this element was then actively involved in the recruitment of regulatory proteins during the differentiation of ES cells through to the pro-B cell stage. Therefore, it seems likely that the marking of regulatory regions for histone marks may be fundamental to the recruitment of transcription factors and the assembly of pre-initiation complex at the SCL locus when pluripotent cells differentiate into haematopoietic progenitors and SCL is turned on. It is interesting to note that the neural enhancer at the +23 region (Gottgens et al. 2000) was also detected with H3 K9/14 diacetylation in ES cells which implies that not only the haematopoietic regulatory elements may be primed, but all elements which may participate in the regulation of SCL in distinct compartments of its normal expression are primed for subsequent expression of SCL. Furthermore, in E14 cells, ChIP-chip assays for H3 K4 methylation identified a number of novel regions across the SCL locus, many of which were located within the SIL gene and were specifically enriched for monomethylation of H3 K4. Based on this observation, it could be suggested that a number of unknown regions were also marked in ES cells and which may represent novel putative regulatory regions which become active at later stages of development. Although the function of these regions is not clear, it is possible that these regions are associated with the regulation of SCL or SIL.

5.17.4 Nucleosome depletion occurs across the SCL locus in all human cell lines

The experiments performed to determine the levels of histone H3 across the SCL locus in SCL expressing and non-expressing human cell lines provide additional evidence that nucleosome depletion occurs at active genes at the SCL locus. A description of results obtained from further analytical approaches which support this evidence is provided in chapter 6.

By observing H3 histone profiles in SCL-expressing and non-expressing cell lines, differences in nucleosome levels could be correlated to the activities of various SCL regulatory regions. For example, only in K562 were depletions at the +51 erythroid enhancer observed. Thus nucleosome differences across the SCL locus between the cell lines studied here are likely to be due to real differences in gene activity rather than just general features of the different cell lines (and not related gene expression). To verify that nucleosome dynamics are responsive to gene activity, it would be necessary to observe the changes in nucleosome density which accompanies the repression or activation of genes in an appropriate cell system. This has been demonstrated in yeast (Lee et al. 2004), but not in human systems to date. However, it is still not clear how this depletion of nucleosomes occurs to make the underlying chromatin available for binding

of the RNA polymerase complex, transcription factors and other regulatory proteins. Mechanisms including sliding or complete eviction of nucleosomes have been suggested (Hamiche et al. 1999; Lorch et al. 1999; Korber et al. 2004). The results in K562 support the idea of nucleosome sliding (see chapter 4 discussion) and the analysis of the three other human cell lines described in this chapter supports this idea.

5.17.5 SCL may be self-regulated in a positive feedback loop

Significant levels of GATA-1, SCL and Ldb-1 enrichments at the +51 erythroid enhancer provided the first evidence that these three transcription factors are involved in the activity of this enhancer during erythroid development. In HSCs, activity of the SCL stem cell enhancer (+20/+21 in human) is independent of SCL expression (Sanchez et al. 1999). The data described here is therefore consistent with a model where once SCL expression has been established in HSCs, maintenance of SCL during differentiation towards erythroid cells is at least partly achieved through a positive feedback loop acting through the +51 erythroid enhancer. However, given that these interactions were shown in the cultured cell line K562, support for this model and the functional relationship between GATA-1, SCL and Ldb-1 at the erythroid enhancer would need to be confirmed in primary primitive erythroblasts, where SCL is under the control of the erythroid enhancer (Delabesse et al. 2005). The presence of the highly conserved binding site for the composite GATA/E-box of the erythroid SCL complex (Wadman et al. 1997) suggests that GATA-1, SCL and Ldb-1 may act together at this site in a multi-protein complex. To further support this evidence, it will also be important to confirm that the transcription factors E2A and LMO-2 bind to this region, once appropriate antibodies are identified which work well in ChIP.

The ChIP-chip enrichments obtained for p300/CBP and HDAC2 at the +51 region and at the SCL promoter region, in addition to the binding events described for SCL, Ldb-1 (this chapter) and GATA-1 (Chapter 4), provide evidence that the regulatory events at these regions may be very complex. GATA-1 interacts with CBP *in vivo* and *in vitro* (Weiss et al. 1997) and SCL has been found to interact with p300 (Huang et al. 1999) and HDAC1 (Huang et al. 2000). Therefore, the presence of p300/CBP and HDAC2 at sites where SCL and GATA-1 are also present is consistent with these previous observations.

5.17.6 Binding interactions of other transcription factors across the SCL locus

5.17.6.1 Elf-1 and the regulation of the SIL gene

Elf-1 is mainly a lymphoid-specific transcription factor but is also known to regulate SCL expression through its stem cell and endothelial enhancers (Gottgens et al. 2002;

Gottgens et al. 2004). ChIP-chip enrichments for Elf-1 one kilobase upstream of the SIL promoter (SIL -1 region) in both lymphoid (Jurkat and HPB-ALL) and myeloid (K562) cell types were unexpected given that the SIL gene is ubiquitously expressed. This suggests that Elf-1 may exert a haematopoietic-specific control on SIL expression, although it cannot be excluded that this interaction is found in cultured cells but not in primary cells *in vivo*. Given that there may be a number of putative novel regulatory regions located in the SIL gene which could control either SIL or SCL expression (as evidenced by regions showing various histone modifications), Elf-1 may be required to modulate SIL expression to ensure appropriate activity of these regulatory regions. The interactions of Elf-1 at the SIL -1 region needs to be further studied in haematopoietic cell types to determine whether this interaction is biologically relevant and whether these putative novel regulatory regions control SIL or SCL expression.

5.17.6.2 CTCF and its role in insulator function at the SCL locus

It was discussed in section 5.16 that based on the location of the +51 erythroid enhancer, the appropriate regulation of SCL, MAP17 and the cytochrome P450 gene cluster may require the involvement of insulator elements. CTCF is known to bind at most known mammalian insulators (Bell et al. 1999) and ChIP-chip enrichments for this transcription factor at the +57 region in both K562 and Jurkat cell lines suggest that this region may represent an insulator element at the SCL locus. It has been suggested that insulator elements possess roles as both enhancer-blockers and as barriers between chromatin domains (West et al. 2002). Whether CTCF performs both functions at the +57 region is not known. Support for its enhancer-blocking function at the SCL locus is evidenced by the fact it also binds to the MAP17 promoter in K562. These interactions indicate that the activity of the MAP17 promoter may be modulated through looping interactions with the +57 region. Alternatively, CTCF may bind directly to the promoter to control MAP17 expression. The binding of CTCF upstream from the MAP17 promoter at the +39 region in K562 suggests that +39 itself could be an insulator which partitions the MAP17 gene from the +51 SCL enhancer, on one side, and the SCL gene and its more proximal regulatory sequences, on the other side. Interestingly, CTCF interactions in Jurkat were not seen at the MAP17 promoter or at +39 - although both K562 and Jurkat express SCL and MAP17 (Delabesse et al. 2005). However, given that SCL expression in Jurkat is inappropriate, the regulation of SCL and MAP17 may be different in the two cell lines.

5.17.7 Issues in interpreting ChIP-chip data

The work described in this chapter illustrates how ChIP-chip can be used to characterize regulatory features at the SCL locus. Although ChIP-chip provides a powerful and high-

throughput method of mapping DNA-protein interactions in this way, it is worth noting at this point, some issues in interpreting the data derived from these experiments. These issues are relevant even if ChIP-quality antibodies are used in ChIP-chip assays which are specific for detecting the appropriate interactions.

i) **Heterogeneity in cell populations:** A ChIP-chip assay provides only a snapshot of *in vivo* DNA-protein interactions within a cell population at a particular point in time or under specific conditions. In other words, ChIP-chip profiles are a composite survey of an entire cell population (which are often growing in culture asynchronously) and the regulatory features observed could be occurring in different cells at any given time. This could also mean that binding events occurring in a small proportion of cells may go unnoticed because they are being masked by events in the rest of the cell population. Thus, multiple interactions occurring at the same sequence, as observed in ChIP-chip assays, does not necessarily mean that these proteins are acting concertedly in a complex within the same cells at the same time.

ii) **Resolution:** Taking this last point further, even when multiple interactions are occurring in the same cells at the same time, it can be difficult to determine whether they bind to the same regulatory element. This is a question of array resolution, and by developing an array of sufficiently high resolution, the location of binding sites can be pinpointed more accurately. Yet, even with a high resolution array as that described in this thesis, proteins can bind in close proximity to a given sequence without the factors existing in the same complex (Metivier et al. 2003).

iii) **Allelic variations in expression:** Similarly, binding interactions at specific gene alleles cannot be inferred from ChIP-chip profiles since the composite profiles report the interactions occurring at all alleles within the genome. For example, in the present study, both K562 and Jurkat showed multiple copies of the SCL locus. Although it is known that SCL expression is mono-allelic in Jurkat (Leroy-Viard et al. 1994), even the allele which does not express SCL may show regulatory events which differ from that at the expressing locus. However, all of these interactions will appear in the composite profile for all of the alleles in a given cell.

iv) **Indirect DNA-protein interactions:** It is possible that enrichments obtained in ChIP-chip are due to regulatory proteins (such as transcription factors) bound to one genomic sequence (for example an enhancer) interacting to a second sequence (for example a promoter) via chromatin looping. This is due to the fact that cross-linking of DNA-protein interactions can physically bind proteins of interest to other sequences with which they interact. These secondary sequences obtained through

such indirect DNA-protein interactions may even lack *bona fide* binding sites for the proteins of interest. This is often suggested as a reason why not all regions which show enrichments in ChIP-chip assays show the requisite binding site for the transcription factors of interest.

v) **Biological and Non-biological Noise:** In this chapter and in chapter 4, it was shown that nucleosome density is variable across the SCL locus, and that this information could be used to help interpret histone acetylation levels (section 4.7.3). The degree to which noise affects ChIP-chip results and the best strategies for data normalization to account for this is a controversial topic and one on which few laboratories agree (Lee et al. 2004; Bernstein et al 2004; Pokholok et al. 2005).

vi) **Antibody efficiencies:** Although antibodies may perform well in ChIP-chip, it is difficult to compare levels of enrichments across assays as a measure of the true levels of histone modifications, or binding stoichiometries of regulatory proteins. This is because antibodies may have variable efficiencies in different assays for enriching for the relevant genomic sequences. It is possible, however, to relate enrichments within an assay to aid in deducing relative binding efficiencies of proteins and relative levels of histone modifications.

5.17.8 Conclusions

The work presented in chapter 5 represents the first ever detailed elucidation of DNA-protein interactions at the SCL locus and one of the most extensive characterization of regulatory interactions of any locus in a mammalian genome. The data obtained from these ChIP-chip experiments provides a comprehensive dataset from which to further explore the relationships of histone modifications with each other and with the underlying DNA sequences. Furthermore, the identification of a number of novel putative regulatory regions and binding interactions of transcription factors at the SCL locus provide opportunities to further characterize the SCL locus to understand its regulation. Chapters 6 and 7 of this thesis show how information from these ChIP-chip profiles is used in these ways.