Chapter 2

Materials and Methods

Materials

2.1. Composition of solutions

Sterile HPLC-grade water (BDH) was used to prepare all solutions.

10 X PCR buffer

- 500 mM KCl
- 50 mM Tris pH 8.5
- 25 mM $MgCl_2$

10 mM dNTP mix for PCR

- 10 mM each dNTP (dCTP, dGTP, dATP, dTTP)

10 X dNTP mix used in DNA labeling

The following mix was used in the labeling reactions that were used with microarray hybridizations set-up using a Tecan HS 4800 hybridization station

- 1 mM dCTP
- 2 mM each of dGTP, dTTP and dATP

20 X SSC

The following were dissolved in 800 ml water:

- 175.3 g NaCl
- 88.2 g Sodium citrate

The volume was adjusted to 1000 ml and the pH adjusted to 7.0

Tecan HS 4800 hybridization station buffer

- 50% formamide (Fluka)

- 5% dextran sulphate
- 0.1% Tween 20 (BDH)
- 2 X SSC
- 10 mM Tris pH 7.4

PBS/0.05% Tween 20 (Hyb wash solution 1)

PBS/0.05% Tween 20 for washing the arrays was prepared by dissolving the following salts in one litre of HPLC water

- 7.33 g NaCl
- 2.36 g Na_2HPO_4
- 1.52 g $NaH_2PO_4H_2O$
- 500 µl Tween 20

Solutions for ChIP:

Cell lysis buffer (CLB)

- 10 mM Tris-HCl pH 8.0
- 10 mM NaCl
- 0.2% NP-40 (Igepal)
- 10mM Sodium butyrate
- 50 µg/ml PMSF
- 1 µg/ml leupeptin

Nuclear lysis buffer (NLB)

- 50 mM Tris-HCl pH 8.1
- 10 mM EDTA
- 1% SDS
- 10 mM sodium butyrate
- 50 µg/ml PMSF

- 1 μg/ml leupeptin

IP dilution buffer (IPDB)

- 20 mM Tris-HCl pH 8.1
- 150 mM NaCl
- 2 mM EDTA
- 1% Triton X-100
- 0.01% SDS
- 10 mM sodium butyrate
- 50 μg/ml PMSF
- 1 μg/ml leupeptin

IP wash buffer 1 (IPWB1)

- 20 mM Tris-HCl pH 8.1
- 50 mM NaCl
- 2 mM EDTA
- 1% Triton X-100
- 0.1% SDS

IP wash buffer 2 (IPWB2)

- 10 mM Tris-HCl pH 8.1
- 250 mM LiCl
- 1 mM EDTA
- 1% NP-40
- 1% deoxycholic acid

IP elution buffer (IPEB)

- 100 mM $NaHCO_3$
- 1% SDS

TE (pH 8.0)

- 10 mM Tris base (pH 8.0)
- 1 mM EDTA

1 X PBS

1 X PBS used for washing the cells in ChIP assay was prepared by dissolving the following salts in 1 litre of HPLC water and the pH was adjusted to 7.4

- 8 g NaCl
- 0.2 g KCl
- 1.44 g Na_2PO_4
- 0.24 g KH_2PO_4

Solutions for western blotting procedure:

Nuclear protein extraction buffer

- 20mM HEPES pH7.9
- 0.2M EDTA
- 25% Glycerol
- 1.5mM $MgCl_2$
- 0.42M NaCl

4 x NuPAGE® LDS sample buffer

5 ml of 4 x NuPAGE® LDS sample buffer was prepared by dissolving the following in 4 ml of water:

0.003 g EDTA

2 g glycerol

0.4 g LDS

0.33 g Tris HCl

0.34 g Tris base

0.375 ml 1% SERVA Blue G250 solution

0.125 ml 1% phenol red solution

The volume was adjusted to 5 ml with water and stored at 4°C.

20 X NuPAGE® MOPS SDS running buffer

A 20X stock solution was prepared by dissolving the following in 800 ml of water:

- 209.2 g MOPS
- 121.2 g Tris base
- 20 g SDS
- 6.0 g EDTA

The volume was adjusted to 1000 ml with water and stored at 4°C. This buffer was diluted to 1 X with water for electrophoresis.

20 X NuPAGE® transfer buffer

250 ml of 20 X transfer buffer was prepared by dissolving the following in 200 ml water:

- 20.4 g bicine
- 26.2 g Bis-Tris
- 1.5 g EDTA

The volume was adjusted to 250 ml with water and stored at 4°C. This buffer was diluted to 1 X with water for western transfer.

1 litre of 1 X NuPAGE transfer buffer was prepared as follows and stored at 4°C:

- 50 ml 20 X NuPAGE transfer buffer
- 850 ml HPLC water
- 100 ml methanol

Blocking buffer

The following were dissolved in 90 ml water and stored at 4°C.

- 10 ml 10 X TBS
- 5 g non-fat dry milk
- 100 µl Tween-20

10 X TBS buffer

The following were dissolved in 1 litre of HPLC water, the pH was adjusted to 7.6 and the solution was stored at 4°C.

- 24.4 g Tris base
- 80 g NaCl

1 X TBST buffer

The following were dissolved in 900 ml of HPLC water:

- 100 ml 10 X TBS
- 1 ml Tween-20

2.2. Reagents

Antibodies

Factor	Supplier	Catalogue number
H2B unmodified	Abcam	ab1790
H3 unmodified	Abcam	ab1791
H3ac	Upstate Biotechnoloy	06-599
H3K4me1	Abcam	ab8895
H3K4me2	Abcam	ab7766
H3K4me3	Abcam	ab8580
H3K9ac	Upstate Biotechnoloy	07-352
H3K9me1	Abcam	ab9045
H3K9me2	Upstate Biotechnoloy	07-212
H3K9me3	Upstate Biotechnoloy	07-523
H3K9ac	Upstate Biotechnoloy	07-352
H3K18ac	Upstate Biotechnoloy	07-354
H3K27ac	Upstate Biotechnoloy	07-360
H3K27me1	Upstate Biotechnoloy	07-448
H3K27me2	Abcam	ab24684
H3K27me3	Upstate Biotechnoloy	07-449
H3K36me1	Abcam	ab9048
H3K36me2	Upstate Biotechnoloy	07-274
H3K36me3	Abcam	ab9050
H3K79me1	Abcam	ab2886
H3K79me2	Abcam	ab3594
H3K79me3	Abcam	ab2621
H4K5ac	Abcam	ab1758
H4K8ac	Abcam	ab1760
H4K16ac	Abcam	ab1762
CTCF	Santa Cruz Biotechnoloy	sc15914x
mSin3a	Abcam	ab3479
USF1	Santa Cruz Biotechnoloy	sc229x
USF2	Santa Cruz Biotechnoloy	sc862x

Enzymes

- Proteinase K, ≥ 20 units/mg (GibcoBRL)
- RNase A, ≥ 50 Kunitz units/mg (ICN Biochemicals)
- Taq polymerase, 5 units/μl (Perkin Elmer-Cetus)
- Klenow fragment, 40 units/μl (Invitrogen)

Fluorophores

- Cy3-dCTP (GE Healthcare)
- Cy5-dCTP (GE Healthcare)

Primer pairs

- Sequences of primer pairs used to construct the SCL tiling path microarray are available in the PhD thesis of Dr. Pawandeep Dhami (University of Cambridge, 2005). Primer pairs used to amplify PCR products for the ENCODE array are available at ftp://ftp.sanger.ac.uk/pub/encode/microarrays/
- Sequences of primer pairs used for quantitative SYBR green PCR are available in Appendix 1.

Other reagents

- Human C_ot 1 DNA (Invitrogen)
- Herring sperm DNA (Sigma)
- Phenol solution (Sigma)
- Chloroform solution (BDH)

2.3. Cells and cell lines

The human cell lines K562 (Lozzio and Lozzio, 1979) and U937 (Sundstrom and Nilsson, 1976) were obtained from Dr. Pawandeep Dhami (Wellcome Trust Sanger Institute). K562 cells were established from a patient with myelogenous leukaemia in which myelo-proliferation was of the erythrocyte precursors (erythroleukemia). The U937 cell line was established from a patient with diffuse histiocytic lymphoma and

displays many monocytic characteristics. Human CD14+ monocytes were isolated from peripheral blood mononuclear cell samples using Stem Cell Technologies RoboSep CD14 beads by Nicola Foad and obtained courtesy of Dr. Willem Ouwehand (Department of Haematology, University of Cambridge). H9 Human Embryonic Stem Cells (Thomson *et al*., 1998) were grown and sorted for expression of SSEA3 by Dr. Enrique Millan as described in section 2.4.1 and obtained courtesy of Prof. Roger Pedersen (Cambridge Institute for Medical Research).

Methods

2.4. Tissue culture

2.4.1. Culturing of cell lines

K562 and U937 cell lines were cultured in suspension in 50 ml of media (Sigma) with the appropriate amount of fetal bovine serum (GibcoBRL) and other supplements (Sigma) in 75 cm^2 tissue culture flasks with vented caps (Corning). K562 cells were cultured in DMEM, supplemented with 10% fetal calf serum, 1% penicillin-streptomycin, and 2mM L-glutamine, while U937 cells were cultured in RPMI, supplemented with 10% fetal calf serum, 1% penicillin-streptomycin, and 2mM L-glutamine. K562 and U937 cells were cultured under 5% CO_2 at 37°C. Note: H9 Human ES cells were cultured in chemically defined medium as described previously (Vallier *et al*., 2007). Differentiation of human embryonic stem cells was induced by embryoid body formation as described (Vallier *et al*., 2007). Cells positive and negative for expression of stage specific embryonic antigen 3 (SSEA3) were isolated by fluorescence-activated cell sorting as described (Henderson *et al*., 2002).

Once K562 or U937 cells reached confluency (i.e. 7-8 x 10^5 cells/ml of culture media), media was replenished and sub-culturing was carried out as follows:

1. 25 ml of fresh media was added to each flask and any clumps of cells were gently broken up using a syringe.
2. The culture was then distributed between three new 75 cm^2 flasks (Corning) and further 25 ml of fresh media was added to each flask, effecting a 1/3 dilution of the confluent starting culture.

3. The number of cells needed for standard chromatin immunoprecipitation (ChIP) experiments was 100 million. Therefore, culture volumes to obtain the required number of cells per flask were suitably scaled up in 175 cm^2 culture flasks with vented caps (Corning).

2.4.2. Cell cryopreservation

For frozen storage, cells were pelleted at 259 g for 5 to 8 minutes, and resuspended at approximately $1X10^7$ cells/ml in 10% (v/v) DMSO in FBS (GibcoBRL). The resulting cell mixture was transferred into polypropylene cryotubes which were cooled overnight in an isopropanol bath to -70°C. The cryotubes were then transferred to the gas phase of a liquid nitrogen vessel (approximately -180°C) for permanent storage. To reconstitute cultures, cells were thawed rapidly at 37°C, washed once with fresh media and finally resuspended in 10 ml of fresh media.

2.4.3. Chromatin immunoprecipitation (ChIP)

K562, U937, human monocytes and human ES cell line H9 were used for chromatin immunoprecipitation. Fresh K562 and U937 cultures were grown and used to prepare chromatin. Aliquots of K562 and U937 cells used for each ChIP experiment were subjected to flow-sorting (Cytomation MoFlo High Performance Cell Sorter, Dako Cytomation) in parallel for cell-cycle analysis (Performed by Bee Ling, Wellcome Trust Sanger Institute). For this, 3×10^6 cells were harvested, washed with 10 ml of PBS and then fixed in 5 ml of 70% ethanol. An equal volume of Hoechst 33342 dye (2mg/ml) was added to samples, incubated at 37°C for 15 minutes and analysis of fluorescence was used to determine DNA content of the cells. This step was performed to ensure that the percentage of actively dividing cells was consistent for the preparation of each batch of chromatin.

Approximately 1×10^8 K562 and U937 cells were harvested for each ChIP procedure. The number of human monocytes obtained from donors varied across three biological samples obtained (Table 2.1). 3.49×10^6 SSEA3+ and 2.89×10^6 SSEA3- Human embryonic stem cells were obtained for ChIP.

Fixation

1. The cells were collected by centrifuging at 259 g for 8 minutes at room temperature and resuspended in 50 ml of serum free media in a glass flask.
2. DNA-protein and protein-protein interactions were cross-linked by adding formaldehyde (37%, BDH AnalaR). 500 µl, 1010 µl or 1355 µl formaldehyde was added drop-wise to a final concentration of 0.37%, 0.75% for histone modifications and 1% for transcription factors respectively.
3. The cross-linking was carried out at room temperature with constant but gentle stirring for 10 minutes (for histone modifications) or 15 minutes (for transcription factors).
4. 3.15 ml, 3.41 ml or 3.425 ml (for 0.37%, 0.75% or 1% formaldehyde concentration respectively) of ice-cold 2M glycine was added to a final concentration of 0.125M with constant but gentle stirring for 5 minutes at room temperature to stop the cross-linking reaction.
5. Cells were transferred to 50 ml falcon tubes and kept on ice whenever possible. The cells were pelleted by centrifuging at 259 g for 6-8 minutes at 4°C and washed with 1.5 ml of ice-cold PBS.
6. After washing, the cells were pelleted at 720 g at 4°C for 5 minutes and the supernatant was removed.

Cell and nuclei lysis

7. Cells were lysed by adding 1.5 x pellet volume of ice-cold cell lysis buffer (CLB). The cell pellets were gently resuspended and incubated on ice for 10 minutes.
8. The nuclei were recovered by centrifuging the samples at 1125 g for 5 minutes at 4°C.
9. After carefully removing the supernatant, the nuclei were lysed by resuspending the pellet in 1.2 ml of nuclei lysis buffer (NLB) and incubating on ice for 10 minutes. Note: 0.6 ml of NLB was used when preparing chromatin from human embryonic stem cells as pellet volumes were much smaller.

Sonication

10. 0.72 ml of IP dilution buffer (IPDB) was added and the samples were transferred to 5 ml glass falcon tubes (Falcon 2058).

11. The chromatin was sonicated to reduce the DNA length to an average size of 600 bp using the Sanyo/MES Soniprep sonicator. The tip of the probe was dipped to reach approximately halfway down the total level of the liquid sample and the tube was kept constantly on ice (Conditions for sonication like number of bursts, length of bursts and power setting depend on the sonicator tip used). The settings used for the sonicator were:

    - Amplitude: 14 microns
    - Number of bursts: 8
    - Length of bursts: 30 seconds

The samples were allowed to cool on ice for 1 minute between each pulse (5 µl of the sheared chromatin was run on an agarose gel to check sonication, see step 32).

12. The sonicated chromatin was transferred to 2 ml microfuge tubes and spun down at 18000 g for 10 minutes at 4ºC.

Immunoprecipitation

13. The supernatant was transferred to a 15 ml falcon tube and 4.1 ml of IPDB (NLB:IPDB ratio is 1:4) was added[a]. Note: 1.68 ml of IPDB was added to human embryonic stem cell chromatin preparations to maintain a 1:4 ratio of NLB:IPDB.

14. The chromatin was precleared by adding 100 µl of normal rabbit IgG (Upstate Biotechnology). The samples were incubated for I hour at 4°C on a rotating wheel.

15. 200 µl of homogeneous protein G-agarose suspension (Roche) was added to the precleared chromatin and the samples were incubated for 3-5 hours at 4°C on a rotating wheel.

16. The samples were centrifuged at 1620 g for 2 minutes at 4°C to pellet the protein G-agarose beads and the supernatant was used to set up various immunoprecipitation (IP) conditions in 2 ml microfuge tubes. [Note: the following applies when setting-up standard ChIP conditons with K562 and U937 cells (i.e. 10^7 cells and 5-10 µg of

antibody are used)]. An aliquot of 270 µl of chromatin was stored at -20°C to be used as input sample for array hybridisations. An NLB:IPDB buffer at the ratio of 1:4 was freshly prepared and used to ensure that the final volume of all ChIP conditions was 1350 µl. Experimental and control ChIP conditions were set up as follows:

- Normal species specific IgG control (matching the species from which antibody used in ChIP condition(s) was derived) – 675 µl chromatin + 675 µl NLB:IPDB buffer + 10 µg species specific IgG
- No chromatin control – 1350 NLB:IPDB buffer
- No antibody control - 675 µl chromatin + 675 µl NLB:IPDB buffer
- ChIP conditions – 675 µl chromatin + 675 µl NLB:IPDB buffer + 5-10 µg* of antibody

(*5-10 µg for antibodies raised against histone modifications and 10 µg for the antibodies raised against specific transcription factors).

16a. See Table 2.1 for alterations when setting up ChIP conditions with fewer numbers of cells.

Sample	**Number of cells used in chromatin preparation**	**Final volume of chromatin preparation (µl)**	**Number of cells used per ChIP condition**	**Volume of chromatin used in ChIP (µl)**	**Volume of NLB:IPDB buffer (µl)**	**Antibody amount (ug)**	**Volume of chromatin used to prepare Input DNA (µl)**
K562	10^8	6000	10^6	60	1290	0.1 - 10	270
K562	10^8	6000	10^5	6	1344	0.1 - 10	270
K562	10^8	6000	10^4	0.6	1350	0.1 - 10	270
Monocytes (BR 1)	5.6×10^7	6000	10^6	100	1250	1	540
Monocytes (BR2)	2.5×10^7	6000	10^6	200	1150	1	1000
Monocytes (BR3)	5.0×10^7	6000	10^6	125	1225	1	540
hESCs: SSEA3+	3.49×10^6	3000	10^5	86	1264	0.5	1000
hESCs: SSEA3-	2.89×10^6	3000	10^5	104	1246	0.5	1000

Table 2.1: Summary of chromatin, antibody and buffer amounts used in reduced cell number ChIP experiments. BR= biological replicate, hESCs = human embryonic stem cells, SSEA3 = stage specific embryonic antigen 3.

17. The samples were incubated at 4°C overnight on a rotating wheel.
18. The samples were centrifuged at 18000 g for 5 minutes at 4ºC and the samples were transferred to fresh 2 ml microfuge tubes. 50 µl of homogeneous protein G-agarose suspension was added to each sample and the samples were incubated at 4°C for at least 3 hours on a rotating wheel.
19. The samples were centrifuged at 6800 g for 30 seconds at 4°C to pellet the protein G-agarose beads.
20. The supernatant was removed and the protein G-agarose beads were carefully washed. For each wash, the wash buffer was added, the samples were vortexed briefly, were centrifuged at 6800 g for 2 minutes at 4°C and left to stand on ice for 1 minute before removing the supernatant. The washes were carried out in the following sequence:
    a) The beads were washed twice with 750 µl of cold IP wash buffer 1. The beads were transferred to a 1.5 ml microfuge tube after the first wash.
    b) The beads were washed once with 750 µl of cold IP wash buffer 2.
    c) The beads were washed twice with 750 µl of cold TE pH 8.0.

<u>Elution</u>

21. DNA-protein-antibody complexes were eluted from the protein G-agarose beads by adding 225 µl of IP elution buffer (IPEB). The bead pellets were resuspended in IPEB, briefly vortexed and centrifuged at 6800 g for 2 minutes at room temperature.
22. The supernatant was collected in fresh 1.5 ml microfuge tubes. The bead pellets in the original tubes were resuspended in 225 µl of IPEB again, briefly vortexed and centrifuged at 6800 g for 2 minutes. Both the elutions were combined in the same tube.

Reversal of cross-links

23. The reversal of cross-links step was carried out on the Input sample which was stored at -20°C previously. 0.1 µl of RNase A (10 mg/ml, 50 Kunitz units/mg[*], ICN Biochemicals) and 16.2 µl of 5M NaCl (to the final concentration of 0.3 M) was added to the Input DNA sample.
24. Similarly, 0.2 µl of RNase A (10 mg/ml, 50 Kunitz units/mg[*]) and 27 µl of 5M NaCl (to a final concentration of 0.3 M) was added to each of the IP test samples. All the samples including the Input DNA sample were incubated at 65ºC for 6 hours to reverse the cross-links.
25. 9 µl of Proteinase K (10 mg/ml, 20 U/mg, GibcoBRL) was added to each sample and incubated at 45ºC overnight[b].

Extraction of DNA

26. 2 µl of yeast tRNA (5 mg/ml, Invitrogen) was added to each sample just before adding 250 µl of phenol (Sigma) and 250 µl of chloroform[c].
27. The samples were vortexed and centrifuged at 18000 g for 5 minutes at room temperature. The aqueous layer (top layer) was collected in fresh 1.5 ml microfuge tubes and 500 µl of chloroform was added to each sample.
28. The samples were vortexed and centrifuged at 18000 g for 5 minutes at room temperature. The aqueous layer was transferred to a fresh 2.0 ml microfuge tubes.
29. 5 µg of glycogen (5 mg/ml, Roche), 1 µl of yeast tRNA (5 mg/ml, Invitrogen) and 50 µl of 3M NaAc (pH 5.2) was added to each sample and mixed well. The DNA was precipitated with 1375 µl of 100% ethanol and incubating at –70ºC for 30 minutes (or -20°C overnight).
30. The samples were centrifuged at 20800 g for 20 minutes at 4ºC. The DNA pellets were washed with 500 µl of ice-cold 70% ethanol and air-dried for 10-15 minutes.
31. The DNA pellets of the IP samples were resuspended in 50 µl of sterile filtered HPLC water and 100 µl for the Input DNA samples.
32. 5 µl of each sample was run on a 1% agarose 1XTBE gel and visualised with ethidium bromide to check DNA size. Samples were stored at –20°C.

[a]The sheared chromatin can be snap frozen in liquid nitrogen at this stage and the frozen samples should be stored at -70°C. When needed, the samples should be thawed on ice and the experiment carried on as per the protocol.
[b]The samples can be stored at -20°C after the step no. 22. When needed the samples can be thawed at room temperature and the DNA extracted as per the protocol.
[c]Safety Note – The phenol/chloroform steps were carried out in a fume cabinet.
*The amount of enzyme causing the hydrolysis of RNA at a rate such that k (velocity constant) equals unity at 25°C and pH 5.0.

2.5. Construction of the SCL tiling path microarray and the ENCODE tiling path microarray

2.5.1. Generation of tiling path amplicons

Primers pairs used to amplify PCR products for the SCL tiling array were designed from the relevant genomic sequence of chromosome 1 and PCR amplicons were generated as described (Dhami, PhD thesis, University of Cambridge, 2005). A 5'-(C6) amino-link was added to all forward primers to allow binding to Codelink slides (GE) (Dhami *et al*., 2005; Dhami, submitted). Similarly primer pairs used to amplify PCR products for the ENCODE array were designed using Primer3 and PCR amplicons were generated as described (Koch *et al*., 2007).

2.5.2. Microarray slide printing and processing

Spotting buffer was added at final concentration of 0.25M sodium phosphate pH 8.5 and 0.00025% sodium sarkosyl (BDH) to PCR products prior to arraying. The PCR products were then filtered through multiscreen-GV 96 well filter plates (Millipore) and aliquotted into 384 well plates (Genetix). PCR products were then spotted onto Codelink slides in a 16 block format for the SCL array and a 48 block format for the ENCODE array using a Microgrid II robotic arrayer (Biorobotics/Genomics solutions). In order to generate microarrays containing single-stranded DNA elements, all PCR amplicons were printed on arrays and processed as described (www.sanger.ac.uk/Projects/Microarrays/arraylab/methods.shtml). Arrays were subject to quality control analysis by hybridization of random oligonucleotide sequence coupled to

Alexa-647 (Panomer-9, Invitrogen) according to manufacturer's instructions and visualised by scanning with a Scanarray 4000 confocal laser-based scanner (Perkin-Elmer). Arrays that passed this quality control step were stored at room temperature in a desiccated environment until ready for hybridisation.

2.6. Hybridisation of microarrays

2.6.1. Labelling by random priming of DNA samples

The SCL genomic tiling path array and ENCODE array hybridisations using DNA obtained by chromatin immunoprecipitation were set up using the Tecan HS 4800 hybridisation station (an automated hyb-station). The DNA was labeled using BioPrime Random Labeling Kit (Invitrogen) as described below.

Labelling method used for Tecan HS 4800 hybridisation set-up

1. The following reagents were mixed on ice in a 1.5 ml microfuge:
    - 60 µl 2.5 X Random primer solution
    - x µl DNA *
    - (70.5-x) µl sterile H_2O

* The DNA amount labeled was different for input and ChIP samples. The amount of DNA labeled was 40% of unamplified ChIP DNA and approximately 200 ng of Input DNA (2-5% of recovered Input DNA).

2. This mixture was heated at 100°C for 10 minutes to denature the DNA and then snap-chilled on ice. The following reagents were added to the tubes on ice:
    - 15 µl 10 X dNTP mix
    - 1.5 µl 1 mM Cy3/Cy5 labeled dCTP (1 mM Cy3-dCTP, 1 mM Cy5-dCTP, GE Healthcare). Input samples were labeled with Cy5 dCTP and ChIP samples were labeled with Cy3 dCTP
    - 3 µl Klenow fragment (40 U/µl)

The final volume per labeling reaction was 150 µl.

3. The reagents were mixed gently but thoroughly and incubated at 37°C overnight.
4. 15 µl stop buffer was added to the reaction mix to terminate the reaction.

Note: Only one labeling reaction was required for the SCL genomic tiling path array but two labeling reactions were used for the ENCODE array hybridisations.

2.6.2. Purification of labelled DNA samples

Labelled DNA samples were purified using the protocol described below.

1. Micro-spin G50 columns (GE Healthcare) were used to remove the unlabeled nucleotides from the labelled DNA samples.
2. Three columns were used for each of the 150 μl labelling reactions.
3. The resin was resuspended in the columns by vortexing gently. The caps were loosened and the bottom of the tubes snapped off. The columns were placed in 2.0 ml microfuge tubes and centrifuged at 1700 g for 1 minute.
4. 50 μl of sterile filtered HPLC water was applied to the resin-bed and the columns were centrifuged at 1700 g for 1 minute.
5. The columns were placed in fresh 1.5 ml microfuge tubes and the labelled DNA samples were carefully applied to the resin-bed. The columns were then centrifuged at 1700 g for 2 minutes.
6. The purified DNA samples were collected in the 1.5 ml microfuge tubes and the samples from the same labeled reaction were pooled together. The final volume for the labelled DNA samples was approximately 180 μl.
7. 5 μl of each labelled DNA was analyzed on a 1% agarose 1 X TBE gel and stained with ethidium bromide for visualization. The samples were used for hybridisation and stored in the dark at –20ºC.

2.6.3. Array hybridisation set-up using the Tecan HS 4800 hybridisation station

Tecan HS 4800 hybridisation station (automated hyb-station) was used to set-up ChIP-chip hybridisations (on the SCL genomic tiling path array and ENCODE array). The Tecan HS 4800 is a fully automated hyb station where the microarray slides are loaded on the Tecan's slide holder and the hybridisation mix is agitated to ensure even hybridisation. The SCL array area was 2 x 2 cm and consequently the smaller chambers

were used on the Tecan to set up the hybridisations. However, larger chambers were required for the ENCODE microarrays

Preparation of the hyb station

1. The slide holders and the slide chambers were carefully cleaned and the slides were loaded on to the slide holder.
2. The wash solutions were prepared (described in section 2.1) and poured into the wash bottles of the Tecan and the hybridisation station was primed to remove any air bubbles in the liquid channels and tubing.

Preparation of pre-hybridisation and hybridisation solutions

Note: The following volumes are for hybridisation of the SCL tiling path array.

3. Two 2.0 ml microfuge tubes were set up for each slide - one with the pre-hybridisation mix and the second with the hybridisation mix. The following reagents were mixed together on ice and kept in the dark as much as possible:

Tube 1 (Pre-hybridisation mix)

- 40 µl Herring sperm DNA (10 mg/ml, Sigma)
- 67.5 µl Human Cot 1 DNA (Invitrogen)
- 12.5 µl 3M NaAc (pH 5.2)
- 300 µl 100% ice-cold ethanol

Tube 2 (Hybridisation mix)

- 180 µl Cy3-labeled DNA
- 180 µl Cy5-labeled DNA
- 135 µl Human Cot 1 DNA (Invitrogen)
- 55 µl 3M NaAc (pH 5.2)
- 1200 µl 100% ice-cold ethanol

Note: The following volumes are for hybridisation of the ENCODE array.

4. Three 2.0 ml microfuge tubes were set up for each slide - one pre-hybridisation mix and two hybridisation mixtures. The following reagents were mixed together on ice and kept in the dark as much as possible:

Tube 1 (Pre-hybridisation mix)

- 80 µl Herring sperm DNA (10 mg/ml, Sigma)
- 135 µl Human Cot 1 DNA (Invitrogen)
- 25 µl 3M NaAc (pH 5.2)
- 800 µl 100% ice-cold ethanol

Tube 2 (Hybridisation mix x 2)

- 180 µl Cy3-labeled DNA
- 180 µl Cy5-labeled DNA
- 135 µl Human Cot 1 DNA (Invitrogen)
- 55 µl 3M NaAc (pH 5.2)
- 1200 µl 100% ice-cold ethanol

5. The DNA samples were precipitated by incubating at –70ºC for 60 minutes (or -20°C overnight) and then centrifuged at 18000 g for 20 minutes at room temperature. The pellets were washed with 500 µl of 80% ethanol and air dried.
6. The pre-hybridisation and hybridisation DNA pellets were resuspended in 120 µl of Tecan hyb buffer each (for hybridisation on the SCL array) and 3 µl of yeast tRNA (100 µg/µl, Invitrogen) was added to the hybridisation mix.
7. For hybridising the ENCODE arrays, the pre-hybridisation DNA pellet was resuspended in 180 µl of Tecan-hyb buffer. Each hybridisation pellet was resuspended in 90 µl of Tecan-hyb buffer and then combined to give a total volume of 180 µl. 3 µl of yeast tRNA (100 µg/µl, Invitrogen) was added to the hybridisation mix. Heating the solutions at 70ºC helps to resuspend the pellets properly.

Array hybridisation and washing

8. The pre-hybridisation and hybridisation solutions were denatured at 100°C for 10 minutes. The hybridisation solution was snap-chilled on ice and then pre-annealed at 37°C for 1 hour. The pre-hybridisation solution was kept at 70°C until applied to the slide.
9. After vortexing, using a positive displacement pipette, 100 µl of the pre-hybridisation solution (for SCL array) or 160 µl of the pre-hybridisation solution (for ENCODE array) was injected onto the slide very slowly and carefully to avoid any air bubbles.
10. The pre-hybridisation step was performed at 37°C for 1 hour. The slides were washed once with PBS/0.05% Tween and dried with short blasts of nitrogen gas.
11. 100 µl or 160 µl of the hybridisation solution was injected slowly onto the slide using a displacement pipette. The hybridisation step was performed at 37°C for 48 hours.
12. The slide washing was carried out on the Tecan which was programmed to perform the washes in the sequence listed in Table 2.2.

Steps	Wash Solutions	Temperature	No. of Washes	Wash Duration	
				Wash time	Soak time
1	PBS/0.05% Tween	37°C	10	1 min	30 secs
2	0.1 X SSC	52°C	5	1 min	2 min
3	PBS/0.05% Tween	R/T	10	1 min	30 secs
4	HPLC water	R/T	2	30 secs	

Table 2.2: Wash steps for hybridisations performed on the Tecan. The solutions were prepared in advance (section 2.1) using HPLC water. R/T = room temperature.

13. The slides were dried on the Tecan using nitrogen gas and stored in a dark, low temperature, low humidity environment until ready for scanning to prevent loss of fluorescent signal.

2.6.4. Scanning and processing of ChIP-chip data

After the images were scanned, Scanarray Express software (Perkin Elmer) was used to analyse the scanned images as described below. The analysis output from both the softwares were analysed using an excel spreadsheet.

1. Cy3 and Cy5 images at 5 µm resolution were acquired using the Scanarray 4000 confocal laser-based scanner (Perkin-Elmer) using a laser power of 100% and a photo multiplier tube (PMT) value of between 70%-85%.
2. ScanArray Express (Perkin Elmer) was used to quantitate the fluorescent intensities of the spots using the adaptive circle quantitation and the TOTAL normalisation methods. This software can automatically locate the spot position on the scanned image of the array to obtain the signal intensity values. Mean intensity ratios (intensity-background) were reported for each spot representing an array element. Those spots identified as not found were removed from the data set.
3. Further analysis of the ChIP-chip data was carried out in a Microsoft Excel spreadsheet in which each array element was associated with its genomic sequence position information.
4. The SCL microarray data was visualised by plotting the mean ratios of all array elements along the Y-axis and the respective genomic positions along the X-axis. ENCODE data was visualised in the UCSC genome browser (Kuhn *et al*., 2007) by uploading a 'wiggle' file which contained chromosome start and end coordinates and fold enrichment ratios or a 'bed' file which contained the location of ChIPOTle peaks only.
5. The median ratio of technical or biological replicates performed using the ENCODE array were calculated using a script written in the R programming language (http://www.sanger.ac.uk/PostGenomics/encode/data-access.shtml).
6. The baseline levels of each SCL or ENCODE data set was normalised to a value of one, so that all the experiments could be directly compared from this baseline value. This was done by calculating the median ratio for each experiment and dividing all the ratios (obtained in that experiment) by this number.

2.7. Analysis of ChIP-chip data with respect to genomic features

The ChIPOTle program (Buck *et al*., 2005) was used to define peaks of enrichment in ChIP-chip data sets by using a sliding window approach and then estimating the significance of enrichment for a genomic region using a standard Gaussian error function. ChIPOTle assigns a p-value to the average $\log_2$ ratio within each window and are corrected for multiple comparisons using the Bonferroni correction. The significance p-value cut-off assigned for analysing histone H3 acetylation, H4 acetylation and H3K4 methylation data sets was p0.0005. A p-value cutoff that produces about 50 times more significant regions than significant negative regions was suggested to be a satisfactory cut-off for the majority of applications (Buck *et al*., 2005). A window size of 2000bp and a step size of 500bp (which should be approximately ¼ the window size) was used analyzing ENCODE histone modification data sets. A p-value cut-off of p0.00001 was used for the analysis of ENCODE transcription factor data sets (analysis performed by Dr. Rob Andrews, Wellcome Trust Sanger Institute).

GENCODE annotations were downloaded from GENCODE genes version 02.2 from ftp://genome.imim.es/pub/projects/gencode/data/havana-encode/current/44regions/44 regions_CHR_coord.gtf. The distribution of histone modifications and transcription factors with respect to GENCODE annotated features such as transcription start sites was performed using scripts written in the R programming language (performed by Dr. Rob Andrews). All ChIP-chip data was analysed on NCBI human genome build 35 (hg17).

2.8. Statistical analysis

Statistical analysis of data was performed using Microsoft Excel. This included the calculation of standard deviations, mean coefficient of variations (CVs), and Pearson correlation coefficients. Standard deviations were calculated using the STDEV function, Pearson correlation coefficients were calculated using the PEARSON function. CVs were calculated by dividing the standard deviation of a set of values by the mean of those values and then expressed as percentage by multiplying by 100.

Significant enrichments in SCL array data were calculated by determining the mean background ratios of two regions which did not contain any known regulatory elements in K562 (Dhami, PhD thesis, Cambridge Univeristy, 2005). These regions spanned

47262287 bp to 47343557 bp, and 47424426 bp to 47489321 bp on chromosome 1. The significant enrichment threshold was three standard deviations above the mean background ratios. Significant enrichments in ENCODE array data sets were also calculated by determining the mean ratio of the entire data set and the significant enrichment threshold was three standard deviations above the mean ratio of the data set.

Over-representation of ChIPOTle sites with respect to genomic features was performed using a randomisation strategy. Each experimental ChIPOTle dataset was randomised 100 times to generate 100 random data sets within the ENCODE regions, conserving the size and number of ChIPOTle sites. The mean overlap with genomic features was then calculated and compared to experimental values (performed by Dr. Rob Andrews).

2.9. Gene expression analysis

Four replicate K562 Affymetrix U133 plus 2.0 gene expression data sets were provided by Dr. Christoph Koch and Dr.Philippe Couttet (Wellcome Trust Sanger Institute). All Affymetrix analyses were performed by Dr. Rob Andrews using the affy package in Bioconductor with default parameters (Gentleman *et al*., 2004). Normalised MAS5 data was generated using the affy package, which was then used for Present/Absent calls per probe set. Robust Multichip Average (RMA) data analysis was also performed using the affy package, in which background correction and normalization was performed to give an expression value per probe set. The RMA expression values for ENCODE genes present on the Affymetrix array were ranked in order of expression as high (100-75%), low (75%-50%), indeterminate (50%-25%), and off (25%-0%). The RMA values of these four classes of genes were compared to MAS5 absent and present expression calls. Genes which were called as present by MAS5 and which were in the top 50% of ranked RMA values were considered as expressed genes (either as high or low expression based on the above criteria). Genes which were called by MAS5 as absent and which were in the bottom 50% of ranked RMA values were considered as not expressed. Genes which showed discrepancies between MAS5 and RMA data were classified as indeterminate.

Two replicate human CD14+ monocyte illumina human ref-8 expression beadchip data sets were provided by Dr. Nick Watkins (Department of Haematology, University of Cambridge). Illumina gene expression analyses were performed in Bioconductor and

BeadStudio packages. Present/absent analysis was performed using the lumi package within Bioconductor where present is called when the detection score provided by BeadStudio was greater than 0.95. Transcript abundance analysis was performed using the lumi package using quantile normalization (performed by Dr. Rob Andrews).

2.10. Receiver operator characteristic (ROC) curve analysis

The association of histone modifications and transcriptions factors at TSSs with gene expression state was examined by plotting ROC curves for each of the histone modifications and transcription factors within 1kb around the TSS (analysis performed by Dr. Ulas Karaöz, Boston University). Present and absent MAS5 calls from 238 Affymetrix probe-sets were used to define the on/off state. The ROC curves illustrate the predictive accuracy of histone marks or the presence of a transcription factor on classifying the expression state of genes. A threshold was applied to each histone modification or transcription factor level and a prediction of the on (or off) state of a gene is made if the level is higher (or lower) than the threshold. The true positive rate is plotted against the false positive rate for each threshold and all possible thresholds are applied so that a curve is obtained (with each point of the curve corresponding to a threshold). The best operating point is the point on a ROC plot which lies on a 45 degree line closest to the northwest corner of the ROC plot.

2.11. DNA sequence motif analysis

Motif matrices from JASPAR (Bryne *et al*., 2007) and TRANSFAC (Matys *et al*., 2006) databases were used to search for enrichment in transcription factor ChIPOTle sites. The DME program (Smith *et al*., 2005) was also used to search for novel motifs in ChIPOTle sites. 1000bp centred sequences were extracted from ChIPOTle sites and defined as foreground sequences, i.e. sequences in which a binding motif was believed to be present. 1000 bp sequences flanking ChIPOTle sites were also extracted and defined as background sequences, i.e. sequences in which no specific binding motif was believed to be present. Motifs which distinguished foreground sequences from background sequences with a low relative error rate were identified. The sensitivity (Sn) and specificity (Sp) of these motifs was also calculated. The sensitivity associated with a motif and p-value cut-

off is the proportion of true foreground sequences that were classified as foreground; the specificity is the proportion of true background sequences classified as background sequences. The relative error rate (Error) for a given motif and associated with a particular p-value cut-off was 1-(Sn+Sp)/2 (Smith *et al*., 2005).

2.12. Real-time PCR

A. Primer design

1. Primer pairs for all the real-time PCR assays, performed for this study, were designed by using the Primer Express software version 2.0 (Applied Biosystems).
2. Primer pair sequences were compared against the entire human genome using e-PCR (Schuler 1997). The amplicons generated by these primer pairs were between 70 bp to 150 bp in length.
3. Standard curves were generated for the primer pairs used in the ChIP verification assays. From these, the PCR yields were calculated for each of the tested primer pairs.

The complete lists of all the primer pair sequences, used in the real-time PCR assays, are provided in Appendix 1.

B. Real-time PCR amplification

The chromatin immunoprecipitated (ChIP) DNA samples were used to set-up quantitative real-time PCR as follows:

1. The ChIP DNA samples were diluted to 1 in 10 dilution i.e. 5 µl of the sample was resuspended in 45 µl of sterile filtered HPLC water.
2. The SYBR green PCR was set-up in a 96-well plate (Applied Biosystems) in a 25 µl reaction, in triplicate for each sample, by mixing the following reagents on ice:
    - 2.5 µl Water
    - 5 µl 1.5 µM forward and reverse primer mix
    - 12.5 µl SYBR green PCR mix (Applied Biosystems)
    - 5 µl ChIP DNA samples

3. PCR was performed on a 7700 sequence detection system (Applied Biosystems) .The following thermal cyclic conditions were used: 50ºC for 2 min; 95ºC for 10 min; then 40 cycles of: 95ºC for 15 sec and 60ºC for 1 min.
4. C_T values were extracted using Sequence Detector 1.7a (Applied Biosystems) with the same threshold and the ΔC_T values were determined as follows:

   $\Delta C_T = C_T$ input - C_T ChIP sample
5. Fold enrichments were calculated by using the following formula:

   Fold enrichment = $(1+\text{PCR yield})^{\Delta CT}$

   Mean fold enrichments were calculated for each assay and data sets were normalised to a median of 1.

2.13. Western blotting procedure

<u>Nuclear lysate preparation</u>

Note: All buffers were freshly prepared and stored at 4ºC. Extractions were carried out on ice.

1. 10^7 K562 cells were harvested by centrifuging at 259g for 5 minutes. The supernatant was removed and the cell pellet was washed in PBS.
2. The cells were then spun down as before, the supernatant was removed and the pellet was resuspended in 1 ml of cell lysis buffer containing 10 µl protease inhibitor cocktail (Sigma).
3. The cell pellet was transferred to a 1.5ml tube (pre-cooled on ice) and left on ice for 5 minutes to ensure efficient cell lysis.
4. Nuclei were spun down by centrifuging for 1 minute at 11000g at 4ºC. The supernatant was then removed and the nuclei were washed in 1 ml of cell lysis buffer containing 10 µl of protease inhibitor cocktail.
5. The nuclei were spun down by centrifuging for 1 minute at 11000g at 4ºC and the supernatant was removed

6. 1µl 0.1M DTT & 1µl Sigma protease inhibitor cocktail was freshly added to 98µl of nuclear protein extraction buffer. The pelleted nuclei were resuspended in 70µl of this extraction buffer.
7. The samples were placed in a small box of ice on top of a vortexer, taped down and vortexed at the lowest setting possible for 30 minutes to avoid foam formation.
8. The samples were centrifuged for 1 minute at 11000g at 4ºC to pellet any debris. The supernatant containing nuclear proteins was transferred to a fresh tube and store at -70ºC.

Determination of protein concentration

A Bradford assay was performed as follows to determine protein concentration:

9. 0.8 ml of HPLC water was added to seven 1.5 ml spectrophotometer cuvettes (Biorad).
10. 0, 2, 5, 10, 15, 20g µl of BSA (1 µg/µl) was added to individual cuvettes to generate a standard curve. 2 µl of protein extract was added to the final cuvette.
11. 200 µl of proein assay dye reagent concentrate (Biorad) was added to each cuvette and mixed well by pipetting.
12. The cuvettes were kept in the dark for 20 minutes and the optical density was measured at 595 nm using a spectrophotometer.

Electrophoresis of protein samples

13. NuPAGE™ 4-12% Bis-Tris polyacrylamide gels (Invitrogen) and the XCell Surelock™ Mini-Cell (Invitrogen) were used for the electrophoresis of protein samples.
14. 1 litre of 1 X MOPS SDS running buffer was prepared and placed at 4°C to cool
15. The comb and tape at the bottom of the gel were removed and the wells were washed with 1 X MOPS running buffer using a syringe to remove acrylamide traces.
16. The gel was then slotted into the upper chamber (cathode) of the tank and locked into place. The upper chamber was filled with 200 ml of 1 X MOPS running buffer.

17. The lower buffer chamber (anode) was half-filed by pouring 300 ml of 1 X MOPS running buffer.
18. Protein samples were thawed at 37°C for 5 minutes if stored at -70°C and 20 µg protein samples were diluted in 4 X LDS sample buffer (maximum loading volume for 1.5 mm 10 well gels is 37 µl).
19. Protein samples were heated at 70°C for 10 minutes and then placed back on ice.
20. The Samples were then loaded onto the gel, along with 10 µl of SeeBlue® Plus2 pre-stained standard.
21. The gel tank was then placed in a cold-room (4°C) and the gel was run at 200 V constant, with a starting current of 125 mA, for 90 minutes.

Transfer of proteins onto PVDF membrane (Western Blotting)

22. 1 litre of 1 X NuPAGE transfer buffer was prepared while the gel was running and cooled at 4°C.
23. The gel, blotting pads, and filter paper (Whatman, cut to size 7.5 x 8 cm) were equilibrated in 1 X NuPAGE transfer buffer for 5 minutes in a fume hood.
24. PVDF transfer membrane (Sigma, 0.45 µm) was cut to size (7.5 x 8 cm) equilibrated in 100% methanol for 5 minutes and then in 1 x NuPAGE transfer buffer for 5 minutes.
25. The transfer 'sandwich' was then assembled in the blot module as shown:
    - Top (+)
    - 2 x blotting pads
    - Filter paper
    - Transfer membrane
    - Gel
    - Filter paper
    - 2 x blotting pads
    - Bottom (-)

26. Air bubbles in the blotting pads and between the gel and membrane were carefully removed. The blot module was then held together and slid into the guide rails on the lower chamber of the tank.
27. The blot module was filled with 1 x NuPAGE transfer buffer until the gel/membrane 'sandwich' was covered with buffer.
28. The transfer was performed in a cold room at 35 V constant with a starting current of 170 mA for 90 minutes.

Blocking the membrane

29. The PVDF transfer membrane was placed in blocking buffer for 60 minutes at room temperature on an orbital shaker and rinsed briefly in TBST.

Incubation with primary antibody

30. The membrane was then incubated with the primary antibody diluted 1:2000 in 15 ml blocking buffer. Membranes where incubated at 4°C over-night on an orbital shaker.
31. The membrane was then washed four times with TBST (15 minutes per wash).

Incubation with secondary antibody

32. The membrane was incubated with the appropriate Horseradish peroxidase-conjugated secondary antibody diluted 1:20000 in 20 ml blocking buffer at room temperature for 60 minutes on an orbital shaker. The membrane was then washed 4 x 15 minutes with TBST.

Immunodetection

33. ECL Plus™ Western Blotting detection reagents (GE healthcare) were used for immunodetection. Solution A (acridan substrate solution containing tris buffer) was mixed with solution B (acridan substrate solution in dioxane and ethanol) in a 40:1 ratio.
34. The detection solution was applied to the membrane for 1 minute (dark room), then placed onto an ECL Hyperfilm™ (GE healthcare) in an X-ray film cassette. The film was exposed for 5-10 seconds and developed using an automated X-ray film developer (Xograph).