

**Identification of Target Genes
of an Erythroid Transcription Factor
Complex Containing SCL (TAL1)**

Hau Ling Jim

Homerton College



**UNIVERSITY OF
CAMBRIDGE**

Thesis submitted for the degree of
Doctor of Philosophy

University of Cambridge

2008

Table of Contents

Chapter 1 Introduction

1.1 Regulation of gene expression	1
1.1.1 Cis-acting regulatory elements	2
1.1.1.1 Promoters.....	3
1.1.1.2 Enhancers.....	4
1.1.1.3 Silencers.....	5
1.1.1.4 Insulators.....	6
1.1.1.5 Locus control regions.....	6
1.1.1.6 Scaffold/Matrix attachment regions (S/MARs).....	6
1.1.2 Trans-acting proteins involved in transcriptional regulation.....	6
1.1.2.1 RNA polymerase.....	7
1.1.2.2 Basal/General transcription factors.....	7
1.1.2.3 Sequence-specific transcription factors.....	8
1.1.2.4 Coactivators/ Corepressors.....	10
1.1.2.5 Chromatin modifying factors.....	11
1.1.3 Transcription factor.....	13
1.2 Transcriptional regulatory networks	14
1.2.1 Properties of a transcriptional network.....	15
1.2.2 Experimental and computational approaches for deciphering regulatory networks.....	16
1.3 Experimental and computational approaches to understand transcriptional regulation	17
1.3.1 Gene perturbation by RNA interference.....	17
1.3.1.1 Discovery and mechanism of RNA interference.....	17
1.3.1.2 Inducing RNAi in various organisms.....	20
1.3.1.3 Non-specific effects of RNAi.....	22
1.3.1.4 Applications of RNAi.....	23
1.3.2 Gene expression profiling.....	26
1.3.2.1 Conventional methods.....	27
1.3.2.2 Sequencing-based methods.....	28
1.3.2.3 Microarray-based methods.....	29
1.3.2.4 Applications of microarrays in gene expression profiling.....	31
1.3.3 Characterisation of regulatory elements.....	32
1.3.3.1 Conventional methods.....	32
1.3.3.2 ChIP-based methods.....	35
1.3.3.3 High-throughput ChIP applications.....	36
1.3.3.4 Alternative ChIP approaches.....	40
1.3.4 Computational approaches to study gene regulation.....	41

1.3.4.1 Promoter prediction.....	41
1.3.4.2 Transcription factor binding site prediction.....	41
1.3.4.3 Comparative sequence analyses.....	42
1.4 Haematopoiesis.....	42
1.4.1 Embryonic origin and lineages of haematopoiesis.....	43
1.4.2 Regulation of haematopoiesis.....	45
1.4.2.1 SCL.....	46
1.4.2.2 GATA1.....	55
1.4.2.3 E2A/TCF3.....	62
1.4.2.4 LMO2.....	64
1.4.2.5 LDB1.....	65
1.4.2.6 Transcription regulatory networks in haematopoietic development.....	66
1.5 Aims of thesis.....	67
Chapter 2 Materials and Methods	
2.1 Composition of solutions.....	68
2.2 Reagents.....	72
2.3 Cell lines.....	73
2.4 Tissue culture.....	73
2.4.1 Culturing and propagation of cell lines.....	73
2.4.2 Cryopreservation of cell lines.....	73
2.5 Transfection of siRNA.....	74
2.6 RNA extraction.....	74
2.6.1 Total RNA extraction by TRIZOL.....	74
2.6.2 DNase treatment for RNA samples.....	75
2.6.3 Phenol-chloroform extraction and ethanol precipitation.....	75
2.7 Reverse transcription.....	76
2.8 Quantitative real-time PCR.....	76
2.8.1 Primer Design.....	76
2.8.2 Quantitative real-time PCR.....	76
2.8.3 Data analyses.....	76
2.9 Protein extraction.....	78
2.9.1 Protein extraction.....	78
2.9.2 Protein quantification.....	78
2.10 Western Blotting.....	79
2.10.1 Sample preparation.....	79
2.10.2 SDS-PAGE.....	79
2.10.3 Blotting.....	79

2.10.4 Immunoblotting and detection.....	80
2.11 Flow cytometry analysis for fluorescent oligo transfection.....	81
2.12 Cell morphology studies.....	82
2.13 Growth arrest studies.....	83
2.14 Glycophorin A expression detection in K562 and HEL 92.1.7.....	83
2.15 Differentiation study of K562.....	84
2.16 Affymetrix GeneChip expression analysis.....	84
2.16.1 Eukaryotic target preparation.....	84
2.16.2 Target hybridisation.....	89
2.16.3 Fluidics station setup.....	89
2.16.4 Probe array washing, staining and scanning.....	90
2.16.5 Data analysis.....	90
2.17 Chromatin immunoprecipitation on microarrays (ChIP-on-chip).....	92
2.17.1 Chromatin preparation.....	92
2.17.2 Immunoprecipitation.....	93
2.17.3 Reversal of crosslinking and DNA extraction.....	94
2.17.4 Labelling of ChIP DNA and input DNA with cyanine dyes.....	94
2.17.5 Hybridisation of the human transcription factor promoter array.....	95
2.17.6 Slide washing, scanning and data analyses.....	96
2.18 Sequence analysis of promoters.....	98
2.18.1 Motif discovery of putative targets.....	98
2.18.2 Conserved transcription factor binding sites identification.....	98
2.19 Transcriptional network generation.....	98

Chapter 3 Developing working siRNAs for members of the SCL erythroid complex in K562

3.1 Introduction.....	99
3.1.1 The SCL erythroid complex and its downstream regulation.....	99
3.1.2 The RNA interference system.....	100
3.1.2.1 Comparison between RNAi and traditional knockouts.....	100
3.1.2.2 Components of a good RNAi system.....	100
3.1.2.3 Advantages of using siRNA in the current study.....	104
3.1.3 The cell culture system under study.....	105
3.2 Aims of this chapter.....	106
3.3 Overall strategy.....	106
3.4 Results.....	107
3.4.1 Developing working siRNA assays against TFs in the SCL erythroid complex (SEC).....	107

3.4.2	Characterisation of antibodies for western blotting analyses.....	110
3.4.3	Determination of the transfection efficiency of siRNAs.....	114
3.4.4	Changes in K562 cell growth and morphology induced by siRNA transfection.....	116
3.4.5	Time-course study of siRNA knockdown.....	120
3.4.5.1	Knockdown of SCL.....	122
3.4.5.2	Knockdown of GATA1.....	124
3.4.5.3	Knockdown of E2A.....	126
3.4.5.4	Knockdown of LDB1.....	129
3.4.5.5	Knockdown of LMO2.....	132
3.5	Discussion.....	133
3.5.1	siRNA delivery.....	133
3.5.2	Effect of siRNA transfection on phenotypic changes.....	134
3.5.3	siRNA-induced knockdown of the SCL erythroid complex in time-course study.....	135
3.6	Conclusions.....	139

Chapter 4 Expression profiling analyses of siRNA knockdowns of the SCL erythroid complex

4.1	Introduction.....	140
4.1.1	Information generated using expression profiling of perturbation of transcription factors.....	140
4.1.2	Expression profiling studies of the SCL erythroid complex in literature.....	141
4.1.3	The Affymetrix GeneChip expression array.....	142
4.1.4	Microarray data analyses.....	146
4.1.5	Confirmation and validation of data.....	149
4.2	Aims of this chapter.....	150
4.3	Overall strategy.....	150
4.4	Results.....	152
4.4.1	Preparation and quality control of samples.....	152
4.4.2	Quality control of Affymetrix GeneChips.....	155
4.4.3	Data analysis of Affymetrix GeneChips.....	161
4.4.3.1	Normalisation and statistical analyses of Affymetrix GeneChip data.....	162
4.4.3.2	Differentially-expressed genes and comparison of two siRNAs.....	165
4.4.3.3	Validation of selected differentially-expressed genes by quantitative PCR.....	166
4.4.3.4	Further study and classification of differentially-expressed genes.....	173
4.4.3.5	Co-regulation of transcription factors in the SCL erythroid complex.....	178
4.4.4	Motif discovery of co-regulated putative target genes.....	180
4.5	Discussion.....	182
4.5.1	Affymetrix GeneChips as a platform of expression profiling.....	182
4.5.2	The SCL erythroid complex regulates transcription factors.....	184

4.5.3 Identification of haematopoietic-related genes regulated by members of the SCL erythroid complex.....	184
4.5.4 Auto-regulation of the SCL erythroid complex.....	185
4.5.5 Motif discovery at target genes.....	185

Chapter 5 ChIP-on-chip analyses of the SCL erythroid complex

5.1 Introduction.....	187
5.1.1 ChIP-on-chip: principles and issues.....	187
5.1.2 Human transcription factor promoter array platform.....	191
5.1.3 ChIP studies of transcription factors in the SCL erythroid complex in the literature.....	194
5.2 Aims of this chapter.....	195
5.3 Overall strategy.....	195
5.4 Results.....	197
5.4.1 Quality control of various steps of chromatin-immunoprecipitation.....	197
5.4.1.1 Culturing of cells.....	197
5.4.1.2 Preparation of cross-linked chromatin.....	198
5.4.1.3 Extraction of ChIP DNA.....	198
5.4.1.4 Labelling of input and ChIP DNA.....	199
5.4.1.5 Hybridisation and analyses of the transcription factor promoter array.....	201
5.4.2 Evaluation of working antibodies by positive control elements of the array.....	202
5.4.3 Data analyses of enriched promoters.....	205
5.4.3.1 Overall strategy of statistical analyses.....	206
5.4.3.2 Data analyses for the selection of putative target genes.....	208
5.4.3.3 Classification and literature review of putative target genes.....	210
5.4.4 Characterisation of a subset of putative target genes.....	214
5.4.4.1 Criteria for selection of subset of genes for further studies.....	214
5.4.4.2 Transcription factor binding sites (TFBS) studies and comparative genomic analyses of enriched promoters.....	216
5.4.4.3 ChIP-qPCR validation of promoter binding events.....	218
5.4.4.4 Comparison of ChIP-on-chip, ChIP-qPCR and motif analyses.....	225
5.5 Discussion.....	228
5.5.1 Validation of promoter-binding events.....	228
5.5.2 Validation of known target genes.....	230
5.5.3 Novel targets of the SCL erythroid complex.....	230
5.5.4 The sequences of the putative binding sites of the SCL erythroid complex.....	231
5.5.5 Biological roles of novel targets of the SCL erythroid complex.....	232
5.5.6 Autoregulation of members of the SCL erythroid complex.....	233
5.5.7 Limitations of the ChIP-on-chip studies.....	233

5.6 Conclusions	235
------------------------------	-----

Chapter 6 Further characterisation of putative target genes of members of the SCL erythroid complex

6.1 Introduction	236
6.1.1 Expression profiling of siRNA-induced knockdown of the SCL erythroid complex.....	236
6.1.2 ChIP-on-chip study of the SCL erythroid complex.....	237
6.1.3 Auto-regulation of the SCL erythroid complex.....	237
6.1.4 Regulation of the LYL1 gene.....	238
6.1.5 Generation of transcription networks.....	238
6.2 Aims of this chapter	239
6.3 Results	240
6.3.1 Further characterisation of putative target genes identified in Affymetrix and ChIP-on-chip studies.....	240
6.3.1.1 Comparison between putative target genes of Affymetrix expression and ChIP-on-chip studies....	240
6.3.1.2 Expression time-course study of the putative target genes during siRNA knockdown of the SCL erythroid complex.....	243
6.3.1.3 Further evidence for auto-regulation of the SCL erythroid complex at the level of gene expression.....	249
6.3.1.4 ChIP-on-chip study of GATA1 knockdown.....	253
6.3.1.5 Off-promoter regulation of SCL and LYL1.....	258
6.3.2 Integration of expression and ChIP data: Derivation of transcriptional interaction networks in the erythropoietic lineage.....	263
6.3.2.1 Networks generation based on ChIP-on-chip data.....	263
6.3.2.2 Integration of ChIP-qPCR data into networks based on ChIP-on-chip interactions.....	265
6.3.2.3 Integration of expression information into interaction networks.....	269
6.3.2.4 Networks generation with an integration of all experimental studies.....	275
6.3.2.5 Identification of network motifs.....	277
6.4 Discussion	279
6.4.1 Low correlation rate between Affymetrix and ChIP-on-chip studies.....	279
6.4.2 The SCL complex transcription network.....	281
6.5 Conclusions	282

Chapter 7 Summary and future work

7.1 Summary of work presented in this thesis	283
7.1.1 Developing working siRNAs for members of the SCL erythroid complex in K562 (Chapter 3).....	283
7.1.2 Expression profiling analyses of siRNA knockdowns of the SCL erythroid complex (Chapter 4).....	284
7.1.3 ChIP-on-chip analyses of the SCL erythroid complex (Chapter 5).....	285

7.1.4 Further characterisation of putative target genes of members of the SCL erythroid complex (Chapter 6).....	285
7.2 Future work	287
7.2.1 Confirmation of putative target genes in primary cells.....	287
7.2.2 Investigation of histone modifications associated with the regulatory pattern.....	288
7.2.3 Identification of all possible regulatory elements bound by the SCL erythroid complex.....	289
7.2.4 Knockout and ChIP studies of interesting putative target genes.....	289
7.2.5 Identification of other interacting partners in the SCL erythroid complex.....	290
7.2.6 Functional assays for the putative LYL1 enhancer.....	291
7.2.7 Studies of changes in expression or promoter binding of target genes at different stages of erythroid differentiation.....	292
7.3 Final thoughts	293
References	294
Appendices	323
Appendix 1A: Sequences of primer pairs used in qPCR to analyse expression of the SCL complex during siRNA knockdown assays.....	323
Appendix 1B: Sequences of primer pairs used in qPCR to analyse expression of housekeeping genes during siRNA knockdown assays.....	323
Appendix 1C: Sequences of primer pairs used in qPCR to analyse expression of ChIP-on-chip target genes during time-course siRNA knockdown assays.....	323
Appendix 1D: Sequences of primer pairs used in ChIP-qPCR to analyse enrichments of ChIP-on-chip target genes.....	324
Appendix 1E: Sequences of primer pairs for the negative control regions used in ChIP-qPCR.....	325
Appendix 2: Gene Ontology classification of differentially expressed genes in Affymetric GeneChip analysis of siRNA knockdown study.....	326
Appendix 3A: Characterisation of antibodies for the SCL erythroid complex for western blot.....	341
Appendix 3B: Characterisation of antibodies for the SCL erythroid complex for ChIP-on-chip.....	342
Appendix 4: Multiple sequence alignments of transcription factor binding sites at promoter regions of selected putative target genes.....	343
Appendix 5: Confirmation of GATA1 knockdown by qPCR and western blotting in GATA1 ChIP-on-chip study.....	348

Abstract

Identification of Target Genes of an Erythroid Transcription Factor Complex Containing SCL (TAL1)

Haematopoiesis is the process whereby haematopoietic stem cells give rise to mature blood cell lineages. The SCL (TAL1) gene encodes a key transcription factor (TF) which is expressed in various blood lineages and is essential for haematopoietic development. It has been shown that the SCL protein forms a multi-protein complex during erythroid development with other TFs (GATA1, E2A, LDB1, and LMO2) which binds to a sequence-specific motif to regulate its target genes. Two complementary approaches were used here to identify novel target genes regulated by this TF complex during erythroid development.

In the first approach, short interfering RNAs (siRNAs) were transfected into the K562 cell line to knockdown transiently each of five TFs found in this complex. For these five TFs, a knockdown efficiency of at least 70% was confirmed at the mRNA and protein level within 48 hours after transfection. The biological consequences of these knockdowns were studied using Affymetrix GeneChips in order to identify gene expression changes of downstream targets. In the second approach, chromatin immunoprecipitation (ChIP) was performed for the five TFs of the complex in the K562 cell line and the resultant ChIP material was hybridised to a human transcription factor promoter microarray. A number of novel target genes for the SCL erythroid complex were identified and verified independently using both approaches. The data presented in this thesis revealed that members of the SCL-containing erythroid complex are involved in auto-regulation and regulate genes which have key roles in haematopoiesis or control chromatin structure and function. These findings demonstrate that the expression of this TF complex is tightly controlled and point to an important role for it in orchestrating fundamental biological processes which have profound effects on gene expression in erythroid development.