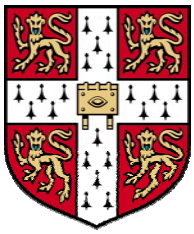
Identification and Characterisation of Regulatory Elements on Human Chromosome 20q12-13.2

Pelin Akan

A thesis submitted in partial fulfilment of the requirements of University of Cambridge for the degree of Doctor of Philosophy



Clare Hall, University of Cambridge

September 2006

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration. The dissertation does not exceed the length limit set by the Biology Degree Committee

ABSTRACT

A nearly-finished sequence of the human genome was published in 2004 (IHGSC, 2004). The task ahead is now to complete the structural and functional annotation of the genome. Although much progress is being made in annotating the coding part of the genome, identification of regulatory regions remains a challenge in current genomics. The aim of this thesis is to identify and characterise promoters and other regulatory elements in a 10 Mb region on human chromosome 20q12-13.2. This region was chosen because of (i) its biological importance, as it is associated with a number of medical conditions such as type II diabetes and obesity (ii) the availability of a detailed transcription map of the region, which is particularly valuable for the experimental and computational analyses carried out in this study.

Firstly, I describe the identification and characterization of core promoter elements using computational methods. Here, promoters are studied at the sequence and structural level in an attempt to discover novel signals for promoter identification *in silico*. Candidate promoters are also correlated to genomic features and expression data from two cell lines, HeLa S3 and NTERA-2 clone D1.

In the subsequent chapter, I describe the systematic validation of annotated (candidate) promoters using dual luciferase reporter assays in the two cell lines, HeLa S3 and NTERA-2 clone D1. Analyses include the assessment of promoter activities in synergy with the SV40 enhancer. The differential response of core promoters to the enhancer is then associated with the presence of transcription factor binding motifs predicted by a transcription factor binding motif prediction program (MAPPER).

In the final results chapter, I present my findings in chromatin immunoprecipitation (ChIP) studies carried out on an in-house spotted DNA array. This array was constructed with 2kb overlapping plasmid clones spanning 3.5 Mb of the investigated

region (gene rich segment; 72 protein-coding genes). ChIP analyses (both cell lines, each in triplicate) included 7 antibodies against modified histones, one antibody against RNA polymerase II and one antibody against the transcription factor CTCF to investigate the histone code and transcriptional activity of the region in two cell lines. Additionally, the sequence features of potential distal regulatory elements are studied *in silico* to recognize any common features of such elements.

Acknowledgements

I would like to thank my supervisor Panos Deloukas, who always believed in me and helped me in every way throughout my PhD. I feel truly privileged to work with him.

I would also like to thank Rhian Gwilliam for her help with the construction of the array and guiding me during my years at Sanger, Phillippe Couttet for his invaluable advice, Team 66, Team 67, and Team 62 for allowing me use their hybridisation station. I would also like to show my appreciation for both members of my thesis committee, David Vetrie and Richard Clarkson.

My friends, Paula, Lawrence, Samuel (dear Sammy), Mark (the funniest guy ever, can't live without his jokes) and Joel! You were always there for me, and I wouldn't have finished if it wasn't for you. Paula, I can't find words to describe your loyal friendship, Mark is the luckiest guy ever! Lawrence, our long conversations, complicated feelings and past, tennis, psychology, philosophy… You will always be a part of me. I love you! Sammy, oh my dear Sammy, you are my hero, your integrity, strength, warm personality, there is nothing in the world you cannot have… Mark, you are the most sane among all of us, your acute comments always fine-tuned my thoughts sometimes even my actions! And your *Best of Pink Floyd* album is on its way! Joel, a bird who adores freedom, I miss you very much, I am still waiting for the credits to finish before leaving the movie theatre sometimes! When are we going to cook together again and watch your wonderful pictures around the world?

Friedel, your common-sense, great taste in music and everything, really. When I think about you, I feel like it is all worth it since I have such a friend like you. Please send more pictures of cutest little Karl Kasimir!

And more, Caroline (my gorgeous blonde and smart friend with a beautiful personality, ok, I think I just described a perfect girl-friend), Amina, Richard, Rhian… I cannot thank you enough…

My sweetheart Nils Jean Nikolaj Martin! First my dear friend then my love. I feel like you are my other half, I don't think anyone can understand me as the way you do. You brighten my life, my heart. I now understood why I had taken such painful decisions in the past, it was all for you. I adore your good heart. Seni çok seviyorum.

And my brother, Ozgun, and dad, without your support and confidence in me, I cannot get through my most difficult times during these years. Ozgun, you always said right things to me, your thoughts always amazed me, my most talented brother, I love you. Dad, you were not only my father but my friend who always listened to me and believed in me and gave me strength…

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PCR	Polymerase Chain Reaction
E. coli	Escherichia Coli
DD	Double distilled
EDTA	Ethylene diamine tetraacetic acid
TE	Tris – EDTA
LB	Luria Bertani
CTD	C-terminal domain
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
DMEM	Dulbecco's Modified Eagle Medium
FBS	Foetal Bovine Serum
IVT	In Vitro Translation
CMV	Cytomegalovirus
SV40	Simian Virus 40
TSS	Transcription Start Site
CpG	A cytosine nucleotide immediately followed by a guanine nucleotide on DNA sequence
FirstEF	First Exon Finder

cDNA	complementary DNA
EST	Expressed Sequence Tag
ORF	Open Reading Frame
D. melanogaster	Drosophila melanogaster
OMIM	Online Mendelian Inheritance in Man Database
RT	Representative Transcript
AT	Alternative Transcript
K	lysine
Ac	Acetylated
Me	Methylated
H3K4me	mono-methylated lysine 4 of histone H3
H3K4me2	di-methylated lysine 4 of histone H3
H3K4me3	tri-methylated lysine 4 of histone H3
H3K9me2	di-methylated lysine 9 of histone H3
H3K27me3	tri-methylated lysine 27 of histone H3
H3Ac	Acetylated lysines 9 and 14 of histone H3
H4Ac	Acetylated lysines 5,8,12 and 16 of histone H4
PolII	RNA polymerase II
CCTF	CCCTC-binding protein
WTSI	Wellcome Trust Sanger Institute
FA	Formaldehyde