In silico and functional analyses of the iron metabolism pathway

Natalie Judith Strickland

Dissertation presented for the degree of Doctor of Philosophy (PhD)

in Genetics

at

Stellenbosch University

Supervisor: Monique G. Zaahl (Department of Genetics, Faculty of Science)

Co-supervisor: Ann Louw (Department of Biochemistry, Faculty of Science)

March 2013

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Summary

Iron is an essential micronutrient that is an absolute requirement for correct cellular function in all eukaryotic organisms. However, ferrous iron has the ability to catalyze the formation of potentially toxic reactive oxygen species and regulation of iron metabolism is therefore of critical importance. Currently, there is little known about the co-ordinated regulation of the plethora of genes coding for proteins involved in this biochemical pathway, with the exception of the well characterized post-transcriptional IRE/IRP system. Regulation of gene expression in eukaryotic organisms is a highly intricate process. Transcriptional regulation is the first step and is controlled by the presence of specific *cis*-regulatory regions (*cis*-motifs), residing within the promoter region of genes, and the functional interactions between the products of specific regulatory genes (transcription factors) and these *cis*-motifs. A combinatorial bioinformatic and functional approach was designed and utilized in this study for the analysis of the promoter architecture of genes of the iron metabolic pathway.

The upstream non-coding region (~2 kb) of 18 genes (*ACO1*, *CP*, *CYBRD1*, *FTH1*, *FTL*, *HAMP*, *HEPH*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF*, *SLC11A2*, *SLC40A1*, *STEAP3*, *TF*, *TFRC*, *TFR2*), known to be involved in the iron metabolism pathway, was subjected to computational analyses to identify regions of conserved nucleotide identity utilizing specific software tools.

A subset of nine (*CYBRD1*, *FTH1*, *HAMP*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF*, *TFRC*) of the genes were found to contain a genomic region that demonstrated over 75% sequence identity between the genes of interest. This conserved region (CR) is approximately 140 bp in size and was identified in each of the promoters of the nine genes. The CR was subjected to further detailed examination with comparative algorithms from different software for motif detection. Four specific *cis*-motifs were discovered within the CR, which were found to be in the same genomic position and orientation in each of the CR-containing genes. *In silico* prediction of putative transcription factor binding sites revealed the presence of numerous binding motifs of interest that could credibly be associated with a biological function in this pathway, including a novel MTF-1 binding site in five of the genes of interest.

Validation of the bioinformatic predictions was performed in order to fully assess the relevance of the results in an *in vitro* setting. Luciferase reporter constructs for the nine CR-containing genes were designed containing: 1) the 2 kb promoter, 2) a 1.86 kb promoter with the CR removed and 3) the 140 bp CR element. The expression levels of these three reporter gene constructs were monitored with a dual-luciferase reporter assay under standard culture conditions and simulated iron overload conditions in two different mammalian cell lines. Results of the luciferase assays indicate that the CR promoter constructs displayed statistically significant variation in expression values when compared to the untreated control constructs. Further, the CR appears to mediate transcriptional regulatory effects *via* an iron-independent mechanism. It is therefore apparent that the bioinformatic predictions were shown to be functionally relevant in this study and warrant further investigation.

Results of these experiments represent a unique and comprehensive overview of novel transcriptional control elements of the iron metabolic pathway. The findings of this study strengthen the hypothesis that genes with similar promoter architecture, and involved in a common pathway, may be co-regulated. In addition, the combinatorial strategy employed in this study has applications in alternate pathways, and could serve as a refined approach for the prediction and study of regulatory targets in non-coding genomic DNA.

Opsomming

Yster is ‘n noodsaaklike mikrovoedingstof wat ‘n vereiste is vir korrekte sellulêre funksie in alle eukariotiese organismes. Yster (II) of Fe^{2+} het egter die vermoë om die vorming van potensiële toksies reaktiewe suurstof spesies te kataliseer en dus is die regulasie van die yster metaboliese padweg van kardinale belang. Tans is daar beperkte inligting oor koördineerde regulasie van die gene, en dus proteïene waarvoor dit kodeer, in hierdie padweg. ‘n Uitsondering is die goed gekarakteriseerde na-transkripsionele “IRE/IRP” sisteem. Regulasie van geenuitdrukking in eukariotiese organismes is ‘n ingewikkelde proses. Transkripsionele regulasie is die eerste stap en word beheer deur die teenwoordigheid van spesefieke *cis*-regulatoriese elemente (*cis*-motiewe), geleë in die promotor area van gene, en die funksionele interaksies wat plaasvind tussen die produkte van spesifieke regulatoriese faktore (of transkripsie faktore) en hierdie *cis*-motiewe. ‘n Gekombineerde bioinformatiese en funksionele benadering was ontwerp en daarna gebruik in dié studie vir die analise van die promotor argitektuur van gene wat ‘n rol speel in die yster metaboliese padweg.

Die stroomop nie-koderende streek (~2 kb) van 18 gene (*ACO1, CP, CYBRD1, FTH1, FTL, HAMP, HEPH, HFE, HFE2, HMOX1, IREB2, LTF, SCL11A2, SLC40A1, STEAP3, TF, TFRC, TFR2*), bekend vir hul betrokkenheid in die yster metabolisme padweg, was bloodgestel aan bioinformatiese analises om die streke van konservering te identifiseer met die hulp van spesifieke sagteware.

Slegs nege (*CYBRD1, FTH1, HAMP, HFE, HFE2, HMOX1, IREB2, LTF, TFRC*) van die geanaliseerde gene het ‘n genomiese area bevat wat meer as 75% konservering getoon het. Hierdie gekonserveerde area (GA) is 140 bp in lengte en is geïdentifiseer in elk van die promotors van die nege gene. Die GA was verder bloodgestel aan analises, met die hulp van spesifieke sgateware, wat gebruik maak van vergelykende algoritmes vir motief karakterisering. Vier *cis*-motiewe is identifiseer en kom voor in dieselfde volgorde en oriëntasie in elk van die gene. *In silico* voorspelling van moontlike transkripsie faktor bindingsplekke het getoon dat daar talle bindingsmotiewe van belang teenwoordig is en dié

motiewe kan gekoppel word aan biologiese funksies in hierdie padweg, insluitend ‘n nuwe MTF-1 bindingsplek in vyf van die gene van belang.

Die bioinformatiese analises is verder gevalideer om die relevansie van die resultate in ‘n *in vitro* sisteem ten volle te assesseer. Luciferase rapporteerder konstrukte is vir die nege gene ontwerp wat die volgende bevat: 1) die 2 kb promotor, 2) ‘n 1.86 kb promotor met die GA verwyder en 3) die 140 bp GA element. Die vlakke van uitdrukking van hierdie drie rapporteerder konstrukte was genormaliseer met ‘n dubbele-luciferase rapporteerder assay onder standaard kultuur kondisies en gesimuleerde ysteroorlading kondisies in twee verskillende soogdier sellyne. Resultate van die luciferase assays dui aan dat die GA promotor konstrukte statisties betekenisvolle variasie toon in vergelyking met die onbehandelde kontrole konstrukte. Verder, die GA blyk om transkipsionele regulatoriese effekte te medieer *via* ‘n yster-onafhanklike meganisme. Dit blyk duidelik dat die bioinformatiese voorspellings ook funksioneel getoon kon word en was dus relevant in dié studie en regverdig verdere ondersoek.

Hierdie eksperimentele ontwerp verteenwoordig ‘n unieke en omvattende oorsig van nuwe transkripsionele beheer elemente wat voorkom in die yster metaboliese padweg. Die resultate van dié studie versterk die hipotese dat gene met soortgelyke promotor argitektuur en wat betrokke is in ‘n gemene padweg saam gereguleer kan word. Daarbenewens, die gekombineerde strategie wat in hierdie studie gebruik is het toepassings in alternatiewe metaboliese paaie, en kan dien as ‘n verfynde benadering vir die voorspelling en studie van die regulerende teikens in nie-koderende genomiese DNS.

Acknowledgements

I would like to express my sincere thanks to the following institutions and individuals without whom this study would not have been possible:

The National Research Foundation (Thuthuka) and the University of Stellenbosch for providing financial support.

The University of Stellenbosch and the Department of Genetics for providing the infrastructure and facilities utilized in the completion of this study.

My supervisor, Dr MG Zaahl, for affording me the opportunity to be part of her research group and for critical reading of this manuscript.

My co-supervisor, Prof A Louw, for critical reading of this manuscript.

My fellow researchers in Lab 242, past and present, for providing constant amusement, a wonderful working environment and invaluable advice.

Miss Marika Bosman for always being there to talk to.

My father and mother for providing financial assistance, emotional support and encouragement throughout the writing of this thesis and my years at the University of Stellenbosch. Thank you for making my dreams a reality.

Matt, for your constant support, encouragement, patience and love.

All of my friends, for your support and patience with my long absences during the writing of this thesis.

Table of Contents

CHAPTER TWO: Materials and Methods

CHAPTER FIVE: Conclusion and Future Directions

CHAPTER SIX: References

APPENDICES

List of Figures

CHAPTER ONE: Literature Review

CHAPTER TWO: Material and Methods

CHAPTER THREE: *In silico* analyses of promoter regulatory targets in the iron metabolism pathway

CHAPTER FOUR: Functional analyses of promoter regulatory targets in the iron metabolism pathway

APPENDIX THREE: pGL4 Vector Maps

APPENDIX FOUR: *In Silico* Promoter Analysis Supplementary Data

APPENDIX FIVE: Functional Promoter Analysis Supplementary Data

List of Tables

CHAPTER TWO: Materials and Methods

CHAPTER THREE: *In silico* analyses of promoter regulatory targets in the iron metabolism pathway

CHAPTER FOUR: Functional analyses of promoter regulatory targets in the iron metabolism pathway

APPENDIX FIVE: Functional Promoter Analysis Supplementary Data

List of Abbreviations and Symbols

α	alpha
β	beta
~	approximately
°C	degrees Celsius
>	greater than
≥	greater than or equal to
$^{\bullet}OH$	hydroxyl radical
<	less than
µg/ml	microgram per millilitre
µg/L	microgram per litre
µl	microlitre
µM	micromolar
%	percentage
+	plus
±	plus or minus
®	registered trademark
x	times
3’	three prime
5’	five prime
A	adenosine
ACO1	aconitase 1 gene
ACD	anaemia of chronic disease
ALAS	δ-aminolevulinic acid synthase
Amp^r	ampicillin resistance gene
ANOVA	analysis of variance
AP-1	activator protein 1
AP-2	activator protein 2

apo-TF	apo-transferrin
ATG	transcription initiation codon
ATP	adenosine triphosphate
ATRAF	atransferrinemia
bp	basepair
BC	before Christ
BMP	bone morphogenetic protein
BMPR	bone morphogenetic protein receptors
BSA	bovine serum albumin
bZIP	basic leucine zipper
C	cytosine
$C_{10}H_{16}N_2O_8$	ethylenediaminetetraacetic acid
$C_2H_4O_2$	acetic acid
CBP	CREB1-binding protein
cDNA	complementary DNA
C/EBP	CCAAT enhancer binding protein
C/EBPα	CCAAT enhancer binding protein alpha
cells/ml	cells per millilitre
ChIP	chromatin immunoprecipitation
cis	lying on the same plane
cm^2	centimetre squared
c-MYC	myelocytomatosis viral oncogene c
CNS	central nervous system
CO_2	carbon dioxide
COS-1	African green monkey kidney cells
CP	ceruloplasmin gene
CP	ceruloplasmin
CR	conserved region
CRM	*cis*-regulatory module
CYBRD1	cytochrome b reductase 1 gene

CYBRD1	cytochrome b reductase 1
Cybrd1	murine *CYBRD1* homologue
Da	Dalton
dATP	2'-deoxy-adenosine-5'-triphosphate
DCT1	divalent cation transporter 1
dCTP	2'-deoxy-cytidine-5'-triphosphate
DCYTB	duodenal cytochrome b
dGTP	2'-deoxy-guanosine-5'-triphosphate
dH_2O	distilled water
DLR	dual-luciferase reporter
DMEM	Dulbecco's Modified Eagles Medium
DMT1	divalent metal transporter 1
DNA	deoxyribonucleic acid
dNTP	2'-deoxy-nucleotide-5'-triphosphate
dTTP	2'-deoxy-thymidine-5'-triphosphate
E-value	expect value
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ENCODE	Encyclopedia of DNA Elements
EPD	Eukaryotic Promoter Database
ER	endoplasmic reticulum
et al.	and others
EtBr	ethidium bromide
FAC	ferric ammonium citrate
FCS	Fetal Calf Serum
Fe	iron
Fe/kg	iron per kilogram
Fe^{2+}	ferrous iron

Fe^{3+}	ferric iron
firefly	*Photinus pyralis*
FISH	fluorescent *in situ* hybridisation
FOXM1	forkhead box protein M1
FOXO3a	forkhead box protein O3a
FPN1	ferroportin 1
FTH1	ferritin, heavy polypeptide 1 gene
FTH1	ferritin, heavy polypeptide 1 protein
FTL	ferritin, light polypeptide gene
FXN	frataxin protein
g	gram
g	gravity
G	guanosine
GATA-1	GATA-binding protein 1
GATA-2	GATA-binding protein 2
GATA-3	GATA-binding protein 3
GFP	green fluorescent protein
g/L	gram per litre
GPI	glycol-phospho-inositol
GR	glucocorticoid receptor
GTf	general transcription factor
GWAS	genome-wide association study
H^+	hydrogen ion
H_2O_2	hydrogen peroxide
HAMP	hepcidin antimicrobial peptide gene
HCl	hydrochloric acid
HCP1	haem carrier protein 1
HepG2	human hepatocarcinoma liver cells
HEPH	hephaestin gene
HEPH	hephaestin

Heph	murine homologue of hephaestin
HFE	haemochromatosis gene
HFE	haemochromatosis protein
HFE2	hemojuvelin gene
hgDNA	human genomic DNA
HH	hereditary haemochromatosis
HIV-1	human immunodeficiency virus-1
HIV-2	human immunodeficiency virus-2
HJV	hemojuvelin
HLH	helix-loop-helix
HMOX1	haem oxygenase 1 gene
HMOX1	haem oxygenase 1
hmox1	rat homologue of human haem oxygenase 1 gene
HNF-1α	hepatocyte nuclear factor 1 alpha
HNF-1β	hepatocyte nuclear factor 1 beta
HNF-3α	hepatocyte nuclear factor 3 alpha
HNF-3β	hepatocyte nuclear factor 3 beta
HNF-4α	hepatocyte nuclear factor 4 alpha
IDA	iron deficiency anaemia
IDT	Integrated DNA Technologies
INF-γ	interferon gamma
in silico	refers to research conducted using computers only
in vitro	Latin phrase meaning “outside the body” or outside a living organism
in vivo	Latin phrase meaning “in the body” or within a living organism
IRE	iron responsive element
IREG1	iron-regulated transporter 1
IREB2	iron regulatory binding protein 2 gene
IRP1	iron regulatory protein 1
IRP2	iron regulatory protein 2

kb	kilo basepair
kcal	kilocalorie
kDa	kilo Dalton
LB	Luria-Bertani
LIP	labile iron pool
LPS	lipopolysaccharide
LTF	lactotransferrin gene
LTF	lactotransferrin protein
luc	luciferase gene
M	molar
MCR	multiple cloning region
MEM	Minimum Essential Medium
MEF2A	myocyte-specific enhancer factor 2A
mg	milligram
mg/day	milligram per day
mg Fe/day	milligram iron per day
mg/L	milligram per litre
mg/ml	milligram per millilitre
$MgCl_2$	magnesium chloride
MHC	major histocompatibility complex
ml	millilitre
mM	millimolar
MPS	mononuclear phagocytic system
MRE	metal responsive element
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MTF-1	metal transcription factor 1
MTP1	metal transporter 1

NADPH	nicotinamide adenine dinucleotide phosphate
ng	nanogram
NHGRI	National Human Genome Research Institute
NKX2-5	NK 2 homeobox 5
NKX3-1	NK 3 homeobox 1
nm	nanometre
NNPP	Neural Network Promoter Prediction
NO	nitric oxide
NRAMP2	natural resistance-associated macrophage protein 2
O_2	oxygen
O_2^-	superoxide radical
O_2^{2-}	peroxide ion
OH^-	hydroxyl anion
p	probability value
p	short arm of chromosome
PAX2	paired box 2
PAX4	paired box 4
PAX5	paired box 5
PAX6	paired box 6
PAX8	paired box 8
PCR	polymerase chain reaction
pH	potential of hydrogen
pmol	picomole
PolII	vertebrate RNA polymerase II
PPAR	peroxisome proliferator-activated receptor
PPARγ	peroxisome proliferator-activated receptor gamma
PPRE	peroxisome proliferator hormone response element
PWM	position weight matrix
q	long arm of chromosome

RE	restriction endonuclease
Renilla	*Renilla reniformis*
RGD	arginine-glycine-aspartic acid motif
RGM	repulsive guidance molecule
RLU	relative luciferase unit
ROS	reactive oxygen species
RNA	ribonucleic acid
rpm	revolutions per minute
RXR-α	retinoid X receptor alpha
RXR-β	retinoid X receptor beta
RXR-γ	retinoid X receptor gamma
S	sulphur
SAP	shrimp alkaline phosphatase
SD	standard deviation
SE	standard error
SEM	standard error of the mean
sHJV	soluble hemojuvelin
SINE	Short Interspersed Nuclear Element
sla	sex-linked anaemia
SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporter), member 2 gene
SLC11A3	solute carrier family 11 (proton-coupled divalent metal ion transporter) member 3
SLC40A1	solute carrier family 40 (iron-regulated transporter) member 1 gene
SLC40A1	solute carrier family 40 (iron-regulated transporter) member 1
SLC46A1	solute carrier family 46 (folate transporter), member 1
SMAD-1	mothers against decapentaplegic homolog 1
SMAD-3	mothers against decapentaplegic homolog 3
SMAD-4	mothers against decapentaplegic homolog 4
SNP	single nucleotide polymorphism
SOFeX	Southern Ocean Iron Experiment

SP1	specificity protein 1
STAMP1	six transmembrane protein of prostate 1 gene
STAT5A	signal transducer and activator of transcription 5A
STAT5B	signal transducer and activator of transcription 5B
STEAP3	six-transmembrane epithelial antigen of the prostate 3 gene
STEAP3	six-transmembrane epithelial antigen of the prostate 3
sTF	serum transferrin
T	thymidine
T_a	annealing temperature
TAF	TATA-binding protein associated factor
Taq	*Thermus aquaticus*
TAE	tris-acetate-EDTA
TBI	transferrin-bound iron
TBP	TATA-binding protein
TF	transferrin gene
TF	transferrin
Tf	transcription factor
TfBS	transcription factor binding site
TFRs	transferrin receptors
TFR1	transferrin receptor 1
TFR2	transferrin receptor 2 gene
TFR2	transferrin receptor 2
TFRC	transferrin receptor 1 gene
™	trademark
T_m	melting temperature
trans	lying on opposite planes
tRNA	transfer RNA
TSS	transcription start site

U	unit
U/ml	units per millilitre
USF	upstream transcription factor
USF2	upstream transcription factor 2
UTR	untranslated region
V	volt
v	version
via	by way of
vice versa	Latin phrase meaning “the other way round”
vs	versus
v/v	volume per volume
WHO	World Health Organisation
w/v	weight per volume
YY1	ying yang 1

CHAPTER ONE

Literature Review

1.1 IRON AND IRON METABOLISM

1.1.1 Iron: Historical Perspective and Global Importance

Iron (chemical symbol; Fe) is a chemical element which forms part of the first transition series on the periodic table. Of all the metal elements, iron has the best described history. With regard to human civilization, iron has played a significant role in the advancement of modern societies. The Iron Age of Man which began around 1200 BC, following the earlier Stone and Bronze Ages, saw the development of the ability to heat and forge iron. This led to the manufacture of weapons and cutting instruments, which concurrently with changes in agricultural practices and religious beliefs led to the development of modern civilizations in Europe, Asia and Africa (Wertime and Muhly, 1980). In contrast to the Stone and Bronze Ages, few artefacts of this time remain due to the instability of iron in conjunction with water and oxygen.

Iron is the fourth most abundant chemical element in the Earth's crust, comprising approximately 4.7%, in the form of iron oxide minerals such as hematite, siderite and magnetite. Both the inner and outer core of the Earth are considered to be made of an iron-nickel alloy, constituting 35% of the Earth's mass (Herndon, 2005). The frequent occurrence of iron in rocky planets, such as Earth, is due to the last nuclear fusion reaction in the collapse of high-mass stars (supernovas) that scatters a radioactive precursor of iron abundantly into space (Burrows, 2000). In contrast to the terrestrial abundance of iron, only trace amounts are found in the world's oceans (3×10^{-3} mg/L) due to the insolubility of iron in salt water (Mason and Moore, 1982). Historically, iron was delivered to pelagic systems *via* dust storms (Aeolian processes) from arid deserts carrying 3 – 5% iron. In recent decades this mechanism of iron deposition has decreased by almost 25% due to climatic changes (Cassar *et al.*, 2007). These low ocean iron levels are thought to limit the production of phytoplankton, thereby contributing to the relatively high atmospheric carbon dioxide (CO_2) levels of the present day. This phenomenon is particularly well studied in the Southern Ocean, which is known to exert a major control over atmospheric CO_2 levels (Bishop *et al.*, 2004). In 2002, The Southern Ocean Iron Experiment (SOFeX) was launched, during which 1.7 tonnes of iron sulphate was added to the seawater (Coale *et al.*, 2004). The data collected from this mammoth experiment demonstrated that for each atom of iron added,

10 000 to 100 000 atoms of carbon were able to be pulled out of the atmosphere *via* large phytoplankton blooms. To date, there have been upwards of ten different `iron fertilization' experiments in many of the world's oceans, with some researchers estimating that 15% of the atmospheric CO_2 build up (contributing to global warming) could be removed (Boyd *et al.*, 2000; Boyd *et al.*, 2007).

In addition to the effects of iron on a global scale, it also operates as a micronutrient and is an absolute requirement for a host of cellular functions (replication, differentiation and ultimately, survival) in almost all animals, plants and microorganisms. Members of the lactic acid bacteria, including the *Lactobacillus* and *Lactococcus* family, as well as some *Streptococcus* species are the only known exceptions with regard to iron requirements. These microorganisms utilize other transition metals, residing next to iron in the periodic table, such as manganese and cobalt to fulfil the essential role that iron plays in all other organisms (Crichton, 2009). Early human civilizations attributed therapeutic properties to iron, with the ancient Greeks administering iron to injured soldiers to improve muscle weakness caused by anaemia as a result of blood loss (Beard *et al.*, 1996). In the 16th century, physicians prescribed iron salts to women suffering from chlorosis (an old term used to describe anaemia) often as a consequence of protein deficiency or pregnancy (Poskitt, 2003). Between 1832 and 1843, chlorosis was further defined by low numbers of red blood cells and reduced plasma iron levels (McCay, 1973 as reviewed by Poskitt, 2003). The essential requirements of iron in nutrition were first described by Boussingault in 1872 (as reviewed by Beard *et al.*, 1996). The basics of iron metabolism, particularly with regard to disorders of iron metabolism, as well as the nutritional importance of this mineral have been recognized for centuries. However, it is only in recent decades that the molecular underpinnings of this indispensable biochemical pathway have begun to be elucidated.

1.1.2 Chemical and Biochemical Properties of Iron

Iron occupies a position in the middle of the elements in the first transition series in the periodic table. The chemical properties of iron allow for it to exist in various oxidative states (from –II to +VI) with the predominate forms being Fe(II) (or Fe^{2+}, ferrous iron) and Fe(III) (or Fe^{3+}, ferric iron). The bioavailability of iron is limited by the fact that soluble Fe^{2+} is readily oxidized under aerobic conditions to Fe^{3+}, which is insoluble in water and at the physiological pH of 7 (Papanikolaou and Pantopolous, 2005). The ability of iron to

interconvert between oxidation states is the primary mechanism through which reversible iron-ligand binding takes place, as well as the manner in which electron transport occurs. The binding of iron to specific biological ligands (usually oxygen for Fe^{3+} and nitrogen or sulphur atoms for Fe^{2+}) is the only way in which significant concentrations of water-soluble Fe^{3+} can be attained and made available for use in metabolic functions (Crichton, 2009). The extreme range of the Fe^{2+}/Fe^{3+} redox potential of iron makes it uniquely suited to perform a variety of cellular functions that encompass almost the entire redox range from -0.5 V to +0.6 V. This redox potential can be finely controlled by the binding of iron to different ligands which influence the electron spin speed of the complexes and thus the chemical reactivity of iron (Crichton, 2009). The biological activity of many enzymes is therefore impeded during times of tissue iron deficiency. Iron is essential for a multitude of important biochemical activities such as oxygen transport by haemoglobin, DNA synthesis and electron transport by the cytochrome proteins of the respiratory chain (Aisen *et al.*, 2001).

1.1.3 Toxicity of Iron

The variable redox potential of iron which makes it uniquely suited to perform the complex biological functions required, also means that this metal is potentially toxic to cellular components. The ability of iron to readily interconvert between oxidation states under aerobic conditions results in the creation of hazardous free radicals by means of the chemistry of the Fenton and Haber-Weiss reactions (Fenton, 1984 as reviewed by Papanikolaou and Pantopolous, 2005). During the Fenton process, trace amounts of Fe^{2+} and hydrogen peroxide (H_2O_2, formed by the protonation of the peroxide ion, O_2^{2-}, at physiological pH) result in hydroxyl radical ($^{\bullet}OH$) formation (Figure 1.1 A). The superoxide radical (O_2^{-}) can then reduce Fe^{3+} to molecular oxygen (O_2) and Fe^{2+} (Figure 1.1 B). During the sum of this reaction and the Fenton reaction, O_2^{-} and H_2O_2 in the presence of catalytic amounts of iron produce O_2, $^{\bullet}OH$ and the hydroxyl anion (OH^{-}) – the Haber-Weiss reaction (Figure 1.1 C) (Haber and Weiss, 1934 as reviewed by Koppenol, 2001).

A.	$Fe^{2+} + H_2O_2$	→	$Fe^{3+} + {}^{\bullet}OH + OH^-$
B.	$Fe^{3+} + O_2^-$	→	$Fe^{2+} + O_2$
C.	$O_2^- + H_2O_2$	→	$O_2 + {}^{\bullet}OH + OH^-$

Figure 1.1 Fenton and Haber-Weiss reactions.

(**A**) The Fenton reaction. (**B**) The production of molecular oxygen and Fe^{3+} from the superoxide radical *via* an intermediate reaction. (**C**) The Haber-Weiss reaction; the sum of reactions A and B.

The free radicals produced by Fenton and Haber-Weiss reactions; hydroxyl-, superoxide radicals and hydroxyl anions are termed reactive oxygen species (ROS). Increases in steady state ROS levels, above the antioxidant capacity of an individual, result in the development of oxidative stress. Oxidative stress, in turn, leads to the damage of cellular structures and macromolecules and is associated with tissue injury and disease (Galaris and Pantopolous, 2008; Kell, 2009). In order to mitigate this double-edged sword nature of iron, intricate biological mechanisms of iron homeostasis exist within all organisms that utilize iron for metabolic functions.

1.1.4 Iron Distribution and Circulation

In healthy, adult humans, the total body iron content is approximately 3 – 5 g, constituting roughly 45 mg Fe/kg of body weight in men and a slightly lower level of 35 mg Fe/kg in women (Andrews, 1999). The vast majority of iron (30 mg Fe/kg or 60-70%) is bound to the haem prosthetic group of haemoglobin, the circulating oxygen transport protein, found within the red blood cells (erythrocytes). A further 4 mg Fe/kg (10%) of iron is present in muscles in the form of myoglobin and the remainder of the total body iron, 10 – 12 mg Fe/kg (20 to 30%), which is surplus to immediate cellular requirements, is stored as ferritin and haemosiderin in the liver, spleen and bone marrow (Conrad *et al.*, 1999). In the plasma, a miniscule fraction of total body iron (0.05 mg Fe/kg or ± 0.1%) is transported by forming a complex with transferrin (TF), an iron transport protein that contains two iron binding sites (Emerit *et al.*, 2001). TF-bound iron (TBI) supplies the majority of iron required for the production of haemoglobin by erythroid precursor cells, as well as delivering iron to a host

of other tissues and cells and delivers the maternal supply of iron utilized at high levels during pregnancy. The vast majority of iron required by the erythrocytes (20 mg per day) for erythropoiesis, comes from the destruction of old red blood cells by the mononuclear phagocytic system (MPS). Mature erythrocytes, produced in the bone marrow, circulate in the blood plasma for around 120 days, where after they are taken up by the MPS of the reticuloendothelial system in cells of the spleen and liver (Deiss, 1983). Macrophage iron recycling releases iron from haem molecules into the circulation where it once again binds to TF. The distribution of iron in adults is illustrated in Figure 1.2.

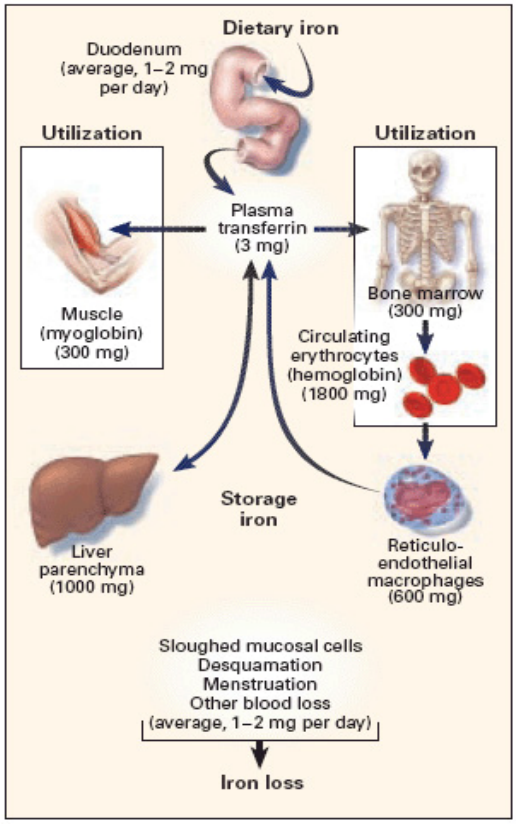


Figure 1.2 Distribution of iron in adults.

In the balanced state, 1 – 2 mg of iron enters and leaves the body each day. Dietary iron is absorbed by duodenal enterocytes. It circulates in the plasma bound to transferrin. Most of the iron in the body is incorporated into haemoglobin in erythroid precursors and mature red cells. Approximately 10 to 15% is present in muscle fibers (in myoglobin) and other tissues (in enzymes and cytochromes). Iron is stored in parenchymal cells of the liver and reticuloendothelial macrophages. These macrophages provide most of the usable iron by degrading haemoglobin in senescent erythrocytes and reloading ferric iron onto transferrin for delivery to cells.

1.1.5 Dietary Iron Uptake

The human body does not possess any regulated mechanism for the excretion of excess iron. Small amounts of body iron stores are lost during the sloughing of mucosal and skin cells (approximately 1 mg/day). Due to regular blood loss during menstruation and childbirth, women lose additional iron from the high concentrations contained in the haemoglobins (Andrews *et al*., 1999). However, this loss of iron is around one tenth of that of other mammals (per kg body weight). This daily iron loss is compensated for by uptake of a similar amount of iron from the diet.

As a consequence of the potential biohazardous effects of iron overload (refer to Section 1.1.3), the control of iron balance within the body is of paramount importance. This balance is primarily maintained at the level of intestinal iron absorption. An inverse relationship exists between iron absorption and iron stores in various tissues of the body (predominately the liver, spleen and bone marrow): as tissue iron stores decrease iron uptake is increased, and *vice versa*. The rate of erythropoiesis is another important factor driving the uptake of iron. Rapid red cell development results in an increased need for iron and consequently the assimilation of iron is heightened (Andrews, 1999). The principle site of iron absorption is in specialized cells (enterocytes) of the upper regions of the gastrointestinal tract, specifically the duodenum and upper ileum. Other factors such as the bioavailability of iron and its composition, pH and mucosal and luminal factors affect the efficient uptake of iron (Hutchinson *et al.*, 2007).

Typical Western diets contribute around 5 – 6 mg of iron per 1000 kcal consumed, amounting to 12 – 18 mg of iron ingested per day. This dietary iron is comprised chiefly of two forms: haem iron and non-haem iron (Reddy *et al.*, 2000).

Haem Iron

Haem iron is derived from haemoglobin and myoglobin proteins associated with meat intake. Haem is a common prosthetic group composed of protoporphyrin IX and a Fe(II) ion. The reaction whereby Fe(II) is incorporated into protoporphyrin IX, takes place in the mitochondria and is catalyzed by the ferrochelatase enzyme. Haem-derived iron is extremely bioavailable and is easily absorbed however, it accounts for only 10% of dietary iron ingestion (reviewed by Anderson *et al.*, 2005). Despite the scarcity of haem iron, it amounts to nearly 50% of the iron absorbed by the body from a standard diet (Majuri and

Grasbeck, 1987). Haem iron is absorbed as an intact metalloporphyrin involving specific haem-binding sites on the brush border of the intestinal enterocytes. Cooking processes such as frying can significantly reduce the bioavailability of haem-iron by breaking down the structure of the haemoglobin molecule, thereby impeding entry into the intestinal cells (Conrad *et al.*, 1999).

Non-haem Iron

The inorganic form of iron is by far the most abundant form of dietary iron and originates principally from plant dietary components. It constitutes almost 90% of ingested dietary iron daily but conversely, only 2 – 20% of the iron that is absorbed, depending on body iron status and the presence of enhancers or inhibitors (Lombard *et al.*, 1997). The foremost enhancers of non-haem iron absorption are meat and organic acids. Ascorbic acid (vitamin C) is the primary enhancer and is able to reduce Fe^{3+} to soluble Fe^{2+} in the presence of the acidic pH provided by the stomach during digestion (Hutchinson *et al.*, 2007). Other enhancers include lactic, citric, malic and tartaric acids (Conrad *et al.*, 1999).

1.1.5.1 Proteins Involved in Dietary Iron Uptake

The myriad proteins associated with the uptake of iron from the gastrointestinal lumen can be grouped into three distinct roles: the uptake of iron from the intestinal lumen, across the brush-border membrane at the apical pole of the enterocytic cell; the transport of iron within the enterocyte, as well as iron sequestration; and the release of iron from the mucosal cell basolateral membrane into the plasma (Fleming, 2005). Knowledge of the involvement of the proteins involved in these processes has greatly increased in recent years with the advent of modern molecular biology techniques, and many proteins that were previously hypothetically predicted to occur have now been identified. However, the understanding of the physiological role that these proteins perform is still somewhat limited (Miret *et al.*, 2003). The proteins illustrated in Figure 1.3 will be discussed further in this Section.

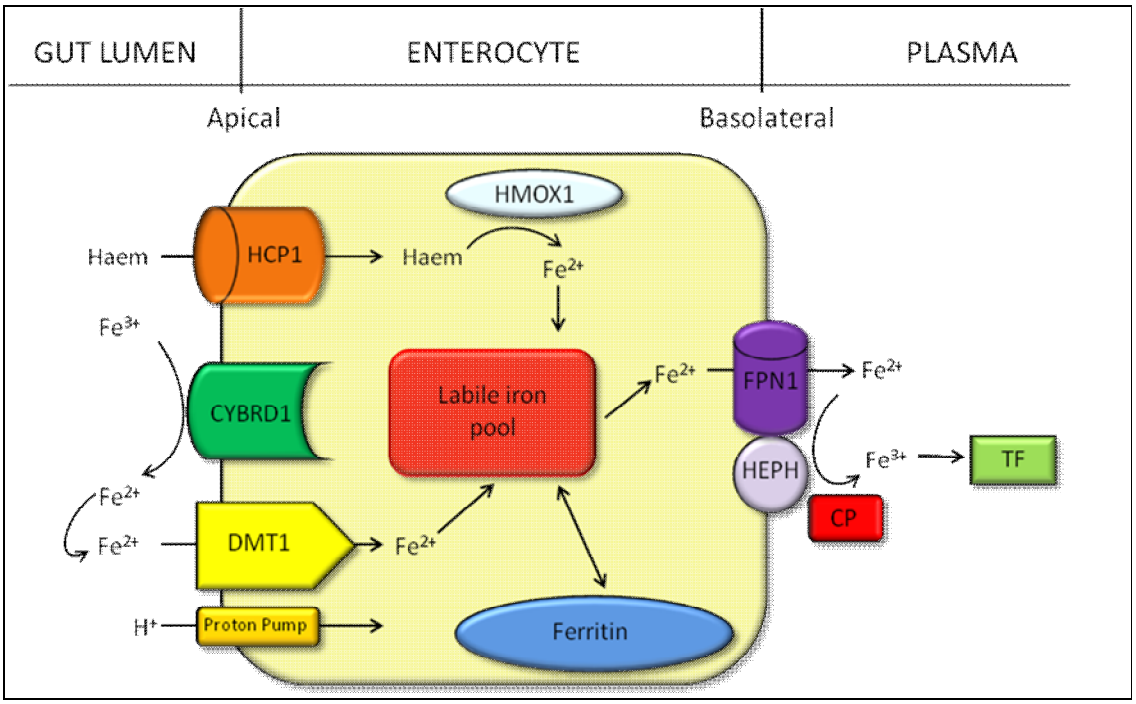


Figure 1.3 Dietary iron uptake.

Haem is hypothesized to be transported from the intestinal lumen *via* HCP1 (currently the protein remains unidentified). Within the enterocyte, Fe^{2+} is released during the degradation of haem by HMOX1. Dietary non-haem iron is reduced by CYBRD1 from Fe^{3+} to Fe^{2+} which is then transported across the apical membrane by DMT1. Intracellular Fe^{2+} is either stored as ferritin or transported out of the cell by FPN1, which is located on the basolateral membrane. Facilitation of iron export takes place with the oxidation of Fe^{2+} to Fe^{3+} by HEPH (membrane-bound) and CP (in the plasma), aiding in the subsequent binding of Fe^{3+} to transferrin.

Abbreviations: CP, ceruloplasmin; CYBRD1, cytochrome b reductase 1; DMT1, divalent metal transporter 1; Fe^{3+}, ferric iron; Fe^{2+}, ferrous iron; FPN1, ferroportin 1; H^+, hydrogen ion; HCP1, haem carrier protein 1; HEPH, hephaestin; HMOX1, haem oxygenase 1; TF, transferrin.

Adapted from Trinder et al. (2002).

Cytochrome b reductase 1 (CYBRD1)

Cytochrome b reductase 1 (CYBRD1), also referred to as duodenal cytochrome b (DCYTB) or duodenal iron-regulated transporter (IREG1), is a brush-border ferrireductase enzyme that is responsible for mediating the reduction of the Fe^{3+} to Fe^{2+} primarily from dietary non-haem iron (McKie *et al.*, 2001). Some studies have suggested that CYBRD1 may utilize ascorbate as an electron donor during the reduction of Fe^{3+} (Oakhill *et al.*, 2008). The protein was first identified in 2000 by McKie and colleagues, by means of a subtractive cloning technique from iron-deficient rats. Further studies conducted on CYBRD1 in rats have shown that

protein levels are increased during iron deficiency and hypoxia, and decreased in response to iron supplementation, indicating the importance that CYBRD1 plays in regulation of iron absorption. Conversely, studies in mice have illustrated that there is little effect on dietary iron uptake when the gene coding for the murine homologue (Cybrd1) of CYBRD1 is disrupted (Frazer *et al.*, 2005; Gunshin *et al.*, 2005). This may indicate that there are other mechanisms responsible for the reduction of dietary non-haem iron that have yet to be identified.

The CYBRD1 protein is 286 amino acids long and 31641 Da in size with six transmembrane domains and four conserved histidine residues. It is transcribed from the *CYBRD1* gene located on chromosome 2q31.1, which is 35.78 kb in size and consists of 4 exons (McKie *et al.*, 2001).

Divalent metal transporter 1 (DMT1)

Following reduction of non-haem dietary iron to Fe^{2+} by CYBRD1, it is then transported across the apical membrane, and into the lumen of the enterocytes (Fleming *et al.*, 1997). The divalent metal transporter 1 (DMT1), also known as divalent cation transporter 1 (DCT1) or natural resistance-associated macrophage protein 2 (NRAMP2) is responsible for the transport of Fe^{2+} as well as the uptake of a broad range of divalent metal ions (manganese, cobalt, cadmium, nickel, vanadium and lead) (Garrick *et al.*, 2003). DMT1 was first identified as the common iron transport enzyme in mammals by positional cloning experiments in rodents with impaired iron transport (Gunshin *et al.*, 1997). The DMT1 transporter is also found in cells other than those of the gastrointestinal system; it plays a role as an endosomal iron transporter in peripheral tissues. The intestinal isoform of DMT1 is characterized by an altered C-terminus and is produced by differential splicing of the DMT1 mRNA. DMT1 is electrogenic, and requires proton co-transport to be able to act as a divalent cation transporter (Gunshin *et al.*, 1997). The low pH provided by gastric acid in the proximal portion of the duodenum provides a proton-rich environment which facilitates this transport. This mechanism provides a plausible explanation for the aberrant iron absorption observed with antacid treatment (Stein *et al.*, 2010).

The DMT1 protein is comprised of 12 helical transmembrane domains and is 568 amino acids in size (Li *et al.*, 2012). It is encoded by the solute carrier family 11 (proton-coupled divalent metal ion transporter), member 2 (*SLC11A2*) gene on chromosome

12q13.12. The *SLC11A2* gene forms part of the SLC superfamily, which includes around 300 genes that encode solute carrier proteins. The SLC genes can be categorized into many smaller families depending on the types of molecules transported by the protein. Each family is indicated by the first number in the gene name (He *et al.*, 2009).

Haem carrier protein 1 (HCP1)

Iron uptake in the form of dietary haem is significant due to the fact that it is more efficiently absorbed than non-haem iron (West and Oates, 2008). However, the mechanism behind haem import is still unidentified to date. For many years, it has been reasonably postulated that haem may be taken up in a receptor-mediated manner (HCP1 in Figure 1.3) but this remains to be shown scientifically. In 2005, Shayeghi and colleagues published a report in the journal Cell, identifying HCP1 (designated as SLC46A1) as the intestinal importer of haem in the enterocytes of mice. The following year, Qui *et al.* (2006) reported that SLC46A1 is actually a transporter of the vitamin folic acid, not haem. They identified a family in which a loss of function mutation in SLC46A1 resulted in hereditary folate malabsorption. This family were shown to have no defects in iron metabolism although they did require folic acid supplementation to maintain proper growth.

There have been descriptions of haem exporters (Krishnamurthy *et al.*, 2004; Quigley *et al.*, 2004), but the elusive haem importer remains unknown. Current scientific theory centres on the haem oxygenase enzyme (HMOX1 – discussed below) as the mechanism behind the liberation of dietary haem iron from protoporphyrin.

Haem oxygenase 1 (HMOX1)

After haem has entered the enterocyte, it is degraded by the enzyme haem oxygenase 1 (HMOX1), located on the surface of the endoplasmic reticulum (ER). The chemical reaction of haem degradation requires NADPH and molecular O_2 and produces bilirubin, carbon monoxide and Fe^{2+} (Stocker, 1990). The Fe^{2+} which is released from haem is either stored as ferritin or enters the circulation across the basolateral membrane. HMOX1 has been shown to play a critical role in the general maintenance of cellular homeostasis due to the fact that excess free haem can lead to the generation of ROS with subsequent cellular and tissue damage (Maines, 2000), in much the same manner as free iron (Jeney *et al.*, 2002).

The human *HMOX1* gene was identified experimentally using a cDNA hybridisation method where rat *hmox1* was hybridized to a human cDNA library that had been treated with hemin (Yoshida *et al.*, 1988). The *HMOX1* gene was later localized to chromosome 22q12 using fluorescent *in situ* hybridisation (FISH) techniques (Kutty *et al.*, 1994). The gene comprises five exons and encodes the 288 amino acid HMOX1 protein. HMOX1 contains a membrane segment composed of hydrophobic amino acids at the C-terminus (Yoshida *et al.*, 1988).

Ferritin

Ferrous iron that enters the intestinal cell joins the so-called "labile iron pool" (LIP) and has one of two fates: either it can be directly exported into the plasma for use elsewhere in the body, or it can be stored within the enterocyte. The proteins responsible for the mucosal sequestration of excess Fe^{2+} are ferritin and, to a lesser extent, haemosiderin. Mucosal cells have an extremely high rate of cellular turnover and iron stored in ferritin is lost when the enterocyte becomes senescent and is sloughed from the intestinal surface (Theil, 1987). Mobilisation of iron from ferritin stores is essential during periods of iron depletion, however, no known mechanism for the movement of Fe^{2+} out of ferritin has been fully elucidated as of yet. It has been hypothesized that ferritin is degraded by lysosomes or proteasomes in order to provide iron for systemic requirements (Aisen *et al.* 2001). Ferritin is present in every cell type in the human body including the leukocytes of the blood where it is termed serum ferritin (Worwood *et al.*, 2008). Serum ferritin is used as a non-invasive diagnostic marker for the estimation of iron stores due to the fact that serum ferritin levels correlate roughly with total body iron stores under normal conditions (Strandberg Pedersen and Morling, 2009).

The ferritin protein was first isolated by a Czech physiologist Vilem Laufberger in 1937 from horse spleen (as reviewed by Wang *et al.*, 2010). Ferritin consists of 24 protein subunits that together, form a hollow spherical molecule which is able to store up to 4500 atoms of iron in a soluble, bioavailable manner. In mammals, heteropolymeric ferritin is made up of two distinct ferritin subunits, known as heavy- (H) and light- (L) chains. H-ferritin is coded for by the ferritin, heavy polypeptide 1 (*FTH1*) gene on chromosome 11q13. L-ferritin is transcribed from the ferritin, light polypeptide (*FTL*) gene in chromosomal position 19q13.33. Both *FTH1* and *FTL* are comprised of four exons but share only approximately 54%

sequence similarity. H-ferritin contains a diiron, ferroxidase centre and is postulated to play a role in the oxidation of Fe^{2+} to Fe^{3+}. L-ferritin is responsible for the nucleation of iron into the ferritin core (Torti and Torti, 2002).

Ferroportin 1 (FPN1)

The iron export protein, ferroportin 1 (FPN1), is the only known mammalian exporter of iron that has been identified to date. It was described, almost simultaneously, by three different research groups (McKie *et al.*, 2000; Donovan *et al.*, 2000; Abboud and Haile, 2000) and designated by the names solute carrier family 40 (iron-regulated transporter) member 1 protein (SLC40A1), solute carrier family 11 (proton-coupled divalent metal ion transporter) member 3 protein (SLC11A3), ferroportin 1 (FPN1), iron-regulated transporter 1 (IREG1) or metal transporter 1 (MTP1). FPN1 is localized to the basolateral membrane of intestinal enterocytes and has 10-12 transmembrane domains. It does not share any sequence homology with DMT1 or any other mammalian proteins and is the only member of the SLC family 40. FPN1 is responsible for mediating the transport of Fe^{2+} across the basolateral membrane (and the membranes of other cell types not involved in dietary iron homeostasis) and into the plasma (Knutson and Wessling-Resnick, 2003).

FPN1 is transcribed from the *SLC40A1* gene which was mapped to chromosome 2q32 using FISH analysis (Haile 2000). *SLC40A1* consists of eight exons and encodes a 571 amino acid protein.

Ceruloplasmin (CP) and Hephaestin (HEPH)

FPN1 works in conjunction with the proteins ceruloplasmin (CP) and hephaestin (HEPH). CP and HEPH are homologous multicopper ferroxidase proteins responsible for the oxidation of Fe^{2+} to Fe^{3+}. CP is the major copper binding protein of the plasma and cation binding sites have been identified adjacent to the mononuclear copper sites, indicating the potential role of CP in iron oxidation (Lindley *et al.*, 1997). It is a blue alpha-2-glycoprotein that is responsible for binding 6 to 7 copper ions per molecule. CP is hypothesized to create ion gradients that favour the export of iron out of cells, thereby assisting in iron export (Harris *et al.* 1999). Both CP and HEPH are thought to aid binding of Fe^{3+} to the plasma iron transporter, transferrin (TF, discussed in Section 1.1.6.1 below) (Hellman and Gitlin, 2002). The role of CP as the primary ferroxidase involved in iron export from intestinal enterocytes

has been contested due to lack of experimental evidence. In patients with aceruloplasminemia (Yoshida *et al.*, 1995), and in CP null mutant mice, iron accumulation occurs primarily in hepatic and pancreatic cells, but no sign of abnormal intestinal iron absorption is observed (Harris *et al*. 1999). Aceruloplasminemia is a disease caused by the complete loss of CP ferroxidase activity and is characterized by the accumulation of iron in the brain (Kono, 2012). CP appears to have more of a physiological role in peripheral tissues and the central nervous system (CNS), as demonstrated by magnetic resonance imaging (MRI) of the brains of individuals with aceruloplasminemia, which show marked iron deposition (Kono and Miyajima, 1997).

The presence of an alternate ferroxidase enzyme was first revealed in mice with X-linked, recessive hypochromic microcytic anaemia (Bannerman and Cooper, 1966; Edwards and Bannerman, 1970). These *sla* (sex-linked anaemia) mice were shown to possess the ability to transport iron from the intestinal lumen, but were unable to transfer iron out of the enterocytes into circulation. Vulpe and colleagues (1999) identified a homologue of murine CP using genome sequencing technologies. In 2002, the human *HEPH* gene was mapped to the X chromosome (Syed *et al.*, 2002). The CP and HEPH polypeptides share 50% amino acid identity with conservation at sites specific for copper binding.

In contrast with CP, HEPH is a membrane-bound enzyme found predominately on the basolateral membrane of intestinal mucosal cells. It is anchored to the membrane by a glycol-phospho-inositol (GPI) anchor at the C-terminus. It represents a fundamental link between copper and iron metabolism in mammals. An alternatively-spliced CP transcript, expressed primarily in astrocyte cells of the brain, is also bound to the cell membrane by a GPI anchor. This form of CP has not been identified in cells of the intestine, liver or spleen (Patel and David, 1997).

The *CP* gene was mapped to the long arm of chromosome 3 (3q21-24) by Southern blot analysis of human-mouse somatic cell hybrids (Naylor *et al*., 1985, Yang *et al*., 1986). *CP* spans a genomic region of approximately 50 kb and contains 19 coding exons (Daimon *et al*., 1995). The *HEPH* gene is localized to chromosome Xq11-12 and has 20 coding exons spanning a genomic region of ~100 kb (Syed *et al.*, 2002). It demonstrates almost 85% nucleotide and amino acid homology with the murine homologue, *Heph*, first identified in *sla* mice.

1.1.6 Systemic Iron Transport and Cellular Iron Uptake

Following absorption from the diet by intestinal enterocytes, iron is transported to peripheral tissues and cells around the body by the blood plasma. In order to minimize the potentially toxic effects of free iron, circulating iron in the plasma is bound to the primary iron transport protein, TF (Ponka *et al.*, 1998). This plasma iron pool amounts to between 3 and 4 mg of iron in an adult human (Andrews, 1999). The demand for iron is highest in the bone marrow, where more than 2 million new erythrocytes are produced every second (Cavill, 2002). This translates to an iron requirement of 20-30 mg per day, which is delivered by TF. Almost 80% of circulating TBI is redirected to the bone marrow erythroblasts (Hentze *et al.*, 2004). In order for TBI to gain access to the specific cells in which iron is required, a process of receptor-mediated endocytosis of the TF-Fe complex is achieved by means of binding to specific receptors on the cell membrane, TFR1 and TFR2. Proteins illustrated in Figure 1.4 will be discussed further.

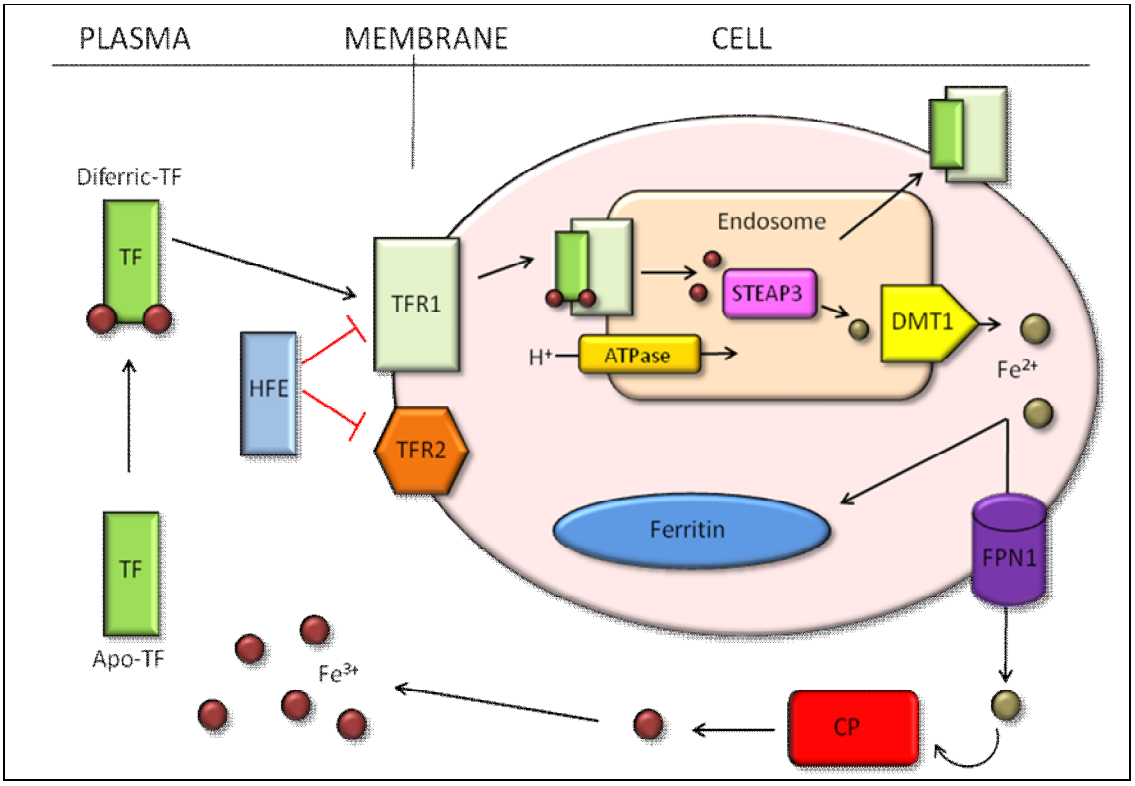


Figure 1.4 Systemic iron homeostasis.

Ferric iron binds to apo-transferrin in the plasma. The diferric transferrin then binds to TFR1 (or TFR2 in the hepatocyte cells of the liver) and is internalized *via* receptor-mediated endocytosis. HFE can compete for binding to TFR1 or TFR2 and subsequently decrease the amount of TBI imported into the cell. A proton pump on the endosomal membrane lowers the pH of the organelle in order to facilitate the release of Fe^{3+}. Within the endosome, STEAP3 reduces Fe^{3+} to Fe^{2+}. DMT1 then transports Fe^{2+} into the cytoplasm where it can be stored in ferritin or transported out of the cell by FPN1. TFR1-apo-TF complexes are returned to the cell surface and recycled for further cellular iron uptake. CP oxidizes Fe^{2+} to Fe^{3+} and facilitates binding to apo-TF. **Abbreviations:** ATP, adenosine triphosphate; CP, ceruloplasmin; DMT1, divalent metal transporter 1; Fe^{3+}, ferric iron; Fe^{2+}, ferrous iron; FPN1, ferroportin 1; H^+, hydrogen ion; HFE, haemochromatosis; STEAP3, six-transmembrane epithelial antigen of the prostate 3; TBI, transferrin-bound iron; TF, transferrin; TFR1, transferrin receptor 1; TFR2, transferrin receptor 2.

Adapted from Lieu et al. (2001).

1.1.6.1 Proteins Involved in Systemic Iron Transport and Cellular Iron Uptake

Transferrin (TF)

As mentioned previously, TF is the major iron transport molecule in the blood plasma of humans. It is responsible for the transport of iron from sites of iron absorption or haem degradation to those of iron storage and utilization (Ponka *et al.*, 1998). As well as TF, other members of the family of transferrin proteins have been identified in physiological fluids of almost all vertebrates studied (Crichton, 2009). The TF family was first recognized during the discovery of a transferrin in egg white, ovotransferrin, in 1900 (Osborne and Campbell, 1900 as reviewed by Montreuil *et al.*, 1997). The subsequent elucidation of an antimicrobial function for TF occurred almost 50 years later (Alderton *et al.*, 1946 as reviewed by Crichton, 2009). All transferrins have been shown to demonstrate a potent antimicrobial effect – due, indirectly, to the deprivation of iron from invading pathogens by the iron-binding capacity of TF.

All TF family members are monomeric glycoproteins with a molecular weight of around 80 kDa. The TF family is divided into classes based on the function and location of the protein in question. They can reversibly bind two atoms of Fe^{3+} with concurrent binding of two anions, usually bicarbonate. In order for the bound iron to be released, acidification must occur as a result of a decrease in the pH of the surrounding fluids (Eckenroth *et al.*, 2011). Characteristically, TF is comprised of two globular protein domains, one at the N-terminus and the other at the C-terminus (Bailey *et al.*, 1988). Each of these lobes is 330 amino acids in length and has identical iron-binding capacity and a very similar fold structure. The N-lobe and the C-lobe are joined together by a three-turn α-helix. The same structure has been confirmed for every member of the TF family in which X-ray crystallography studies have been performed (Anderson *et al.*, 1990).

The process of serum TF (sTF) binding to Fe^{3+} is facilitated by the plasma ferroxidase, CP, in cellular iron uptake; and the membrane-bound homologue, HEPH, during intestinal iron mobilization (see Section 1.1.5.1). It is however, unknown whether these ferroxidases interact with TF directly or not. *In vivo*, TF has been shown to exist in three different states: iron-free (apo-transferrin), one iron molecule bound (monoferric transferrin) and two iron molecules bound (diferric transferrin) (Bou-Abdallah, 2012). CP and HEPH oxidize Fe^{2+} to Fe^{3+} in order to permit iron binding to apo-TF and the subsequent protein conformational

change resulting in the generation of monoferric-TF or diferric-TF. This ligand-induced conformational change in TF upon iron binding was first demonstrated with studies on human apo-lactotransferrin (Anderson *et al.*, 1990).

Cellular uptake of diferric-TF takes place upon binding of TF to transferrin receptors on the cellular membranes of target cells (described below). Within the cell, Fe^{3+} is released from TF in a pH-dependent manner. It is thought that acidification of the endosome (to a pH of 5.5) and subsequent protonation of the bicarbonate anion, results in destabilization of the TF protein structure and the release of the iron molecules (Bali and Aisen, 1991). Numerous studies with labelled TF have revealed that the half-life of the TF molecule is much longer than that of TBI (7.6 days *vs* 1.7 hours) indicating that whilst iron is rapidly removed from the circulation, apo-TF is recycled and utilized in upwards of 100 iron transport cycles before degradation occurs (Dautry-Varsat *et al.*, 1983; Yamashiro *et al.*, 1984).

The gene for TF is comprised of 17 exons and is localized to chromosome 3q22.1. Defects in this gene cause atransferrinemia (ATRAF), a rare autosomal recessive disorder resulting in iron overload and hypochromic anaemia. It has been proposed by Park and colleagues (1985) that the present day genes for the TF family arose from gene duplication and fusion events more than 580 million years ago (Lambert *et al.*, 2005). The hypothesis is based on the fact that the two lobes of the TF polypeptide show almost identical structure, and each is transcribed from seven homologous pairs of exons within the gene.

Lactotransferrin (LTF)

Lactoferrin or lactotransferrin (LTF) is a member of the TF family of proteins and has the same characteristic two-lobed structure of other transferrins. However, LTF is distinct from other members of the TF family with regard to function due to its higher isoelectric point and an increased iron-binding affinity (Pakdaman *et al.*, 1998). In contrast to TF, LTF does not release bound-iron in the presence of an acidic pH. In humans, LTF is expressed in milk, and in secretions such as tears and saliva. It is also secreted at high levels by neutrophils, abundant white blood cells in mammals that form an integral part of the innate immune system (Metz-Boutigue *et al.*, 1984). Receptors for LTF have been identified on the brush-border membrane of foetal enterocytes (Kawakami and Lönnerdal, 1991) which has led to

speculation that LTF may form part of an integral alternative mechanism of iron uptake in early postnatal stages (Lopez *et al.*, 2006).

In adults, however, it appears as if the primary biological function of LTF is not as a regulator of iron homeostasis. Studies in mice have demonstrated that when the LTF gene is disrupted, no noticeable effect on relevant iron parameters or levels was observed (Ward *et al.*, 2003). Whilst TF is responsible for the transport of Fe^{3+} into the cytosol of proliferative cells, LTF is believed to act as a molecular "iron scavenger" thereby sequestering iron from invading microorganisms and proliferating tumour cells that require elevated levels of iron for growth and survival (Aguila *et al.*, 2001). The high affinity of LTF for iron provides indirect evidence for its principle biological role as an inhibitor of bacterial growth, tumour development and free radical formation. In addition to the indirect antimicrobial effects mentioned, LTF also contains a bactericidal peptide-binding sequence, which is distinct from iron-binding regions, in the N-terminus (Bellamy *et al.*, 1992).

The *LTF* gene is located on chromosome 3p21.31 and spans a genomic region of ~30 kb. *LTF* comprises 17 coding exons and, like *TF*, is thought to have arisen from a gene duplication event (Seyfert *et al.*, 1994).

Transferrin receptor 1 (TFR1) and Transferrin receptor 2 (TFR2)

In all actively dividing mammalian cells, uptake of TBI is mediated by the presence of transferrin receptors (TFRs) on the cell membrane. Currently two TFRs have been identified; transferrin receptor 1 (TFR1) and transferrin receptor 2 (TFR2). Both TFRs are homodimeric membrane-bound glycoproteins which are able to bind to diferric-TF thereby permitting the cellular uptake of iron.

TFR1 was the first of the TFRs to be identified and has since been extensively characterized (Cheng *et al.*, 2004). TFR1 is comprised of two identical subunits of 90 000 kDa linked by disulfide bridges. It possesses a single cytoplasmic domain, one transmembrane domain and a large extracellular domain. It is ubiquitously expressed in all proliferative cells of the human body, and particularly in erythroid precursor cells which have a high demand for iron for haemoglobin synthesis (Cavill, 2002).

A second TFR, TFR2, was identified by Kawabata and colleagues in 1999. TFR2 shares significant sequence similarity with TFR1 and is also able to mediate cellular uptake of

diferric-TF. Unlike the universal expression of TFR1, the tissue distribution and expression pattern of TFR2 is markedly different (Deaglio *et al.*, 2002). It is transcribed only by cells of the liver, erythroid precursors and to a lesser extent, by cells of the spleen, lungs and muscles. The affinity of TFR2 for TBI is also 30-fold lower than that of TFR1 (Kawabata *et al.*, 2000; West *et al.*, 2000). Unlike TFR1, expression of TFR2 is not regulated by the IRE/IRP system (described below in Section 1.1.7) or by cellular iron levels. The TFR2 hepatocyte mRNA levels in mice fed an iron deficient, normal or iron overloaded diet showed no difference (Fleming *et al.*, 2000). It is thought that disruptions in TFR2 contribute toward the often observed phenomenon of iron loading observed in the liver in disorders of iron metabolism (Fleming *et al.*, 2000).

The process of TBI iron uptake takes place *via* a receptor-mediated endocytosis mechanism (Ponka *et al.*, 1998). In the plasma, Fe^{3+} is loaded onto TF, which is then able to interact and bind with TFR on the target cell membrane. TFRs do not recognize iron-free apo-TF and preferentially bind to diferric-TF. Upon binding of the ligand to TFR, the complex localizes to a clathrin-coated pit and is internalized *via* endocytosis. The endosome containing the TFR-TF complex is then acidified by means of an ATPase proton pump; the resulting decrease in pH aids in the release of iron from the complex during a protein conformational change (Eckenroth *et al.*, 2011). Within the endosome, Fe^{3+} is reduced to Fe^{2+} by STEAP3 and transported into the cytosol by DMT1 on the endosomal membrane (Gkouvatsos *et al.*, 2012). The iron-free TFR-TF complex is recycled back to the cell membrane where the external physiological pH of 7.4 results in the liberation of apo-TF which is free to chelate more iron (Eckenroth *et al.*, 2011).

The gene encoding TFR1, *TFRC*, is located on chromosome 3q29 and is ~50 kb in length. TFRC contains 19 coding exons, with exons 17-19 encoding residues essential for TF binding to occur (Evans and Kemp, 1997). TFR2 is encoded by the *TFR2* gene on chromosome 7q22.1. Nine mutations in *TFR2* have been linked to the recessive disorder hereditary haemochromatosis type III (HH-III). It is also comprised of 19 exons and spans a genomic region of 22 kb (Fleming *et al.*, 2002).

Six-transmembrane epithelial antigen of the prostate 3 (STEAP3)

After initiation of the TFR-TF cycle of cellular iron uptake, inorganic iron is released within the endosome in the presence of an acidic environment. This is then reduced to Fe^{2+} by the ferrireductase enzyme, six-transmembrane epithelial antigen of the prostate 3 (STEAP3), permitting transport of Fe^{2+} across the endosomal membrane by DMT1. STEAP3 was discovered by positional cloning in iron-deficient anaemic *nm1054* mice (Ohgami *et al.*, 2005a; Ohgami *et al.*, 2005b). It is a member of a unique family of transmembrane ferrireductases containing a C-terminal domain comprised of six transmembrane helices. The reduction activity of STEAP3 is thought to occur *via* sequential electron transfer from NADPH to endosomal Fe^{3+} in association with intramembrane haem and flavin derivatives (Sendamarai *et al.*, 2008).

Human *STEAP3* was mapped to chromosome 2q14.2 and is comprised of six exons coding for a 498 amino acid protein (Korkmaz *et al.*, 2002). Korkmaz and colleagues cloned *STEAP3* in the process of discovering differentially expressed genes in a prostate cancer cell line. They named it six transmembrane protein of prostate 1 (*STAMP1*) and demonstrated that it is highly expressed in epithelial tissues of the human prostate gland. They attributed a secretory or endocytic trafficking role to the STAMP1 protein. This same gene was later reclassified as *STEAP3* during work by the Ohgami research group on *nm1504* mice mutants.

1.1.6.2 Regulation of Systemic Iron Homeostasis Cellular Iron Uptake

At organism level, the regulation of iron balance is an essential process in order to ensure the availability of iron to meet body requirements, without the associated risk of toxicity as a consequence of excess iron. The regulation of systemic iron levels is reflected by iron entry into the plasma in response to the iron status of peripheral tissues that are distinct from the intestinal mucosa (Anderson *et al.*, 2007). Various external signals such as hypoxia and inflammation are known to regulate the amount of iron absorbed by cells from the plasma. With regard to low levels of intracellular oxygen (hypoxia), iron absorption is increased, probably as a result of decreased erythropoiesis as a consequence of tissue hypoxia (Pak *et al.*, 2006). In the case of inflammation, cytokines mediate the reduction of iron export from cells by regulating the synthesis of the peptide hormone hepcidin (Nicolas *et al.*, 2002) (discussed in Section 1.1.6.2.1 below). The rate of erythropoiesis is the driving force behind

systemic iron regulation due to the large amount of iron required for haemoglobin synthesis in developing red blood cells (Cavill, 2002).

1.1.6.2.1 Proteins Involved in Regulation of Systemic Iron Homeostasis

Haemochromatosis (HFE)

The haemochromatosis protein, HFE, is a 348 amino acid type I transmembrane glycoprotein which is homologous to members of the class I major histocompatibility complex (MHC) (Lebrón and Bjorkman, 1999). It is known to form complexes with the light chain of β2-microglobulin for cell surface expression. Unlike other class I MHC molecules HFE is unable to bind peptide antigens and has demonstrated no known role within the immune system (Lebron *et al.*, 1998). This is postulated to be due to a narrow peptide binding groove. The HFE protein is composed of three distinct domains; an extracellular domain (α1, α2 and α3 loops), a transmembrane domain and a cytoplasmic tail.

With regard to regulation of iron homeostasis, HFE competes with TF for binding to the transferrin receptors, TFR1 and TFR2, thereby reducing the affinity of TFRs for diferric-TF and subsequently decreasing the levels of iron absorbed by peripheral tissues (Figure 1.4) (Parkkila *et al.*, 1997; Gross *et al.*, 1998; Waheed *et al.*, 1999). The decrease of TFR1 affinity for TBI in the presence of HFE is thought to be as high as 5-10 fold (Feder *et al.,* 1998). The α helices 1 and 2 of HFE, and the helical domain of TFR1 are principally responsible for the interaction of the two proteins in a pH dependant manner. Diferric-TF can displace HFE from TFR1 because they both compete for overlapping binding sites on TFR1. TFR2 has also been implicated in HFE binding and is thought to compete with TFR1 in this regard, although the protein domains involved are different (Goswami and Andrews, 2006). Unlike TFR1, TFR2 has the ability to bind HFE and diferric-TF simultaneously which may suggest a role as an iron-sensing molecule for TFR2 (Gao *et al.*, 2009). Evidence for the role of HFE as a regulator of hepcidin expression has been inferred from experiments in which patients with *HFE* mutations have unexplained low levels of urinary hepcidin in relation to overall body iron status (Bridle *et al.*, 2003). Mouse models with targeted disruptions of HFE also demonstrate decreased hepcidin mRNA levels in liver tissues (Nicolas *et al.*, 2003).

The *HFE* gene was initially mapped to chromosome 6 in 1976 (Simon *et al.*, 1976) but it was not until 1996 that Feder and his team localized it to 6p21.3 using a positional cloning strategy. *HFE* consists of seven exons and spans 11 kb. Mutations in *HFE* are responsible for the primary cause of iron overload in individuals of Northern European descent, hereditary haemochromatosis type I (HH-I). The main causative polymorphisms are C282Y and H63D. C282Y occurs at nucleotide position 845 which corresponds to the α3 domain of HFE and results in disruption of the β2-microglobulin interaction and subsequent misfolding of HFE (Feder *et al.*, 1996). Homozygosity for C282Y occurs in more than 90% of HH cases, whereas H63D is only identified in 4.5% of HH individuals but occurs frequently in the heterozygous state with C282Y (Merryweather-Clarke *et al.,* 1997).

Hepcidin (HAMP)

Hepcidin is a circulating antimicrobial peptide hormone that is the major regulator of systemic iron balance and controls cellular iron export into the plasma (Ganz, 2005). It is produced primarily by the liver and was discovered by accident during an experiment in mice in which an attempt was made to knock-out an adjacent gene (Nicolas *et al.*, 2001). It is a member of the defensin family of proteins and shows high sequence conservation between different species (Ganz, 1999). The biologically active isoform of hepcidin is a 25 amino acid β-sheet hairpin peptide processed from preprohepcidin, an 84 amino acid precursor molecule. Only this form of hepcidin demonstrates iron regulatory activity and was the first variety to be identified in human urine and plasma (Krause *et al.*, 2000; Park *et al.*, 2001).

The biological mechanism whereby hepcidin influences regulation of iron homeostasis is through an interaction with FPN1, the only known mammalian cellular exporter of iron. As previously mentioned (Section 1.1.5.1), FPN1 facilitates the efflux of Fe^{2+} out of the cytosol and into the plasma. Hepcidin is able to bind directly to FPN1 on the cellular basolateral membrane where it results in the internalisation and degradation of FPN1 (Figure 1.5 A). In response to hepcidin binding, FPN1 is phosphorylated and internalized, followed by ubiquitination and trafficking to the lysosome for degradation (De Domenico *et al.*, 2007). An inverse relationship exists between levels of hepcidin and FPN1: when hepcidin levels are high, the rate of FPN1 degradation exceeds its rate of synthesis and it is lost from the cell membranes resulting in internal sequestration of iron within

enterocytes, macrophages and hepatocytes; when hepcidin levels decrease, FPN1 is able to export iron to the plasma with a subsequent decrease in cellular iron stores (Falzacappa and Muckenthaler, 2005).

Hepcidin expression is known to be regulated by the same factors that control overall systemic iron homeostasis; levels are increased in response to high serum iron, iron overload and inflammation. In response to iron deficiency, hypoxia or elevated erythropoiesis rates the expression of hepcidin is diminished (Figure 1.5 B) (Papanikolaou and Pantopoulos, 2005). Hepcidin is therefore a negative regulator of body iron homeostasis. In addition, hepcidin is positively regulated under normal conditions by signalling through the bone morphogenetic proteins (BMPs) and the SMAD4 pathway which activates hepcidin expression (Babitt *et al.*, 2007).

Hepcidin is transcribed from the hepcidin antimicrobial peptide (*HAMP*) gene located on chromosome 19q13.1. The *HAMP* gene consists of only three exons (Krause *et al.*, 2000), of which exon 3 encodes the active peptide, which contains a unique 17-residue stretch with eight cysteine residues which stabilize the peptide (Hunter *et al.*, 2002). Mutations of this gene have been linked to a severe form of HH, juvenile haemochromatosis or HH type II (Merryweather-Clarke *et al.*, 2003).

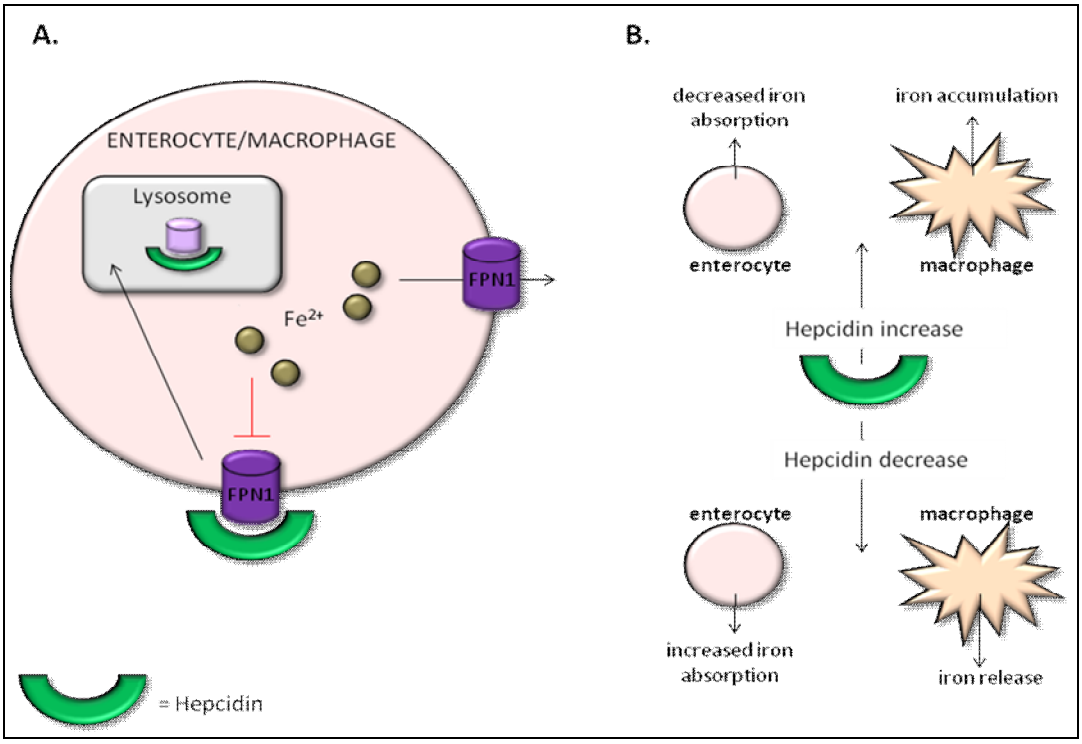


Figure 1.5 Hepcidin – regulator of iron homeostasis.

A.) Binding of hepcidin to FPN1 on the cell surface results in its lysosomal internalization and degradation. A subsequent decrease in Fe^{2+} efflux is observed. **B.**) Erythropoiesis requirements, hypoxia or a reduction in body iron stores results in decreased levels of plasma hepcidin and increased dietary iron absorption and iron efflux from macrophages. Iron overload or inflammation result in increased levels of plasma hepcidin with ensuing inhibition of dietary iron absorption and iron release from macrophages.

Abbreviations: Fe^{2+}, ferrous iron; FPN1, ferroportin 1.

Part A adapted from Andrews (2008); part B adapted from Papanikolaou and Pantopoulos (2005).

Hemojuvelin (HJV)

Hemojuvelin (HJV) is a C-terminal GPI-anchored protein which acts as a co-receptor for BMP with ensuing regulation of hepcidin expression *via* the BMP-SMAD4 signalling cascade (Casanovas *et al.*, 2009). HJV can present in either the membrane-bound, GPI-anchored form or as a soluble molecule (sHJV) (Lin *et al.*, 2008). HJV contains a number of functional sequence motifs including a signal peptide, an arginine-glycine-aspartic acid (RGD) motif and a partial von Willebrand factor type D domain (Papanikolaou *et al.*, 2004). It shares over 50% amino acid identity with members of the repulsive guidance molecule (RGM) family, which are known to act as co-receptors for enhanced BMP signalling (Babitt *et al.*, 2006). Tissue expression of HJV is similar to that of hepcidin and occurs principally in cells of the

liver, skeletal muscle and heart. It has also been identified to a lesser degree in the pancreas and lungs (Papanikolaou *et al.*, 2004). It appears as though the physiological location of HJV action is the liver, despite high expression patterns in skeletal muscle, as evidenced by recent work on murine HJV (Chen *et al.*, 2011).

HJV binds to type I BMP receptors (BMPR) thereby amplifying the signal produced by BMP binding. It can also enhance phosphorylatation of SMAD in conjunction with BMP resulting in elevated hepcidin expression (Figure 1.6). The physiological role that HJV plays in hepcidin regulation was demonstrated by parallel research carried out by two independent groups (Wang *et al.*, 2005; Babitt *et al.*, 2006). Wang and colleagues inactivated SMAD4 and observed a dramatic haemochromatosis phenotype, similar to that observed in hepcidin-knockout mice. Babitt's group studied HJV and showed that treating hepatocytes with BMP in the presence of SMAD and HJV increased hepcidin expression. It is thought likely that SMAD4 binds directly to recognition sequences in the hepcidin promoter after activation by BMP. HJV therefore functions as a positive regulator of hepcidin expression. In contrast, it has been proposed that sHJV acts as an antagonist of GPI-anchored HJV by suppressing the mRNA expression of hepcidin (Figure 1.6) (Lin *et al.*, 2005).

The 426 amino acid protein HJV is transcribed from the haemochromatosis 2 (*HFE2*) gene on chromosome 1q21.1 (Papanikolaou *et al.*, 2004). It was first identified with a positional cloning approach during studies on juvenile haemochromatosis (Papanikolaou *et al.*, 2004). *HFE2* is comprised of four coding exons covering a 4 kb region. Five differentially spliced gene transcripts of *HFE2* have been identified (Severyn and Rotwein, 2010). More than 20 different mutations in *HFE2* have been described which lead to the development of juvenile HH (HH-II) due to perturbation of hepcidin expression (Niederkofler *et al.*, 2005). The same phenotype is observed in patients with mutations in the *HAMP* gene (Papanikolaou *et al.*, 2004).

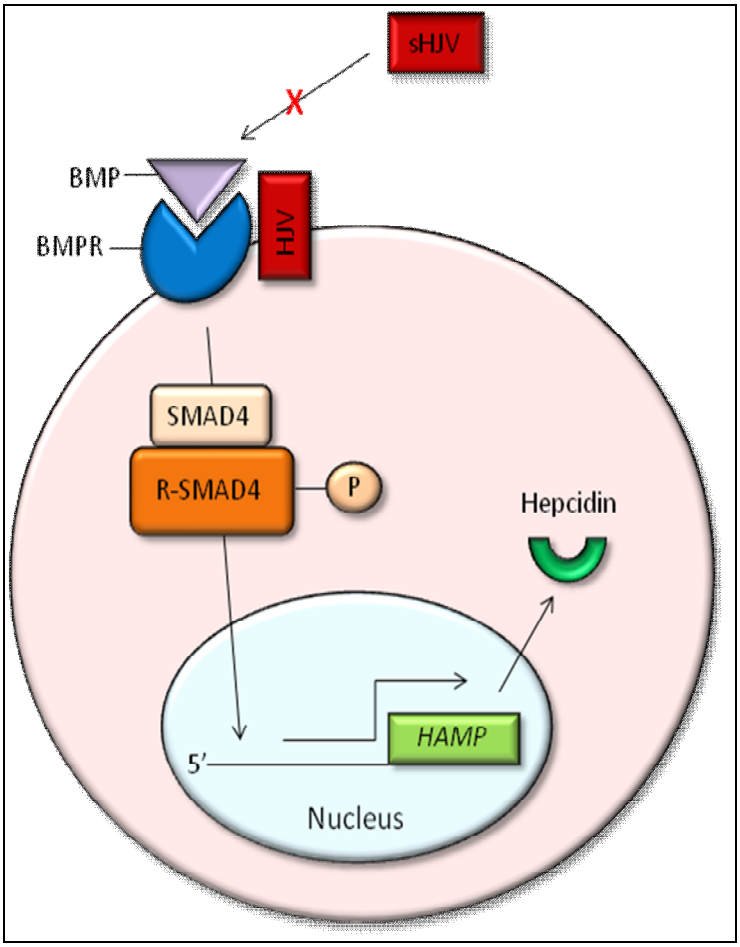


Figure 1.6 Regulation of hepcidin expression by hemojuvelin.

Cell surface BMPRs are activated by binding to BMPs with membrane-bound HJV acting as a co-receptor and increasing the binding efficiency. Activated BMPRs phosphorylate R-SMAD4 which forms a dimer with SMAD4. The R-SMAD4-SMAD4 heterodimer relocates to the nucleus where it binds to nucleotide recognition sites on the *HAMP* promoter and induces elevated hepcidin expression. sHJV has the ability to bind to BMP and inhibit the activation of BMPR resulting in lower levels of downstream hepcidin expression.

Abbreviations: 5', five prime; BMP, bone morphogenetic protein; BMPR, bone morphogenetic protein receptor; *HAMP*, hepcidin antimicrobial peptide gene; HJV, hemojuvelin; sHJV, soluble hemojuvelin; P, phosphate group; R-SMAD4, SMAD4 receptor.

Adapted from De Domenico et al. (2008).

1.1.7 Post-transcriptional Control of Iron Homeostasis

Overall, iron homeostasis is largely regulated at the level of protein synthesis instead of at the level of mRNA synthesis (Meyron-Holtz *et al.*, 2004). This post-transcriptional regulation is mediated by repressors and enhancers in mammalian iron transcribing mRNAs during translation (Johansson and Theil, 2002). These mRNA regulatory recognition sites are located in noncoding, untranslated regions (UTRs), either upstream (5') or downstream (3') of the coding component. Regulatory sequences in the 5'UTR are primarily associated with translation initiation, and those situated in the 3'UTR are involved in mRNA turnover by influencing stability and degradation (Pantopoulos, 2004). In mammals, two iron regulated proteins have been identified; IRP1 and IRP2 (Rouault *et al.*, 1987; Leibald and Munro, 1988).

Iron regulatory protein 1 (IRP1) and iron regulatory protein 2 (IRP2)

Iron regulatory protein 1 (IRP1) and iron regulatory protein 2 (IRP2) are cytosolic RNA binding proteins. They recognize, and bind to, specific nucleotide sequences in the UTRs of mRNA, iron responsive elements (IREs). Several proteins of the iron metabolism pathway are encoded for by mRNAs which contain IREs, which form phylogenetically conserved hairpin structures to which the IRPs bind (Johansson and Theil, 2002). IREs are usually ~30 bp in length and form a characteristic 5'-CAGUGN-3' loop and a stem of moderate stability (Pantopoulos, 2004).

IRP1 and IRP2 are homologous proteins of the iron-sulphur cluster family of isomerases. They share ~70% sequence similarity with each other and are 889 and 963 amino acids in length respectively. During conditions of sufficient iron, IRP1 is bound by an iron-sulphur cluster consisting of four atoms each of Fe and S (4FE-4S). The presence of this cluster renders IRP1 unable to bind to IREs and it also facilitates the conversion of IRP1 into a cytosolic aconitase enzyme (Figure 1.7) (Rouault, 2006; Wallander *et al.*, 2006; Recalcati *et al.*, 2010). IRP1 therefore has dual functions in both iron homeostasis and in the control of redox potential by promoting the formation of cytosolic NADPH.

Under conditions of cellular iron deprivation, the 4Fe-4S cluster is disassembled and IRP1 regains IRE-binding activity. The removal of the Fe-S cluster is also induced by stimuli such as hydrogen peroxide (H_2O_2) and nitric oxide (NO) implicating IRP1 in the inflammatory pathway during ROS formation (Cairo and Pietrangelo, 2000). IRP2 synthesis is triggered

during times of iron depletion and it remains stable and able to bind to IREs under hypoxic conditions. Unlike IRP1, IRP2 undergoes proteosomal degradation when cells are iron replete (Figure 1.7).

IREs in the 5'UTR have been identified in H- and L-ferritin and FPN1 which show decreased expression in times of iron starvation due to IRP binding (Beaumont, 2010). As a consequence, cells minimize iron efflux *via* FPN1 and also enhance iron bioavailability by lowering the levels of iron sequestration by ferritin (Gkouvatsos *et al.*, 2012). In the 3'UTR, an IRE has been found in two isoforms of DMT1 (Gunshin *et al.*, 1997) and there are five 3' IREs in TFR1 (Hentze and Kühn, 1996). IRP binding to these IREs results in stabilization of the DMT1 and TFR1 mRNAs with ensuing cellular iron uptake and transport *via* these proteins under iron deficient conditions. Targeted disruption of the IRP/IRE system is lethal in early embryonic development (Smith *et al.*, 2006), indicating the importance of this regulatory mechanism in sustaining life.

IRP1 is encoded for by the aconitase 1 (*ACO1*) gene on chromosome 9. *ACO1* has 21 exons spanning a genomic region of 3.5 kb (Kühn, 1998). The 963 residue IRP2 protein is transcribed from the iron regulatory binding protein 2 (*IREB2*) gene localized to chromosome 15 (Guo *et al.*, 1995). *IREB2* covers 6.3 kb and contains 22 exons. Both genes are hypothesized to have arisen from a gene duplication event (Wang, 2011).

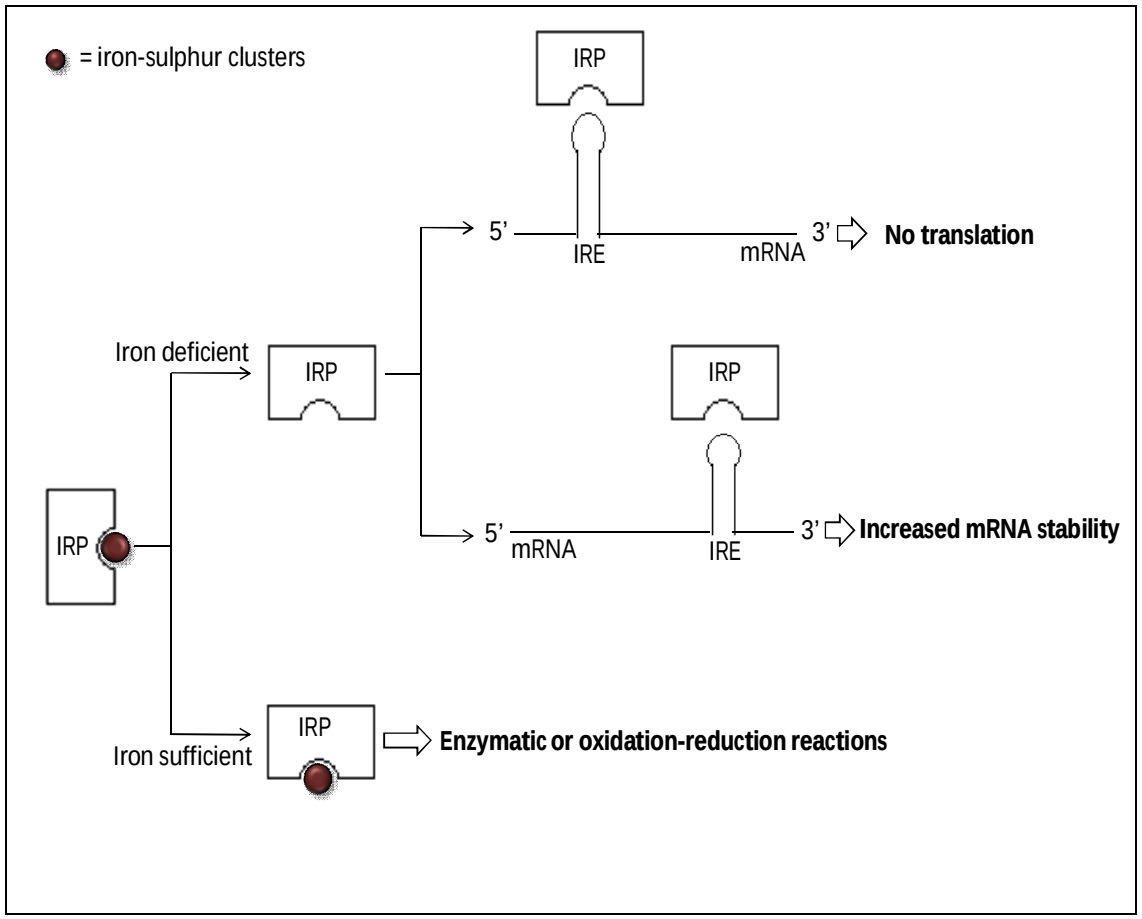


Figure 1.7 IRP/IRE mechanism of post-transcriptional regulation.

Under iron sufficient conditions, IRPs have intact iron-sulphur clusters and function as enzymes during redox reactions. However, under iron deficient conditions, the iron-sulphur cluster is lost from the IRP allowing it to function as a RNA-binding protein. IRPs bind to specific nucleotide sequences called IREs on the mRNA. The binding of an IRP to an IRE located in the 5'UTR of a mRNAs results in translation inhibition. In contrast, the binding of an IRP to a 3'UTR IRE ensures increased translation due to the protection of the mRNA from possible degradation.

Abbreviations: 3', three prime; 5', five prime; IRE, iron responsive element; IRP, iron regulatory protein; mRNA, messenger RNA; UTR, untranslated region.

Adapted from Banerjee et al. (2011).

1.1.8 Disorders of Iron Metabolism

The profound effect of disordered iron homeostasis has been recognized, if not fully understood, for centuries. Disorders of iron metabolism were some of the earliest described human conditions and were often treated with iron supplementation (Sheftel *et al.*, 2012). In recent years, the understanding of iron pathophysiology has improved enormously with modern molecular and genetic tools utilized to elucidate the molecular basis for disruptions of iron metabolism. Studies on mutant mice, rats, zebrafish and humans presenting with inherited disorders of iron metabolism have been enormously beneficial in the identification of novel iron transport proteins (Andrews, 2000). Disorders of iron homeostasis can be classed broadly into those that result in an overload phenotype, and those demonstrating an iron deficiency phenotype.

1.1.8.1 Iron Overload

Iron overload disorders are characterized by the progressive tissue and organ damage as a direct result of the accumulation of intracellular iron stores (Turlin and Deugnier, 2002). The accrual of excess iron leads to the generation of ROS which damage cellular structures and ultimately, result in organ failure if left untreated (Deugnier *et al.*, 2008). Only the most common and well described cause of iron overload, hereditary haemochromatosis, will be discussed here. However, other primary causes of iron overload are well established; aceruloplasminaemia (defects in the *CP* gene resulting in iron deposition in tissues of the CNS), atransferrinaemia (parenchymal iron overload caused by mutations in the *TF* gene) and Friedreich's ataxia (a class of syndromes characterized by iron accumulation in specific cellular compartments) (reviewed by Kushner *et al.*, 2001). Secondary iron overload is also observed in individuals with various haematological disorders such as β-thalassaemia, sickle cell disease and red cell deficiencies (Nemeth and Ganz, 2006).

Hereditary Haemochromatosis (HH)

The German pathologist Friedrich Daniel von Recklinghausen (1889) was the first to use the term "haemochromatose" to describe the bronze-like appearance of the organs of a diabetic patient upon autopsy (as reviewed by Pietrangelo, 2006). It is now known that the characteristic hyperpigmentation of the skin and internal organs occurs as a result of the genetic disorder, hereditary haemochromatosis (HH). HH is an inherited family of disorders

with numerous subtypes that are all characterized by a hereditary origin (usually autosomal recessive), progressively elevated transferrin saturation levels from the normal ~30% to complete saturation, parenchymal iron deposits in endocrine tissues, heart and liver and no noticeable effect on the rate of red blood cell synthesis (Pietrangelo, 2006). In recent years, it has been shown that all subtypes of HH are characterized by inappropriately low levels of hepcidin expression, which results in unrestricted activity of the iron exporter, FPN1. The downstream effects of unrestricted iron efflux are increased sTF saturation and serum ferritin levels, which are the prognostic indicators for HH diagnosis.

The most prevalent form of HH, HH type 1, results from genetic defects in the *HFE* gene (Feder *et al.*, 1996). It occurs at a frequency of ~1 in 200 individuals from North European descent. The most common genetic mutations associated with HH-I are the C282Y (substitution of a Cys for a Tyr resulting from a G to A base change at position 845) and H63D (Waheed *et al.*, 1999). The mechanism underlying C282Y pathogenesis has been described and is attributed to the disruption of a disulphide bond in HFE, preventing association with β2-microglobullin and TFR1 interaction (Feder *et al.*, 1997, Waheed *et al.*, 1999). The molecular effects of H63D are not yet fully understood as it does not impede HFE associations or expression (Waheed *et al.*, 1999).

In addition to classic HH-I, three other subtypes of HH have been described and have assisted further understanding of iron metabolism and regulation (Roetto and Camaschella, 2004).

Juvenile HH (type 2, HH-II) is rare in relation to HH-I but presents at an earlier age and with a much more severe iron overload phenotype. Causative mutations are usually in the *HFE2* (Papanikolaou *et al.*, 2004) and *HAMP* (Roetto *et al.*, 2003) genes. Hepcidin expression is severely impaired in HH-II patients with subsequent excessive iron overload *via* impairment of the FPN1 inhibition.

Mutations in the *TFR2* gene lead to development of the third HH subtype, type 3 (HH-III). The prevalence of HH-III is low compared to the other subtypes and was initially diagnosed in two Sicilian families (Camaschella *et al.*, 2000). The majority of *TFR2* mutations are identified in the heterozygous state and their functional effects have yet to be described. HH-III symptoms and severity parallel those of HH-I, however, the age of onset is similar to that of HH-II presenting in younger individuals.

Type 4 HH (HH-IV) is also called ferroportin disease type B and is caused by relatively rare mutations in the *SLC40A1* gene that transcribes FPN1. Unlike the other HH subtypes, HH-IV is inherited in an autosomal dominant pattern and patients present with a typical haemochromatosis phenotype (Sham *et al.*, 2005).

Treatment of HH-related iron overload begins after correct diagnosis and classification of the molecular basis of the observed phenotype. Advances in genetic testing techniques and imaging technologies allow for the accurate non-invasive determination of HH subtype. Traditional therapeutic phlebotomy is the treatment method of choice for HH (Pietrangelo, 2006). Generally, 400-500 ml of whole blood is removed weekly until serum ferritin levels stabilize at ~30 µg/L and TF saturation is decreased to below 30%. Following this, maintenance therapy is continued by removing 1-2 L of blood each year in order to sustain normal physiological iron levels (Hicken *et al.*, 2003). Phlebotomy therapy is highly effective in the management of iron overload symptoms however, it has little or no effect if started once organ impairment has developed, once again highlighting the importance of accurate and early disease diagnosis. In some cases treatment with iron chelators is indicated, especially for HH-IV individuals as weekly phlebotomy is not well tolerated and can result in moderate anaemia (Brissot *et al.*, 2011).

Research is currently being conducted on the development of specific pharmacological mediators which interact with hepcidin in order to increase production in a targeted manner. However, the efficacy of these antagonists/agonists remains to be fully tested and elucidated *in vivo* (Pietrangelo, 2010).

1.1.8.2 Iron Deficiency

Iron deficiency refers to total body iron deficits that occur when the requirement for iron (primarily for the production of red blood cells) exceeds the supply of iron from depleted stores.

Dietary Iron Deficiency

Dietary iron deficiency is the most common nutritional disorder afflicting developed countries (Stoltzfus, 2003). Despite iron fortification of cereal grain-based foods in most industrial nations, internationally it is estimated that 15% of the population suffer from iron

deficiency anaemia (Looker *et al.*, 1997; Scholl, 2005). Women, and particularly those who have had more than one child, comprise the bulk of iron deficient individuals within a population (WHO, 2004). The major cause of body iron deficit is diets based primarily on plants, with low haem content, and inhibition of dietary non-haem absorption by high fibre contents and other repressors of iron assimilation (Zimmermann and Hurrell, 2007). In developing nations, persistent microbial and parasitic infections result in chronic blood loss and inflammation. Frequently observed vitamin A deficiencies also contribute to the poor absorption of dietary iron (Larocque *et al.*, 2005).

Several different stages categorize iron deficiency: *latent iron deficiency* during which all available iron stores are mobilized, although iron parameters may remain within normal ranges; *iron deficient erythropoiesis* occurs when lack of iron severely limits haemoglobin production; *iron deficiency anaemia* (IDA) where functional concentrations of haemoglobin decrease and lead to the impairment of oxygen transport and changes in red cell morphology (Crichton *et al.*, 2005).

Treatment for iron deficient individuals involves administration of oral iron supplementation, usually in the form of ferrous salts, and is a simple and cost effective strategy (Patterson *et al.*, 2001).

Anaemia of Chronic Disease

Anaemia of chronic disease (ACD) is the phenotypic opposite of haemochromatosis in many respects. ACD is an acquired anaemia second to IDA in prevalence in Western populations (Weiss, 2002) and results from chronic immune system activation during infection, cancer, autoimmune disorders and organ transplantation. During periods of prolonged inflammation or infection, hepcidin production is increased *via* direct cytokine induction, which results in disruptions of intestinal iron absorption and macrophage iron recycling. In addition, diminished serum iron is available for erythropoiesis and macrophage sequestration of iron is observed (Roy and Andrews, 2005).

1.2 TRANSCRIPTIONAL REGULATION OF GENE EXPRESSION

The cellular environment of an individual organism is in a constant state of flux. The ability to co-ordinate the regulation of a complex network of genes is imperative for the growth and survival of an organism (Carter *et al.*, 2004). The complexity of gene regulation reflects the intricacy of the environment in which the products of transcription fulfil a myriad of physiological and functional roles. The molecular control of gene expression allows for certain genes to be turned on or off, both spatially and temporally, thereby determining tissue specificity and cellular development and growth (Butler and Kadonaga, 2002).

In eukaryotic organisms the term "gene expression" encompasses a variety of distinct mechanisms responsible for co-ordinated control of the genome (Pedersen *et al.*, 1999). These include: transcription of DNA to RNA; post-transcriptional modification of mRNA targets (Lareau *et al.*, 2004); translation of mRNA to protein; post-translational processing and modification of protein products; protein folding; epigenetic adjustments such a DNA methylation (Jaenish and Bird, 2003) and chromatin remodelling and adenylation. Different regulatory mechanisms control each of the processes listed above contributing another layer of complexity to the procedure of genetic regulation (Gottesfeld *et al.*, 1997; Steger *et al.*, 2003; Mattick *et al.*, 2009).

Transcriptional regulation is the foremost mechanism by which the genome is able to react and adapt to cellular changes throughout development and in response to external influences (Figure 1.8). Genes are able to react to a cacophony of external stimuli by regulating expression levels appropriately. Each coding gene in eukaryotic genomes is flanked, both upstream and downstream, by regions of genomic DNA that are untranslated. In humans, the amount of this untranslated DNA constitutes ~98% of total genomic content (Elgar and Vavouri, 2008). Prior to the advent of modern sequencing technologies, this so called "junk" DNA was thought to be largely functionless and it was proposed as a type of "molecular scaffold" to simply support the protein coding genes (Carninci *et al.*, 2005). However, with the vast amounts of genomic data that have been generated in the years following the complete sequencing of the human genome (Venter *et al.*, 2001), it is now recognized that these regions are essential for transcriptional regulation.

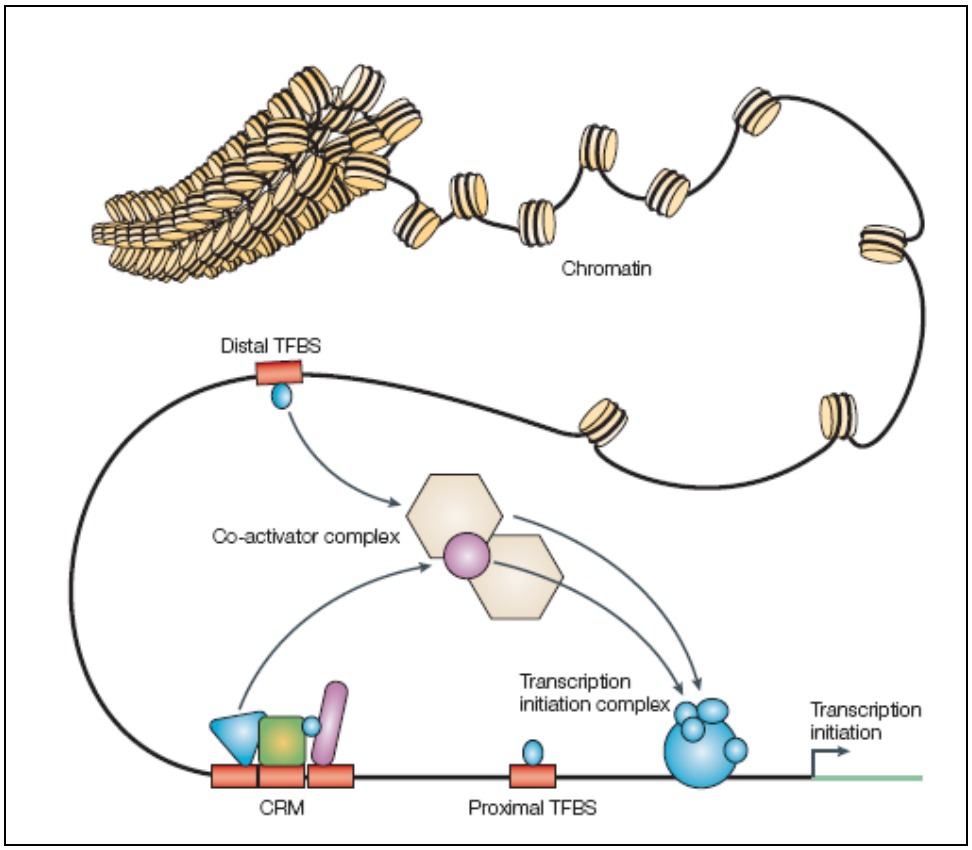


Figure 1.8 Components of transcriptional regulation.

Trans-acting regulatory factors (Tfs) bind to specific TfBSs located either proximal or distal to the TSS. Groups of Tfs can operate together in functional *cis*-regulatory modules (CRMs) to co-ordinate regulation. Cofactors interact with bound Tfs and stabilize the transcription initiation complex assembly to enable regulation of transcription initiation and subsequent gene expression. The three-dimensional chromatin structure can influence the regulation of expression through interference of Tf-binding.

Abbreviations: CRM, *cis*-regulatory module; TF/Tf, transcription factor; TFBS/TfBS, transcription factor binding site; TSS, transcription start site.

1.2.1 Promoter Architecture

The untranslated region upstream (5'UTR) of the coding portion of a gene is responsible for mediating the transcription of the gene and contains the promoter region to which RNA polymerase binds and is directed to the transcription start site (TSS). Exact promoter architecture has proved extremely difficult to define in higher eukaryotes; it may be close to

the transcriptional initiation codon (ATG) or it may be many kb away, or even within another flanking gene. Generally, however, eukaryotic promoters can de described in three sections: the core or basal promoter, to which RNA polymerase II binds, situated about 40 bp upstream of the TSS; the proximal promoter region, ~250-500 bp from the TSS and the distal promoter, which is classified by promoter elements further upstream (Maston *et al.*, 2006).

The core promoter region is defined as the minimum length of sequence required for binding of RNA polymerase and general transcription factors (pre-initiation complex) to initiate accurate transcription (Smale and Kadonga, 2003). Many core promoters contain a sequence recognized by the transcription machinery, the TATA box, which consists of the consensus TATAAAA sequence and directs transcription ~30 bp downstream. It is highly conserved in the promoters of eukaryotes, although mismatches are tolerated in numerous instances. Another class of genes do not contain a TATA-box (TATA-less promoters) and the transcription apparatus is thought to bind to an initiator (Inr) element. TATA-less promoters are generally weaker initiators of transcription (Green, 2005).

Eukaryotes have three different types of RNA polymerases: i) RNA polymerase I transcribes genes for ribosomal RNAs, ii) RNA polymerase II (PolII) transcribes genes encoding mRNAs (the focus of this study) and iii) RNA polymerase III transcribes genes for tRNAs and other small RNA species (Smale and Kadonga, 2003). During *in vitro* transcription, PolII requires the addition of specific auxiliary factors (general Tfs – GTfs) for the initiation of transcription. Binding of a subunit of the GTf, TFIID [comprised of the TATA-binding protein (TBP) and TBP associated factors (TAFs)] to the TATA box initiates the process of transcription by recruiting other GTfs that make up the pre-initiation transcriptional complex (Orphanides *et al.*, 1996).

Transcription mediated by the binding of the pre-initiation assembly to the core promoter proceeds at a basal rate which can be greatly enhanced by the action of additional factors (Figure 1.8).

1.2.2 *Cis*-acting Elements: Transcription Factor Binding Sites (TfBSs)

The mediation of transcription tempo is effected by regulatory sequence elements located upstream of the core promoter (Bolouri and Davidson, 2003). Regulation of gene expression involves occupancy of *cis*-regulatory motifs (CRMs) by transcription factors – often by

recognition of specific binding site motifs, recruitment of co-activators and co-repressors and modifications of the chromatin structure. These specific CRMs are composed of multiple binding sites (transcription factor binding sites, TfBSs) for numerous transcription factors, therefore acting as "molecular switches". Groups of genes may contain identical sequences for particular *cis*-motifs, however, the relative importance of specific motifs may differ in a tissue- and physiological-dependent manner, thereby indicating that not all TfBSs have functional importance *in vivo* (Li and Johnston, 2001).

In general, TfBSs are composed of consensus recognition sites, 6-25 bp in length, which are highly conserved in higher vertebrates (Cooper and Sidow, 2003). The binding of a specific Tf to its cognate TfBS in a sequence specific manner results in activation or repression of transcription (Jones and Takai, 2001). Initiation of transcription is carried out by enhancers, which appear to aid in RNA polymerase II binding and a subsequent increase in transcription, above basal levels (Szutorisz *et al.*, 2005). In contrast, repression of transcription is mediated by silencers. Several types of silencers have been identified that differ on the basis of their mechanism of action. Typically, silencers interfere with positive acting Tfs or they result in modification to chromatin structure and intron splicing which leads to down regulation of gene transcription (Monsalve *et al.*, 2000).

The importance of the interaction between *cis*- and *trans*-acting factors in the regulation of transcription is now widely accepted. Numerous elements in the untranslated regions of genes, particularly the promoter region, contribute to the overall regulation of gene expression with resulting influences on intracellular homeostasis and cell differentiation and growth (Chen *et al.*, 2005).

1.2.3 *Trans*-acting Elements: Transcription Factors (Tfs)

The regulation of transcription by *cis*-acting elements is mediated through binding by *trans*-acting Tfs (Figure 1.8). A multitude of Tfs have been described (>1400) although not all are biologically active in *trans*-regulation in the UTR of all genes (Dröge and Müller-Hill, 2001). The complexity of Tf interaction is illustrated by the fact that out of the ~25 000 genes in the human genome, only ~1800 of them encode Tfs, far less than the number of expression patterns regulated by Tfs for each gene (Maston *et al.*, 2006). Tfs can be classed as either

activators or repressors of transcription depending on interaction with *cis*-regulatory enhancers or silencers respectively.

Activator Tfs are characteristically comprised of two protein domains, a DNA binding domain and an activator domain. Different classes of activators have been identified and are categorized on the basis of structural differences in their DNA binding domains (Scarpulla, 2002). Examples include those containing homeobox, helix-loop-helix (HLH), basic leucine zipper (bZIP) and cysteine rich zinc finger DNA binding domains (Pabo and Sauer, 1992). The activator domain is required for recruitment of the pre-initiation complex, thereby enhancing relative concentration and promoter-binding. Many activators form hetero- or homodimers and binding efficiency is determined by the consensus sequence of the corresponding TfBS. Disruptions to TfBS recognition sequences can dramatically alter the strength of activator binding, influencing regulatory output (Merika and Orkin, 2003). Mutations in numerous TfBSs that result in varied human disorders have been described and illustrate the importance of accurate Tf-TfBS association (summarized in Maston *et al.*, 2006).

Tfs that negatively regulate transcription are termed repressors. In comparison to activators, little about the mechanism underlying transcriptional inhibition by repressor molecules is known. They are postulated to function by a variety of different mechanisms including: recruitment to promoters by other proteins without direct promoter-binding, disruption of the protein levels of the pre-initiation complex, activators or co-activators and direct binding to the activator domain of activator molecules reducing the positive interaction with the pre-initiation complex (Gaston and Jayaraman, 2003).

To complicate the matter further, Tfs (whether activators or repressors) are often modulated by co-regulators themselves (Figure 1.8). These co-regulators do not possess a DNA binding domain, but appear to have functional protein-protein interaction domains and mediate activity of Tfs through direct molecular interactions. In this manner, a single Tf can act as an activator of expression of one gene and as a repressor of transcription of another gene (Spiegelman and Heinrich, 2004).

1.2.4 Prediction of regulatory promoter elements

The biological study of molecular regulatory elements has been conducted for almost 30 years (Nardone *et al.*, 2004). In recent times, the advent of advanced sequencing technologies has led to the deposition of huge amounts of high-quality sequence data into publically available databases (Edgar *et al.*, 2002). Traditional genomic research has seen a shift toward the interpretation of mechanisms of gene regulation and understanding the complexities of the genome (Nardone *et al.*, 2004).

1.2.4.1 *In silico* promoter analysis

The development of computational tools for the discovery and identification of regulatory elements in non-coding DNA has paralleled the exponential rise in sequencing technologies (Wasserman and Sandelin, 2004). This has largely been due to a deficit, in the past, of methods with which to analyze sequence data accurately. The burgeoning field of bioinformatics and *in silico* research has been driven by the need to accurately annotate and characterize genomic information. In addition, relatively little is understood about the regulation of gene expression by the genome as a consequence of limited laboratory experiments on non-coding flanking regions of genes (Pastinen and Hudson, 2004).

In the years following the sequencing and publication of the human genome, numerous projects were started with the goal of investigating the daunting task of annotating the functional elements of the genome, further illustrating the importance of this field of study. The largest of these is the ENCODE (ENCyclopedia of DNA Elements) Project launched by the National Human Genome Research Institute (NHGRI) in 2003 (The ENCODE Project Consortium, 2004). The ENCODE project utilizes a combination of technologies, including bioinformatic analysis, to identify functional elements.

The computational approach to promoter investigation can be viewed in three parts: i) *pattern discovery*, during which the UTRs of a group of co-regulated genes are examined to determine over-represented motifs that are shared among the genes of interest; ii) *pattern matching*, by predicting the location of TfBSs within the given sequence; iii) *pattern recognition*, during which the presence of putative Tfs is inferred from the identified TfBSs (van Helden, 2003).

A huge variety of web-based resources are available for bioinformatic investigation of genomic data (summarized in Nardone *et al.*, 2004; Wasserman and Sandelin, 2004) using different computational algorithms to detect patterns of TfBSs. The degeneracy and short recognition sequences for these *trans*-acting regulatory elements makes their identification in sequence data an overwhelming task and has driven the search for new and more powerful means of analysis. The most widely used algorithm underlying software predication programs is the position weight matrices (PWMs) approach. This is a computationally stringent method of scoring motif matches by generating a matrix of score values for a substring of fixed length (Das and Dai, 2007). However, other algorithms have been developed and are the basis for many other software tools, therefore it is often sensible to utilize different programs for data analysis in order to identify a consensus result. This approach is advisable as it adds weight to the identification of functional elements, as those identified by *in silico* tools with diverse underlying statistical methods may be more functionally relevant (Bulyk, 2003).

Computational analysis of promoters therefore offers a powerful method for uncovering *cis*- and *trans*-acting factors implicated in promoter regulation of gene expression, and can be enormously beneficial in the downstream design of biological experiments.

1.2.4.2 Experimental approaches for identifying regulatory elements

Experimental validation of promoter regulatory targets identified through an *in silico* approach is crucial in order to assess the functional role that these elements may play. Various different *in vitro* assays are utilized for this purpose.

Reporter gene assays

In vitro assays using a reporter gene strategy are among the most versatile techniques available (Maston *et al.*, 2006). DNA fragments of interest are cloned into a plasmid, upstream of a reporter gene, in order to assess the activity of the cloned fragment. Reporter genes are numerous, green fluorescent protein (GFP), luciferase or β-galactosidase are all examples, and are selected on the basis of the particular assay to be performed (Naylor, 1999). The selected reporter gene is quantitatively assayed to measure activity (Alam and Cook, 1990). Within the plasmid, the precise location of the fragment to be cloned depends

on the nature of the functional element. DNA fragments suspected to contain core promoter elements may be cloned directly upstream of the reporter gene (promoterless constructs) (Figure 1.9 A). Proximal promoter elements are cloned upstream of a weak core promoter that drives reporter gene expression at a basal rate and allows for detection of transcriptional enhancement (minimal promoter constructs) (Figure 1.9 B). Enhancer and silencer activity can also be detected upon assay of reporter gene expression levels in plasmids with the appropriate promoters selected (Figure 1.9 C and D).

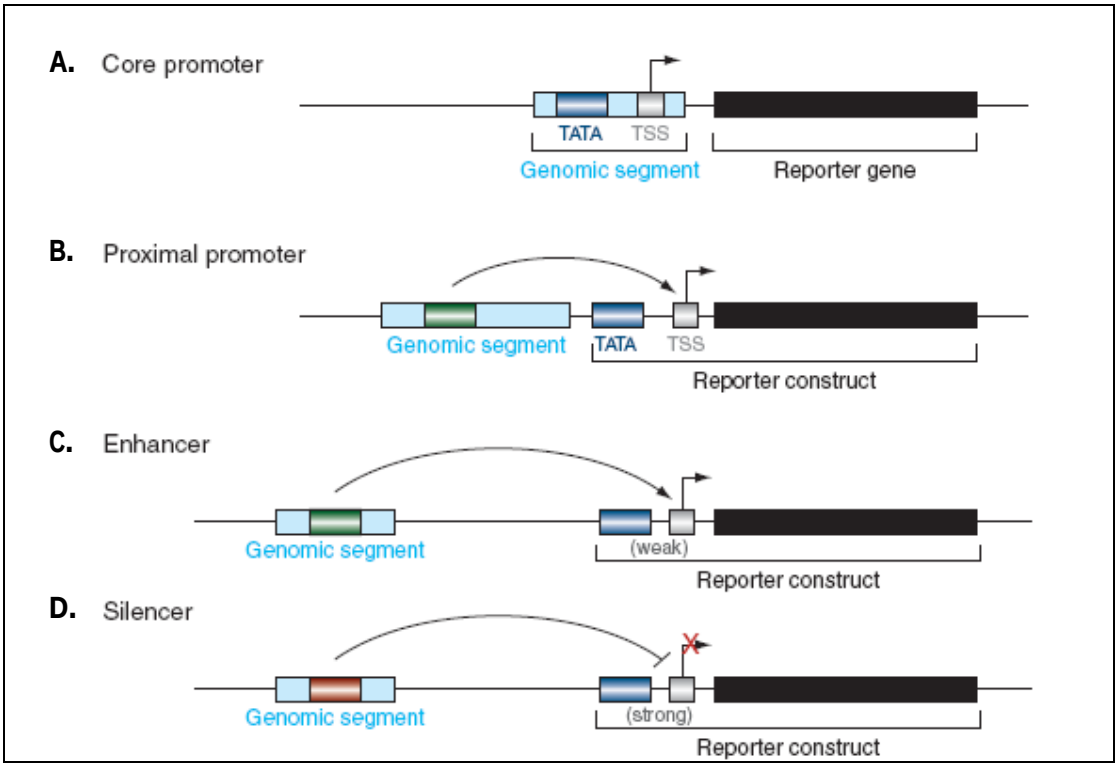


Figure 1.9 Functional assays to measure transcriptional regulation.

A.) In order to assay core promoter activity, the DNA segment of interest (light blue) is cloned into a plasmid upstream of a reporter gene that lacks a functional promoter (promoterless construct). **B – D.**) Proximal promoters, enhancers or silencers are cloned upstream of an appropriate promoter to initiate basal transcription of the reporter gene.

Abbreviations: TATA, consensus binding site for the pre-initiation complex; TSS, transcription start site.

Reprinted by permission from Annual Reviews: [Genomics and Human Genetics] (Maston et al., 2006), copyright (2006).

DNaseI hypersensitive site mapping

DNaseI site mapping is a technique that can be used to identify TfBSs. It is based on the principle that chromatin in genomic regions where particular Tfs have bound undergoes a structural change, and is more sensitive to digestion with DNaseI than unbound chromatin. Mapping of functional regulatory sites by this method has the potential to be a powerful tool in the identification of *cis*-motifs, however it is limited by the fact that it only implies Tf-binding without actual demonstration of the fact (Crawford *et al.*, 2004).

Chromatin immunoprecipitation (ChIP) assays

ChIP analysis allows for the detection of DNA sequences bound by proteins and is an ideal tool for the discovery of Tf-binding. The technique can detect core promoters and provide the possibility of detecting silencers, provided the correct, highly specific antibody for the Tf of interest is used (Kim *et al.*, 2005). ChIP has also been modified for use in large-scale genome wide analysis of Tf-binding by hybridising the DNA purified during ChIP to a DNA microarray (ChIP-chip) (Horak *et al.*, 2002).

1.3 MOTIVATION FOR THIS STUDY

The human genome is a system regulated at many different levels. Transcriptional regulation is the first step in the process of gene expression and is governed by the presence of specific *cis*-regulatory regions (*cis*-motifs), residing within the promoter region of genes, and the functional interactions of *trans*-acting factors (transcription factors-Tfs) and these *cis*-motifs (Maston *et al.*, 2006). Accurate bioinformatic analyses of *cis*-motif architecture combined with experimental validation could offer insights into complex mechanisms governing transcriptional regulation, serving as a refined approach for prediction and the study of regulatory targets (Venter and Warnich, 2009).

Understanding gene regulation and the subsequent analysis of *cis*-motifs in promoters can be a demanding endeavour. It is therefore necessary to explore new technologies such as the use of various *in silico* tools to attempt to formulate putative predictions on how specific *cis*-motifs residing in these regions may influence the expression patterns of specific genes or groups of genes (Wasserman and Sandelin, 2004).

To date, little is known about the co-ordinated transcriptional regulation of genes of the iron metabolism pathway, other than well established post-transcriptional regulation by the IRE/IRP system (Pantopoulos, 2004). Iron is a trace element that is required by almost every living organism. It is used in a variety of cellular functions, and as a result, changes in iron homeostasis within the body can have important clinical consequences (Hentze *et al.*, 2010). Disorders of iron metabolism, be they iron overload or iron deficiency, are amongst the most prevalent conditions affecting individuals globally (Brissott *et al.*, 2011). A combination of factors such as nutrition, environment and genetics play a role in the control of iron levels within an individual. As a result, the cellular mechanisms responsible for iron transport and homeostasis are worthy of significant investigation, as they have the potential to provide future targets for pharmacological intervention thereby benefitting individuals afflicted by disorders of iron metabolism (Pietrangelo, 2010).

The requirement for tight molecular control of iron homeostasis makes this particular biochemical pathway an ideal candidate for the proposed proof-of-concept study of co-ordinated transcriptional regulation. Results of these experiments would represent a unique and comprehensive overview of novel transcriptional control elements of the iron metabolism pathway.

1.4 AIMS AND OBJECTIVES

1.1.4 Aim

To extensively investigate and characterize the promoter and upstream untranslated region of genes implicated in the iron metabolism biochemical pathway, in order to design a strategy for the identification of novel regulatory targets and to contribute toward the understanding of the molecular mechanisms of transcriptional regulation.

1.4.2 Objectives

- Design and establishment of a combinatorial bioinformatic and experimental system for the analysis of gene promoters:

1) Bioinformatic promoter analysis

- Selection of candidate genes involved in the biochemical pathway of iron metabolism and homeostasis.
- Comprehensive *in silico* analyses of the 5' untranslated region of the candidate genes utilizing a variety of open source prediction software.
- Identification and characterization of novel transcriptional regulatory targets.

2) Functional promoter analysis

- Experimental determination of the *in vitro* validity of the bioinformatic results utilizing a functional genetic approach.
- Isolation and molecular cloning of the identified target promoter elements into reporter vectors.
- Transformation of the reporter constructs into a mammalian *in vitro* culture system to evaluate the transcriptional effects of the promoter elements.

CHAPTER TWO

Materials and Methods

2.1 *IN SILICO* PROMOTER ANALYSES

2.1.1 Promoter Sequences

The promoter sequences for 18 of the genes [*ACO1* (ENSG00000122729), *CP* (ENSG00000047457), *CYBRD1* (ENSG00000071967), *FTH1* (ENSG00000167996), *FTL* (ENSG00000087086), *HAMP* (ENSG00000105697), *HEPH* (ENSG00000089472), *HFE* (ENSG00000010704), *HFE2* (ENSG00000168509), *HMOX1* (ENSG00000100292), *IREB2* (ENSG00000136381), *LTF* (ENSG00000012223), *SLC11A2* (ENSG00000110911), *SLC40A1* (ENSG00000138449), *STEAP3* (ENSG00000076351), *TF* (ENSG00000091513), *TFRC* (ENSG00000072274), *TFR2* (ENSG00000106327)] involved in the metabolism of iron were obtained from the Ensembl database (Flicek *et al.*, 2012). 2000 bp (upstream of the initiating ATG) of sequence from the 5'UTR of each gene was utilized for further *in silico* analyses.

2.1.2 Promoter Prediction

Core promoter prediction of the 5'UTR sequence of the 18 iron metabolism genes (Section 2.1.1) was performed to confirm that the sequences retrieved from Ensembl contained transcription start sites (predominately vertebrate PolII promoters), which are key elements of eukaryotic promoter architecture. The 5'UTR sequences were analyzed using the Eukaryotic Promoter Database (Schmid *et al.*, 2005), the Promoter 2.0 program (Knudsen, 1999) and the Neural Network Promoter Prediction program (Reese, 2001).

2.1.3 Pattern Discovery

The promoter sequences of the genes of interest were submitted to mVISTA (Frazer *et al.*, 2004) and YASS (Noe and Kucherov, 2005) in order to search for areas of sequence similarity between the genes. mVISTA is a set of programs developed for the comparison of DNA sequences, up to megabases in length. It provides access to global pairwise, multiple and glocal (global with rearrangements) alignment tools. The program determines the percent identity between two sequences using a sliding window of predefined length, and displays it as a continuous curve. The program also identifies and colours regions of high conservation. YASS is a genomic similarity search tool that produces local pairwise alignments. It is a

heuristic DNA local alignment software which detects potential similarity regions, and then tries to extend them to actual alignments. In contrast to mVISTA, YASS allows only two sequences to be inputted at any one time. For this reason, each gene promoter sequence was compared individually to each of the others in turn, and the results then compared between them.

2.1.4 Pattern Matching

ClustalW (Thompson *et al.*, 1994) was utilized to generate multiple alignments of the regions of similarity in order to prepare a consensus sequence, which was employed in a BLASTn search against the nucleotide (nr/nt) collections and Human ALU repeat elements databases.

The suite of programs, MEME (Bailey *et al.*, 2006) was utilized to further analyze the conserved region (CR) of the gene promoters that was identified in Section 2.1.3.

Motif Discovery

Initially, MEME (Multiple EM for Motif Elicitation) itself was used to discover repeated, ungapped sequence patterns (motifs) that occur in the DNA sequences provided. MEME represents motifs as position-dependent letter-probability matrices which describe the probability of each possible letter at each position in the pattern. MEME allows the user to specify the number of motifs to return in the output. Gaps in motif sequences are not allowed, and as such, MEME splits the motifs with variable-length gaps into one or more separate motifs. For this reason, it is best to search for motifs with a shorter length than that which is specified by the program as default. The parameters used were as follows: Distribution of motif occurrences: one motif per sequence; Number of different motifs: 10; Minimum motif width: 4; Maximum motif width: 15.

2.1.5 Pattern Recognition

In silico analysis was performed in order to identify putative transcription factor binding sites (TfBSs) within each of the identified motifs in the CR of the gene promoters. Several bioinformatic databases are available for *in silico* analysis of the promoter region, of which the following were used in this study: TOMTOM (Gupta *et al.*, 2007), JASPAR CORE (Sandelin

et al., 2004) and TRANSFAC®7 (Wingender *et al.*, 2001). From TRANSFAC®7, two programs were used, namely PATCH and MATCH™ (v1.0) (Kel *et al.*, 2005). Both JASPAR CORE and TRANSFAC®7 contain annotated, matrix-based transcription factor binding site profiles for eukaryotic organisms in an open-access database. Experimentally validated sets of nucleotide sequences that bind transcription factors (Tfs) form the basis of these profiles. The TRANSFAC®7 programs (such as PATCH and MATCH™) use the integrated matrices and site sequences in TRANSFAC®7 to perform matrix-or pattern-based searches of factor binding sites in regulatory DNA sequences. The default parameters were employed in the use of both of these databases. The motifs discovered with MEME were compared to the elements of a database of known motifs using TOMTOM. Matching motifs are ranked by *q*-value. Motifs were searched against the following databases as specified by TOMTOM: JASPAR and UniPROBE, JASPAR CORE vertebrates and All Vertebrates.

In order to compare the putative TfBSs identified by these databases, another program, namely rVISTA (Loots and Ovcharenko, 2004) was also utilized. rVISTA (regulatory VISTA) combines searching the major transcription binding site database TRANSFAC Professional with a comparative sequence analysis and alignment using the global alignment program AVID (Bray *et al.*, 2003).

2.2 DETAILED EXPERIMENTAL PROCEDURES

2.2.1 Polymerase Chain Reaction (PCR) Amplification

Based on the results obtained from the *in silico* promoter analyses, the 5' promoter regions of the *CYBRD1*, *FTH1*, *HAMP*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF* and *TFRC* genes were analyzed for differences in transcriptional regulation and expression. For each gene, three different promoter elements were generated *via* PCR for use in subsequent *in vitro* analyses:
1) an area of 2 kb upstream of the initiating ATG, containing the core- and proximal promoter (termed the 2 kb promoter fragment); 2) a conserved region of 140 bp (termed the CR promoter element); 3) a promoter fragment in which the CR promoter element was removed from the 2 kb promoter fragment (termed the 1.86 kb CR-removed fragment).

Genomic DNA

Human Genomic DNA (hgDNA, Promega Corporation, Madison, WI, USA) was purchased for use in this study in order to minimize possible confounding factors that may exist (e.g. presence of specific disease-related mutations) when selecting individual persons for use as study subjects. The hgDNA was obtained from whole blood samples of multiple anonymous donors and is negative for HIV antibodies (HIV-1/HIV-2) and Hepatitis B surface antigens.

Oligonucleotide Primers

The oligonucleotide primers used in this study were designed to flank the promoter areas of interest using the Primer3 v0.2 program (Rozen and Skaletsky 2000) and are listed in Tables 2.1 – 2.3. All primers were manufactured by Integrated DNA Technologies (IDT, Coralville, IA, USA). The published reference sequence for each gene was obtained from the Ensembl database (Flicek *et al.*, 2012) and is illustrated in Appendix 2, with the relevant positions of the designed primers indicated. Primers that were to be used for cloning were designed to include restriction endonuclease (RE) recognition sites at their 5' ends (either *NheI*, *BglII*, *XhoI* or *HindIII*), flanked by 3 bp random nucleotide overhangs (underlined in Tables 2.1 – 2.2). Appropriate RE sites were assessed by utilizing the reference sequence of the amplicon of interest with an *in silico* RE mapping tool (www.restrictionmapper.org) prior to primer design.

Table 2.1 Oligonucleotide primers designed for PCR amplification of the 2 kb promoter region.

Gene	Forward Primer (a) (5' - 3')	T_m (°C)	Reverse Primer (d) (5' - 3')	T_m (°C)	Product size (bp)	T_a (˚C)
CYBRD1	CAGGCTAGCAAGTGAACGTTCCCAAATGC	63.3	AATAGATCTCTGGCCCCAACTCAGAAAT	59.4	2053	75
FTH1	AATGCTAGCACAACCCTGCCCGATCTT	64.0	AATAGATCTGCCTCTGAGTCCTGGTGGTA	61.6	2232	75
HAMP	CAGGCTAGCCACTCTGCTCACCTTGTGGA	66.3	AATCTCGAGAAACAGAGCCACTGGTCAGG	63.0	2229	75
HFE	CAGGCTAGCAGGAGGAGGAGAATGGAGGA	65.4	AATAGATCTGCTCAGGAGATGCCCAGTAA	60.4	2164	60
HFE2	AATGCTAGCTGGCCTCTAATGGTTGAGAAA	61.3	AATCTCGAGCTGAGGTCAGGGAACCAGAG	64.1	2060	60
HMOX1	AATGCTAGCGCTGCCTCAGAGTCAAGGTC	65.0	CAGAGATCTTGAGGACGCTCGAGAGGA	62.4	2284	75
IREB2	CAGGCTAGCCTCCTAGCATCAAGCGATCC	64.9	AATAGATCTAGGGAGGAAAGAAGGAAGCA	58.5	1998	65
LTF	AATGCTAGCGCTGAGCCTCTTGGTAGTGG	64.9	AATAGATCTAGCAGGACGAGGAAGACAAG	59.0	2166	65
TRFC	ATTGCTAGCAGCCAGGCATGTTAGCTCAT	63.6	CAGAGATCTTTCTAGAAGCCCGCACTCAC	61.7	1985	75

Abbreviations: 5', 5-prime; 3', 3-prime; °C, degrees Celsius; bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; PCR, polymerase chain reaction; T_a, annealing temperature; T_m, melting temperature; *TFRC*, Transferrin receptor protein 1 gene.

Table 2.2 Oligonucleotide primers designed for PCR amplification of the 140 bp CR promoter element.

Gene	Forward Primer (5' - 3')	T_m (°C)	Reverse Primer (5' - 3')	T_m (°C)	Product size (bp)	T_a (˚C)
CYBRD1	CGGGCTAGCGGTTTTTTTTGTTTTTTGTTTT	59.7	CGGAAGCTTCCTGTCTCTACTGAAAATACAAAT	59.7	145	60
FTH1	CGGGCTAGCTCTTTCTTTCTTTCTTTTT	57.7	CGGAAGCTTCAAAAATTAGCCGGGCGTGGTG	66.0	145	60
HAMP	CGGGCTAGCAGCTATTACTATTACTGCTA	59.0	CGGAAGCTTAAATTTAGCCAGGCGTGGTG	63.1	141	65
HFE	CGGGCTAGCATGCCACCACACCCAGCTAA	68.6	CGGAAGCTTCAAAAATTAGCCGGGCGTGATA	63.7	146	65
HFE2	CGGGCTAGCCTTGGTTGCAGTATCTCTTT	63.5	CGGAAGCTTGGAATAACATAATCTTGTGACC	59.2	150	65
HMOX1	CGGGCTAGCAGTTTTTTTTTTTTTTTTTGTTTTT	57.9	GGAAGCTTTAAATTAGCTGGGCATGGTG	60.1	145	65
IREB2	CGGGCTAGCGGACACATTTAATGTCCAAT	62.3	GGAAGCTTTAAAAATTAGCTGAGTGTGGTG	58.3	145	65
LTF	CGGGCTAGCCTTCACTTAGTTACCCTTTT	61.6	GGAAGCTTCAAAAATTAGCCCAGCATGGTG	61.8	147	65
TRFC	CGGGCTAGCTGCTCCTGCTCATATGACTT	64.7	GGAAGCTTAAAATTAGCCAGGCATAGTG	58.0	145	65

Abbreviations: 5', 5-prime; 3', 3-prime; °C, degrees Celsius; bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; PCR, polymerase chain reaction; T_a, annealing temperature; T_m, melting temperature; *TFRC*, Transferrin receptor protein 1 gene.

Table 2.3 Oligonucleotide primers designed for PCR-driven overlap extension of the gene promoters.

Gene	Forward Overlap Extension Primer (c) (5' - 3')	T_m (°C)	Reverse Overlap Extension Primer (b) (5' - 3')	T_m (°C)
CYBRD1	TTTGTTTTTTGTTTTATTTGTATTTTCAGT	51.1	TCAGAAAATACAAATAAAACAAAAAACAAA	50.9
FTH1	CTTTCTTTCTTTTTTCACCACGCCCGGCTA	62.6	TAGCCGGGCGTGGTGAAAAAAGAAAGAAAG	62.6
HAMP	TTACTATTACTGCTACACCACGCCTGGCTA	61.9	TAGCCAGGCGTGGTGTAGCAGTAATAGTAA	61.9
HFE	ACCACACCCAGCTAATATCACGCCCGGCTA	67.1	TAGCCGGGCGTGATATTAGCTGGGTGTGGT	67.1
HFE2	TTGCAGTATCTCTTTGGTCACAAGATTATG	57.1	CATAATCTTGTGACCAAAGAGATACTGCAA	57.1
HMOX1	TTTTTTTTTGTTTTTCACCATGCCCAGCTA	58.7	TAGCTGGGCATGGTGAAAAACAAAAAAAAA	58.7
IREB2	CATTTAATGTCCAATCACCACACTCAGCTA	58.7	TAGCTGAGTGTGGTGATTGGACATTAAATG	58.7
LTF	CTTAGTTACCCTTTTCACCATGCTGGGCTA	61.5	TAGCCCAGCATGGTGAAAAGGGTAACTAAG	61.5
TRFC	CTGCTCATATGACTTCACTATGCCTGGCTA	61.0	TAGCCAGGCATAGTGAAGTCATATGAGCAG	61.0

Abbreviations: 5', 5-prime; 3', 3-prime; °C, degrees Celsius; bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; PCR, polymerase chain reaction; T_a, annealing temperature; T_m, melting temperature; *TFRC*, Transferrin receptor protein 1 gene.

2 kb Promoter and CR Promoter Element PCR Amplification

PCR amplification of the 2 kb promoter fragment and the 140 bp CR promoter element was performed in a gradient thermocycler (Veriti 96 Well Thermocycler, Applied Biosystems, Carlsbad, CA, USA) in reactions with a total volume of 50 µl, consisting of 0.5 U *Taq* polymerase high fidelity, proofreading enzyme [KAPA HiFi™ (HotStart) DNA Polymerase, Kapa Biosystems, Boston, MA, USA], 10 ng hgDNA template, 0.3 mM of each dNTP (dATP, dCTP, dTTP, dGTP) (Kapa), 0.3 µM of each primer and 5 × KAPA HiFi™ Buffer (Kapa). For every round of PCR amplification performed, a negative control (reaction without hgDNA template) was included to indicate the possibility of contamination in the PCR reaction. Individual annealing temperatures for each primer set were optimized and are listed in Tables 2.1 and 2.2. General PCR cycling conditions were: an initial denaturation step at 95˚C for 2 minutes, followed by 35 cycles of denaturation at 98˚C for 20 seconds, annealing at a temperature optimized for each primer set (indicated as the T_a of each amplicon in Table 2.1 and Table 2.2) for 15 seconds and an elongation step of 72˚C for 2 minutes. This was followed by a final extension step of 72˚C for 5 minutes.

PCR-driven Overlap Extension

In order to generate the 1.86 kb CR-removed promoter fragments, the overlap extension PCR technique (Heckman and Pease, 2007) was utilized. This method allows for the deletion of a specified region without the need to introduce foreign nucleotides (as in the case of RE digestion), which have the potential to act as novel TfBSs in the PCR fragment generated. The protocol requires an initial PCR with primers that generate overlapping, complementary 3' ends and results in intermediate segments. These intermediate fragments are then used as template DNA for a subsequent PCR with flanking primers that result in a full length product (Figure 2.1). The primers for the first PCR (b and c) generate overlapping sequences by including nucleotides that span the junction of the AB (dotted line) and CD (dashed line) fragments. The next PCR generates the chimeric gene product AD using flanking primers a and d.

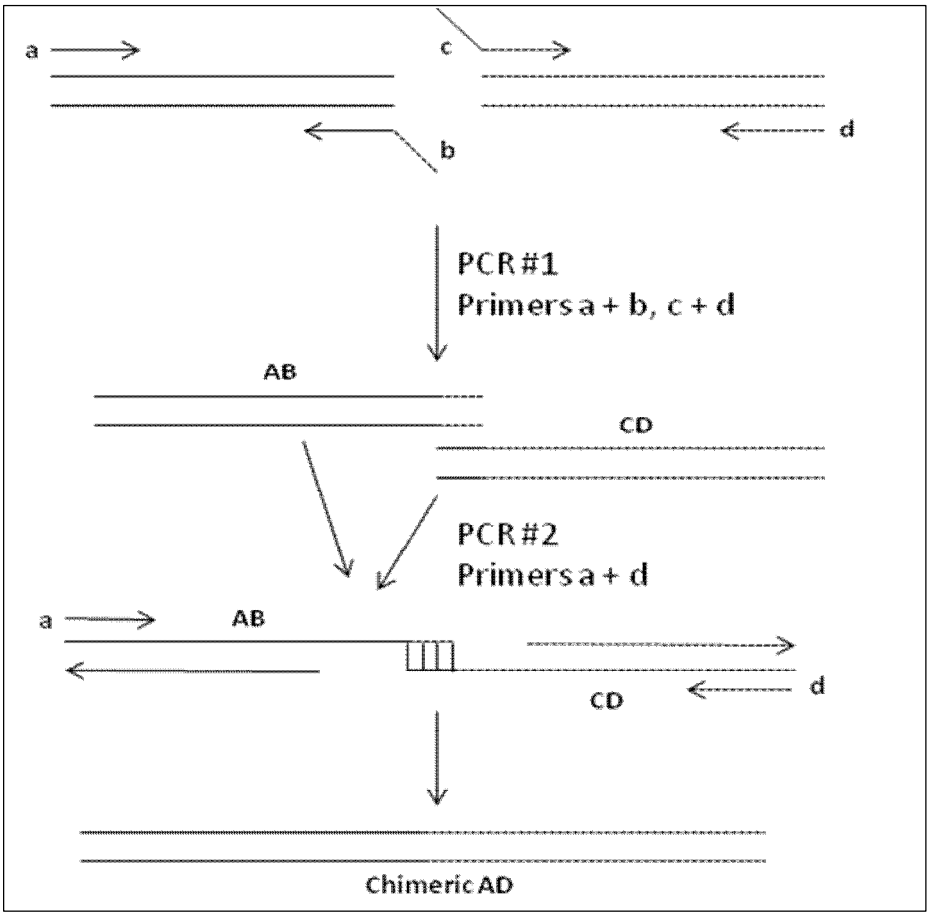


Figure 2.1 Schematic outline of PCR-driven overlap extension.

Adapted from Heckman and Pease (2007).

For the initial PCR, the forward flanking primer (a) was used in combination with the internal overlap primer (b) resulting in the fragment AB; the reverse flanking primer (d) was used with the internal overlap primer (c) to generate the fragment CD. This PCR was performed as a nested reaction; the 2 kb promoter fragment which was previously generated for each gene was used as DNA template. For the second PCR, the two flanking primers (a and d) were used in a PCR reaction with equal concentrations (50 ng) of both the AB and CD fragments as DNA template. Flanking primers (a and d) were the same primers designed for PCR amplification of the 2 kb promoter region (Table 2.1). Internal overlap primers are listed in Table 2.3. The components of the PCR reactions, as well as the cycling conditions were kept the same as those listed in Section 2.2.1.3. The size of each of the PCR amplicons for each gene and their respective annealing temperatures are listed in Table 2.4.

Table 2.4 Product sizes for the fragments generated by the PCR-driven overlap extension.

Gene	Primer Combination	Product size (bp)	T_a (°C)
CYBRD1	Forward Primer (a)/Reverse Overlap Extension Primer (b)	514	55
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	1424	55
	Forward Primer/Reverse Primer (ad)	1938	60
FTH1	Forward Primer (a)/Reverse Overlap Extension Primer (b)	114	55
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	2003	60
	Forward Primer/Reverse Primer (ad)	2117	60
HAMP	Forward Primer (a)/Reverse Overlap Extension Primer (b)	1195	55
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	923	60
	Forward Primer/Reverse Primer (ad)	2118	60
HFE	Forward Primer (a)/Reverse Overlap Extension Primer (b)	338	60
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	1710	60
	Forward Primer/Reverse Primer (ad)	2048	60
HFE2	Forward Primer (a)/Reverse Overlap Extension Primer (b)	851	55
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	1089	55
	Forward Primer/Reverse Primer (ad)	1940	60
HMOX1	Forward Primer (a)/Reverse Overlap Extension Primer (b)	887	55
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	1282	55
	Forward Primer/Reverse Primer (ad)	2169	60
IREB2	Forward Primer (a)/Reverse Overlap Extension Primer (b)	777	55
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	1106	55
	Forward Primer/Reverse Primer (ad)	1883	60
LTF	Forward Primer (a)/Reverse Overlap Extension Primer (b)	769	55
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	1280	55
	Forward Primer/Reverse Primer (ad)	2049	60
TRFC	Forward Primer (a)/Reverse Overlap Extension Primer (b)	612	55
	Forward Overlap Extension Primer (c)/Reverse Primer (d)	1258	55
	Forward Primer/Reverse Primer (ad)	1870	60

Abbreviations: °C, degrees Celsius; bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; PCR, polymerase chain reaction; T_a, annealing temperature; *TFRC*, Transferrin receptor protein 1 gene.

2.2.2 Agarose Gel Electrophoresis

Amplified PCR products were electrophoresed on either 0.8% (w/v) or 1% (w/v) agarose gels [consisting of either 1.6 g or 2 g agarose in 200 ml 1 × Tris-Acetate-EDTA (TAE) (40 mM Tris-HCl, 20 mM acetic acid ($C_2H_4O_2$) and 1 mM EDTA ($C_{10}H_{16}N_2O_8$), pH 8.0) and 0.01% (v/v) ethidium bromide (EtBr)] depending on the size of the relevant DNA fragment. An equal volume of cresol red loading dye solution [2 mg/ml cresol red and 35% (w/v) sucrose] was added to each of the PCR products prior to sample loading in order to resolve each fragment. An appropriate molecular size marker [O'GeneRuler™ 100 bp DNA ladder (Fermentas Life Sciences, Burlington, Ontario, Canada) or O'GeneRuler™ 1 kb DNA ladder (Fermentas Life Sciences)] was electrophoresed alongside the PCR products for accurate size comparison of each specific fragment. Electrophoresis was performed at 120 V for 1 hour in 1 x TAE, after which the DNA fragments were visualized with an ultraviolet (UV) light transilluminator system. The resulting images were captured using the MultiGenius Bio Image System (Syngene, Cambridge, England).

2.2.3 PCR Product and Gel Extraction Purification

The Wizard® SV Gel and PCR Clean-Up System (Promega Corporation) was utilized in order to purify PCR products directly, or to clean-up samples excised from agarose gels using microcentrifugation in a silica membrane-based column system. Incubation and centrifugation times were extended to 5 minutes each. Following agarose gel electrophoresis, the band of interest was excised under UV transillumination and placed in a pre-weighed 1.5 ml microcentrifuge tube (Axygen, Union City, CA, USA). The weight of the gel slice was calculated by subtracting the weight of the empty tube from the total weight. Membrane Binding Solution (4.5 M guanidine isothiocyanate, 0.5 M $C_2H_3KO_2$, pH 5.0) was added at a ratio of 10 µl solution to 10 mg of agarose gel slice and incubated at 55°C for 10 minutes to dissolve the gel slice. For PCR amplifications, an equal volume of Membrane Binding Solution was added directly to the PCR product. The dissolved gel mixture or the prepared PCR product was then added to a SV Minicolumn assembly and incubated for 5 minutes at room temperature. The columns were then subjected to centrifugation (Centrifuge 5415D, Eppendorf, Hamburg, Germany) at 16 000 x *g* for 5 minutes and the flow-through was discarded. The column was then washed by adding 700 µl of Membrane Wash Solution (10 mM $C_2H_3KO_2$, 16.7 µM EDTA, 80% EtOH) and centrifuging again at 16 000

x *g* for 5 minutes. The collection tube was emptied as before and the wash step was repeated with 500 µl Membrane Wash Solution. The flow-through was emptied and the empty column was centrifuged at 16 000 x *g* for 2.5 minutes to remove any traces of ethanol from the wash solution. The SV Minicolumn was transferred to a clean 1.5 ml microcentrifuge tube and 25 µl of Nuclease-Free Water (pre-warmed to 60°C) was added. The samples were incubated at room temperature for 5 minutes followed by a centrifugation at 16 000 x g for 5 minutes. The SV Minicolumn was discarded and the samples stored at 4°C.

2.2.4 DNA Quantification

The concentration of the purified DNA fragments was determined using a spectrophotometer (NanoDrop® ND-100, NanoDrop Technologies Inc., DE, USA) which measures the absorbance of nucleic acids at 260 nm. In this study, A260/A280 ratios in the range of 1.8 – 2.0 and A260/A230 ratios of 2.2 were considered to be high quality, purified and protein-free DNA.

2.2.5 Semi-automated DNA Sequencing Analysis

Purified PCR products were subjected to bidirectional semi-automated DNA sequencing analysis at an analytical facility (CAF, Stellenbosch University or Inqaba Biotec, Pretoria, South Africa) on an ABI Prism 3130XL Genetic Analyser (Applied Biosystems). In order to obtain coverage of the whole region of the 2 kb and 1.86 kb CR-removed amplicons, the sequencing reactions were performed in overlapping fragments using the designed PCR primers (Table 2.1) as well as nested primers that were designed specifically for sequencing analysis (Table 2.5). The 140 bp CR amplicons were sequenced bidirectionally with the relevant PCR primers (Table 2.2).

Analysis of the DNA sequences and chromatograms was performed by alignment with the published reference (wild-type) sequence (accession numbers listed in Section 2.1.1 and Appendix 2) using the BioEdit Sequence Alignment Editor v7.0.9.0 (Hall, 1999) and CLC Sequence Viewer 6.7.1, as well as visual inspection of the electropherograms to detect differences from the reference sequence.

Table 2.5 Internal oligonucleotide sequencing primers.

Gene	Forward Primer (5' - 3')	Reverse Primer (5' - 3')
CYBRD1	GGAGGTGGTCCTTACATG	GAGAACATGTTACATGATG
FTH1	CACCCTAATTCCGTCGGCAAT	GGTGTGTGTGGCTAAGCTGAG
HAMP	CATCGGACTGTAGATGTTAGC	TCAAGACTAGCCTGGGCAAC
HFE	CGGAGCTCTGAACCAGCAAT	GTGCTGAGTTCACTTCGCAG
HFE2	TGACCGTATTTGGAATCGGTC	TCCTTCCTTCCGAGATGAG
HMOX1	GGCCAGTCTCGAACTCAAAG	AATCAATTCTGCATGAGGTGG
IREB2	AATGGACAGAGGGAGGAGGT	CAAGACCAAACACAGCAACG
LTF	CACCATGCTGGGCTAATTTTT	TGGGCAAGATAGGGAAACCCC
TRFC	CAGCAATGCTCCTGCTCATA	AGGTGAGCGGATCACGAG

Abbreviations: 5', 5-prime; 3', 3-prime; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

2.2.6 Luciferase Reporter Vectors

For *in vitro* transfection experiments, two different pGL4 luciferase reporter vectors were selected; the promoterless pGL4.10[*luc2*] vector (GenBank accession number: AY738222) and the minimal promoter pGL4.23[*luc2*/minP] vector (EMBL accession number: DQ904455) (Promega). The pGL4 reporter vectors are optimized for expression in mammalian cells and have been engineered to contain reduced consensus transcription factor binding sites (Appendix 3). Each vector contains the synthetic *firefly luc2* (*Photinus pyralis*) reporter gene. In addition, the pGL4.73[*hRluc*/SV40] (GenBank accession number: AY738229) vector was selected as an internal control reporter for co-transfection with the experimental reporter vectors. The SV40 early enhancer/promoter region contained in this plasmid provides strong constitutive expression of the *Renilla hRluc* (*Renilla reniformis*) reporter gene.

The pGL4.10[*luc2*] vector does not contain any promoter or enhancer elements and the multiple cloning region (MCR) is located immediately upstream of the luciferase reporter gene. This vector was selected in order to clone the 2 kb promoter and 1.86 kb CR-removed fragments. In contrast, the pGL4.23[*luc2*/minP] vector contains a TATA-box promoter element upstream of the luciferase reporter gene and downstream of the MCR, and was used for the cloning of the 140 bp CR promoter elements.

2.2.7 Digestion and Purification

The pGL4.10[*luc2*] and pGL4.23[*luc2*/minP] vectors along with all of the purified promoter fragments of interest (whole 2 kb promoters, the 1.86 kb CR-removed promoters and the 140 bp CR elements) were digested overnight at 37°C in a waterbath in order to generate the correct overhangs for subsequent ligation of the promoter fragments into the revelant vector (Table 2.6). All restriction digests were performed in 20 µl double digestion reactions with 10 U (*NheI*, *XhoI* and *HindIII*; *BglII* was doubled to 20 U as this enzyme only demonstrates 75% performance in buffer 2) of each of the appropriate restriction enzymes for each construct (Table 2.6), 1 x buffer 2 (New England Biolabs, Ipswich, MA, USA), 0.01% BSA (New England Biolabs) and 500 ng of DNA.

The products of the double digest reactions were subjected to purification using the Wizard® SV Gel and PCR Clean-Up System (Promega Corporation) as per Section 2.2.3. The concentration of the digestion products was determined spectrophotometrically as described in Section 2.2.4.

Table 2.6 Double digest restriction enzymes for the pGL4 vectors and each of the promoter elements.

	Enzyme 1	Enzyme 2
Vector		
pGL4.10[*luc2*]	*NheI*	*BglII*
pGL4.10[*luc2*]	*NheI*	*XhoI*
pGL4.23[*luc2*/minP]	*NheI*	*HindIII*
2 kb/1.86 kb CR-removed promoter fragments		
CYBRD1	*NheI*	*BglII*
FTH1	*NheI*	*BglII*
HAMP	*NheI*	*XhoI*
HFE	*NheI*	*BglII*
HFE2	*NheI*	*XhoI*
HMOX1	*NheI*	*BglII*
IREB2	*NheI*	*BglII*
LTF	*NheI*	*BglII*
TFRC	*NheI*	*BglII*
140 bp CR promoter fragments		
CYBRD1	*NheI*	*HindIII*
FTH1	*NheI*	*HindIII*
HAMP	*NheI*	*HindIII*
HFE	*NheI*	*HindIII*
HFE2	*NheI*	*HindIII*
HMOX1	*NheI*	*HindIII*
IREB2	*NheI*	*HindIII*
LTF	*NheI*	*HindIII*
TFRC	*NheI*	*HindIII*

Abbreviations: bp, base pair; CR, conserved region; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; kb, kilobase pair; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

2.2.8 Dephosphorylation of the pGL4 Luciferase Reporter Vectors

Following double digestion, but prior to purification, each of the pGL4 luciferase reporter vectors was treated with the enzyme shrimp alkaline phosphatase (SAP) (Fermentas Life Sciences). SAP removes the 5' terminal phosphate residue in a dephosphorylation reaction that results in the prevention of circularization of the plasmid during ligation. The reaction took place in 1 x SAP dephosphorylation buffer (Fermentas Life Sciences for 30 minutes at 37°C. Heat inactivation of the enzyme was performed at 65°C for 15 minutes. The SAP-treated plasmids were then purified and the concentration determined as described in Sections 2.2.3 and 2.2.4

2.2.9 Ligation into pGL4 Luciferase Reporter Vectors

The purified double digestion products for each promoter fragment were ligated into the appropriate vector in overnight ligation reactions at 4°C. The amount of each insert DNA was calculated using the following equation:

$$\textbf{insert concentration (ng)} = \textbf{vector concentration (ng)} \times \frac{\textbf{insert size (kb)}}{\textbf{vector size (kb)}} \times \textbf{insert:vector ratio}$$

Ligation reactions took place in a final volume of 20 µl consisting of 1 U T4 DNA Ligase (Fermentas Life Sciences), 1 x T4 DNA Ligase buffer [400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP (pH 7.8 at 25°C); Fermentas Life Sciences)], varying concentrations of insert DNA (see below) and 25 ng of vector. Following the overnight incubation step, the T4 DNA Ligase was heat inactivated at 70°C for 5 minutes.

The pGL4.10[luc2] Constructs

The 2 kb promoter elements and the 1.86 kb CR-removed fragments were ligated into the pGL4.10[*luc2*] vector at an insert:vector ratio of 3:1. In order to achieve this ratio with a vector concentration of 25 ng, 33.9 ng (2 kb promoters) or 32.9 ng (1.86 kb CR-removed promoters) of each insert was added to the ligation reaction.

The pGL4.23[luc2/minP] Constructs

The 140 bp CR elements were ligated into the pGL4.23[*luc2*/minP] vector at an insert:vector ratio of 10:1. In order to achieve this ratio with a vector concentration of 25 ng, 8.2 ng of each insert was added to the ligation reaction.

2.2.10 Transformation of *E. cloni* ® 10G Chemically Competent Cells

E. cloni ® 10G Chemically Competent Cells (Lucigen Corporation, Middleton, WI, USA) are derivatives of *E. coli* that have been optimized for high transformation efficiency with the heat-shock method. They contain the *end*A1 and *rec*A1 mutations that result in high yield and high quality plasmid DNA. Before transformation of the *E. cloni* cells, sterile 1.5 ml microcentrifuge tubes were chilled on ice (one for each transformation reaction). The cells were then removed from the -80°C freezer and thawed for 15-20 minutes on ice. A 20 µl aliquot of competent cells was added to each of the chilled 1.5 ml tubes and 5 µl of the specific ligation reaction for each transformation was added. To serve as negative controls, 1 µl unligated vector and/or 5 µl dH_2O were added to the competent cells. The sample was then mixed by gentle pipetting to avoid bubble formation. Incubation for 30 minutes on ice was concluded with a 45 second heat-shock step in a 42°C waterbath, after which the cells were returned to ice for 2 minutes. A volume of 480 µl of room temperature Recovery Medium was added to the cells which were then placed on a platform shaker at 250 rpm at 37°C for one hour. For each transformation reaction, 100 µl of the transformed cells were plated out on Luria-Bertani (LB) plates containing 50 µg/ml ampicillin [Sigma-Aldrich (Pty) Ltd., St Louis, MO, USA]. The plates were incubated overnight at 37°C in order for colony formation to occur.

2.2.11 Clone Selection

Following overnight incubation, colonies were observed on the LB-ampicillin plates. Colony growth is indicative of bacteria that contain the vector-insert combination, as the ampicillin resistance gene (*Amp*^{r}) is contained on the pGL4 vector backbone and can only be expressed in cells that have taken up the recircularized plasmid. Colonies were picked with a sterile pipette tip and inoculated into 5 ml LB-ampicillin medium in a 50 ml tube and incubated overnight at 37°C on a platform shaker at 150 rpm.

The following day: 1) For pGL4.10[*luc2*] constructs, the overnight cultures were used for plasmid miniprep isolation with the GeneJET™ Plasmid Miniprep Kit (Fermentas Life Sciences) as per Section 2.2.12. The purified plasmids (250 ng) were then subjected to overnight double digestion reactions with 10 U of each of the appropriate restriction enzymes for each construct (see Table 2.6), 1 x buffer 2 (New England Biolabs) and 0.01% BSA (New England Biolabs) in a waterbath at 37°C (as per Section 2.2.7). The digested pGL4.10[*luc2*] constructs were resolved on a 1% (w/v) agarose gel in 1 x TAE for 1.5 hours at 100 V. The agarose gel was visualized under UV light transillumination for the identification of plasmids containing the insert of interest.

2) For pGL4.23 [*luc2*/minP] constructs, 1 µl of the overnight culture was used in a colony PCR reaction with a total volume of 25 µl, consisting of 0.5 U *Taq* polymerase enzyme (Fermentas Life Sciences), 0.2 mM of each dNTP (dATP, dCTP, dTTP, dGTP) (Fermentas Life Sciences), 1 x Buffer (Fermentas Life Sciences), 1.5 mM of magnesium chloride ($MgCl_2$), 10 pmol of the vector-specific RV3 forward primer (5' CTAGCAAAATAGGCTGTCCC 3') which flanks the MCR and 10 pmol of the relevant insert-specific reverse primer. Colony PCR amplifications were assessed on a 1% (w/v) agarose gel in 1 x TAE buffer.

2.2.12 Small-scale Isolation of Plasmid DNA (Miniprep)

Following identification of positive colonies, the 5 ml overnight culture was utilized for plasmid DNA extraction using the GeneJET™ Plasmid Miniprep Kit (Fermentas Life Sciences) which makes use of SDS/alkaline lysis to liberate the plasmid DNA. Firstly, the bacterial culture was harvested by centrifugation at 6000 rpm for 2 minutes at room temperature. The supernatant and the remaining medium were decanted and the pelleted cells thoroughly resuspended in 250 µl of Resuspension Solution (with RNase A added). The resuspension was transferred to a 1.5 ml microcentrifuge tube and 250 µl of Lysis Solution was added. The solution was mixed by inverting the tube 4-6 times until viscous, followed by the addition of 350 µl of Neutralization Solution. The tube was immediately inverted 4-6 times to mix the solution. Following a centrifugation step at 16 000 x *g* in a microcentrifuge, the supernatant was transferred to a GeneJET™ spin column without disturbing the white precipitate. The column was then centrifuged for 1 minute at 16 000 x *g*, the flow-through discarded and 500 µl of Wash Solution added. This was followed by a centrifugation step for 1 minute at 16 000 x *g*. This wash step was repeated twice and the flow-through discarded

before another centrifugation step at 16 000 x *g* for 1 minute with the empty column. The GeneJET™ spin column was then transferred to a new 1.5 ml microcentrifuge tube and 50 µl Elution Buffer was added to the centre of the column membrane. The spin column was incubated at room temperature for 2 minutes and centrifuged for an additional 2 minutes at 16 000 x *g* to elute the plasmid DNA. The spin column was discarded and concentration of each extracted plasmid was determined (as per Section 2.2.4). The purified plasmid was stored at -20°C.

The correct orientation and nucleotide sequence of the promoter inserts were verified by DNA sequence analysis (as described in Section 2.2.5) utilizing a plasmid-specific forward primer (RV3; 5' CTAGCAAAATAGGCTGTCCC 3') and an insert-specific reverse primer (Tables 2.1 and 2.2).

2.2.13 Preparation of Glycerol Stocks

Following identification of positive colonies, 850 µl of the 5 ml overnight culture that was prepared for plasmid DNA minipreps was used to create glycerol stocks: 1/10 volume of 100% glycerol and 9/10 volumes of the bacterial culture were mixed vigorously and stored at -80°C.

2.2.14 Large-scale Endotoxin-free Plasmid Isolation (Maxiprep)

Following DNA sequence analysis, the glycerol stocks of the positive clones were used to inoculate 200 ml LB-ampicillin medium which were incubated overnight at 37°C. The overnight bacterial culture was used for endotoxin-free plasmid extraction with the PureYield™ Plasmid Maxiprep System (Promega Corporation). The cells were pelleted by centrifugation at 5000 x *g* for 10 minutes in an ultracentrifuge (Hermle LaborTechnick GmbH, Wehingen, Germany) and the supernatant discarded. The cell pellets were then resuspended in 12 ml of Cell Resuspension Solution after which, 12 ml of Cell Lysis Solution was added and the solution mixed by inverting the tube 3-5 times. This was followed by a 3 minute incubation at room temperature. To the lysed cells, 12 ml of Neutralization Solution was added and mixed by inverting the tube 10-15 times until the solution appeared flocculent. A centrifugation step for 30 minutes at 6000 rpm followed. A column stack was assembled by placing a PureYield™ Clearing Column on top of a PureYield™ Maxi Binding

Column which was then placed onto the vacuum manifold [Vac-Man® Laboratory Vacuum Manifold (Promega Corporation)]. Approximately one-half of the lysate was poured into the column stack assembly and maximum vacuum was applied until all the liquid had passed through both columns. The remainder of the lysate was added and allowed to pass through the columns. The PureYield™ Clearing Column was removed after releasing the vacuum and 5 ml of Endotoxin Removal Wash was added to the PureYield™ Maxi Binding Column. The vacuum was turned on and the solution allowed to pass through the column. This was followed by the addition of 20 ml Column Wash Solution. The column membrane was allowed to dry for 5 minutes under vacuum before it was removed from the vacuum manifold. In order to elute the plasmid DNA, the PureYield™ Maxi Binding Column was placed in a new 50 ml tube (Greiner Bio-one, Frickenhausen, Germany) and 1.5 ml Nuclease-free Water was added to the DNA binding membrane in the column. The column was then centrifuged in a swinging bucket rotor at 2000 x *g* for 5 minutes (Unicen 20 Ultracentrifuge, Orto Alresa, Madrid, Spain). The eluate was collected from the 50 ml tube and transferred to a 1.5 ml microcentrifuge tube. The concentration of each extracted plasmid was determined (as per Section 2.2.4) and stored at -20°C.

2.3 CELL CULTURE, TRANSIENT TRANSFECTIONS AND DUAL-LUCIFERASE REPORTER ASSAYS

2.3.1 Cell Culture

In this study, two different cell types were used to assess the effect of the various iron gene promoter elements on transcriptional activity. Human hepatocarcinoma liver (HepG2, Cat # 85011430) cells and African green monkey kidney (COS-1, Cat # 88031701) cells were purchased from the European Collection of Cell Cultures (ECACC, Wiltshire, England). Both cell types were cultured in 75 cm^2 sterile polystyrene culture flasks (Cellstar®, Greiner Bio-one). Cells were grown in 15 ml Dulbecco's Modified Eagles Medium with 4.5 g/L glucose and 2mM L-Glutamine [DMEM, Sigma-Aldrich (Pty) Ltd.], supplemented with 10% fetal calf serum (v/v) [FCS, Sigma-Aldrich (Pty) Ltd.] and 100 U/ml penicillin and streptomycin solution [Sigma-Aldrich (Pty) Ltd.]. In the case of the HepG2 cells, 1% Non-essential Amino Acid Solution (v/v) [MEM, Sigma-Aldrich (Pty) Ltd.] was added. The cells were kept in an incubator (Heraeus Cell 150, Kendro Laboratory Products, USA) at 37°C with an atmosphere of 5% CO_2. The growth medium was replaced with fresh medium every four days in a sterile laminar flow hood.

In order to determine approximate confluency, cells were examined under an inverted microscope (Nikon DIAPHOT-TMD Biological Inverted Microscope, Nikon Instruments Inc, Melville, NY, USA). Upon reaching 80% confluency, cells were passaged by removing the growth medium and rinsing with Hank's Balanced Salt Solution [Sigma-Aldrich (Pty) Ltd.] to remove any non-adherent or dead cells. Detachment of the cells was achieved by applying 3 ml of 0.25% trypsin-EDTA solution [Sigma-Aldrich (Pty) Ltd.] and incubating the flasks at 37°C for four minutes (COS-1 cells) or 10 minutes (HepG2 cells). Following incubation, 5 ml of fully supplemented culture medium was added and the cell suspension transferred to a sterile 15 ml polypropylene tube (Cellstar®, Greiner Bio-one). The cells were pelleted by centrifugation at 800 rpm for 5 minutes. The supernatant was then discarded and the cells resuspended in another 5 ml of fully supplemented DMEM. Cells counts were performed using an automated cell counter (Countess® Automated Cell Counter, Life Technologies, Carlsbad, CA, USA) with the addition of 0.4% Trypan Blue Stain (Life Technologies) in order

to differentiate between live and dead cells. The cell suspension was then used to seed the cells in new 75 cm^2 flasks at a density of 1 x 10^5 cells/ml, and the flasks were returned to the incubator.

2.3.2 Transient Transfection

Transient transfections of the HepG2 and COS-1 cells with the promoter constructs of interest were achieved with *Trans*IT®-LT1 Transfection Reagent (Mirus Bio LLC, Madison, Wisconsin, USA). *Trans*IT®-LT1 Transfection Reagent is a broad spectrum, low toxicity, serum-compatible reagent that provides high efficiency delivery of plasmid DNA in many primary cell types. Approximately 24 hours prior to transfection, HepG2 and COS-1 cells were plated in 500 µl complete culture medium in 24-well plates (Cellstar®, Greiner Bio-one) at a density of 1 x 10^5 cells/well. The plates were then incubated overnight (37°C, 5% CO_2). After a 24 hour incubation period, the *Trans*IT®-LT1 Transfection Reagent was warmed to room temperature and vortexed gently before use. A total DNA (µg): transfection reagent (µl) ratio of 1:3 was used per well. For each construct, 200 ng (1 µl) of plasmid DNA and 10 ng (1 µl) of the control reporter vector (pGL4.73[*hRluc*/SV40], Promega Corporation) were added to 21 µl of serum- and antibiotic-free medium [100 x the DNA concentration (µg)] and 0.63 µl of *Trans*IT®-LT1 Transfection Reagent. To allow for the formation of the *Trans*IT®-LT1 Transfection Reagent:DNA complexes, the solution was pipetted gently and incubated at room temperature for 30 minutes. Following incubation, 23.63 µl of each complex solution was added dropwise to different areas of the appropriate wells. The 24-well plates were then rocked gently to ensure even distribution of the *Trans*IT®-LT1 Transfection Reagent:DNA complexes and incubated for a further 24 hours (37°C, 5% CO_2).

2.3.3 Exogenous Iron Supplementation

In order to simulate iron overload conditions *in vitro*, cells were treated with ferric ammonium citrate [FAC, Sigma-Aldrich (Pty) Ltd.] as an external stimulus 24 hours after transient transfection. FAC was added to the cells in a dropwise fashion at a final concentration of 200 µg/ml. The same volume of dH_2O was added to the untreated cells. The cells were then incubated for 24 hours (37°C, 5% CO_2) before lysis.

2.3.4 Dual-Luciferase® Reporter Assay

The Dual-Luciferase® Reporter (DLR™) Assay System (Promega Corporation) was used in this study to assay the activities of *firefly* (*Photinus pyralis*) and *Renilla* (*Renilla reniformis*, or sea pansy) luciferases sequentially in a single sample.

Passive Cell Lysis

Following a 24 hour incubation period with FAC, the growth medium was removed by aspiration from each well of the 24-well plate and 200 µl of phosphate buffered saline (PBS) was added to wash the surface of the culture vessel. The plate was swirled to remove detached cells and residual culture medium and the PBS removed. Passive Lysis Buffer (PLB, Promega Corporation), supplied as a 5 x concentrate, was diluted prior to use by adding 1 volume of 5 x PLB to 4 volumes of dH_2O and mixing well. Fresh 1 x PLB was prepared just before use. According to the manufacturer's recommendation, 100 µl of 1 x PLB was added to each well and the culture plates placed on a rocking platform (BellyDancer®, Stovall Life Science Inc., Greensboro, NC, USA) at room temperature for 15 minutes. The plates were then placed at -80°C overnight to ensure complete lysis of the cells.

Dual-Luciferase® Reporter Assay Protocol

Following overnight incubation of the 24-well plates at -80°C, 20 µl of cell lysate was transferred to a 96-well Assay Plate (Costar®, Corning Inc., NY, USA). Luciferase Assay Reagent II (LARII) was prepared by resuspending the lyophilized Luciferase Assay Substrate in 10 ml of the supplied Luciferase Assay Buffer II. In addition, an adequate amount of Stop & Glo® Reagent was prepared just before use for the number of DLR™ assays to be performed (50 µl per assay): 1 volume of 50 x Stop & Glo® Substrate was added to 50 volumes of Stop & Glo® Buffer in a 15 ml polypropylene tube. The plate was then placed within a luminometer (Veritas™ Microplate Luminometer, Promega Corporation) which automatically added 50 µl of LARII to the appropriate wells and monitored the relative light units (RLU) observed using the Veritas v2.0.4401 coupled software program.

2.3.5 Statistical Analysis

The *firefly* luciferase values obtained for each construct were divided by the *Renilla* luciferase readings in the same well in order to obtain normalized values to account for variation in transfection efficiency. For each construct, an average of the normalized values in each experiment was calculated to determine the relative expression of the specific construct under different experimental conditions (i.e. FAC-treated or untreated). Relative expression values for each construct were then divided by those of the constructs designated as control constructs to calculate fold change in luciferase expression.

The fold change values were analyzed using GraphPad Prism® version 5.10 (GraphPad Software). Only values within 20% of one another were used to determine arithmetic means, standard deviation (SD) and standard error (SE) of the means. One-way ANOVA and Dunnett's multiple comparison post-test were used for statistical analysis. Two-tailed paired *t* tests after ANOVA were used when a significant result was obtained in order to evaluate individual *p*-values. Assays for each construct were performed in triplicate in both cell lines and three independent transfection experiments were performed. Error bars represent the standard error of the means (SEM) of the three independent experiments.

CHAPTER THREE

In silico analyses of promoter regulatory targets in the iron metabolism pathway

3.1 ABSTRACT

Transcriptional regulation is governed by the presence of specific *cis*-regulatory regions (*cis*-motifs), residing within the promoter region of genes, and the functional interactions between the products of specific regulatory genes (transcription factors-Tfs) and these *cis*-motifs. Accurate bioinformatic analyses of *cis*-motif architecture could offer insights into complex mechanisms governing transcriptional regulation, serving as a refined approach for prediction and the study of regulatory targets of specific genes or groups of genes, such as the iron metabolism pathway.

The DNA sequence of the upstream non-coding region (refer to Appendix 2 for the respective genes) of 18 genes (*ACO1*, *CP*, *CYBRD1*, *FTH1*, *FTL*, *HAMP*, *HEPH*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF*, *SLC11A2*, *SLC40A1*, *STEAP3*, *TF*, *TFRC*, *TFR2*) known to be involved in the iron metabolism pathway was retrieved from the human Ensembl database. These sequences were subjected to *in silico* analyses to identify regions of conserved nucleotide identity utilising specific software tools.

The sequences of nine (*CYBRD1*, *FTH1*, *HAMP*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF*, *TFRC*) of the 18 genes when examined were found to contain a genomic region that demonstrated over 75% sequence identity between the genes of interest. This conserved region (CR) is approximately 140 bp in size and is common to each of the promoters of the nine genes. This finding adds strength to the hypothesis that genes with similar promoter architecture, and involved in a common pathway, may be co-regulated. The CR was further examined using comparative algorithms from specific platforms for motif detection. Specific combinations of *cis*-motifs were discovered within the CR identified in the promoter regions. *In silico* analysis of putative TfBSs revealed the presence of numerous binding motifs of interest that were detected by more than one bioinformatic tool.

3.2 INTRODUCTION

Within a biological context, iron is an essential micronutrient that is an absolute requirement for life in almost all organisms (exceptions include a few members of lactic acid bacteria and some *Streptococcus* species) (Crichton, 2009). The bioactive form of iron, Fe^{2+}, has the potential to generate free radicals and ROS *via* Fenton and Haber-Weiss chemistry which can lead to cellular and tissue damage, and ultimately disease if untreated (Papanikolaou and Pantopolous, 2005). Globally, disorders of iron metabolism are amongst the most frequent nutritional maladies in humans. The inherited, genetic condition Hereditary Haemochromatosis (HH) is the most common cause of iron overload in Caucasians of North European ancestry with a carrier frequency of 1 in 200 individuals (Feder *et al.*, 1997, Waheed *et al.*, 1999). It has been proposed that the causative HH polymorphisms in different genes, either directly or indirectly, lead to altered levels of the hepcidin protein, thought to function as the "master regulator" of iron metabolism, therefore resulting in the HH phenotype (Nemeth and Ganz, 2006).

In contrast, dietary iron deficiency is the most widespread nutritional disorder affecting industrialized nations, with an estimated 15% of the general population afflicted. In developing countries, this is further exacerbated by chronic microbial and parasitic infections, as well as other vitamin and mineral deficiencies (Looker *et al.*, 1997; Scholl, 2005).

Homeostatic regulation of iron metabolism is of critical importance as there is no known physiological mechanism for the excretion of excess iron. The maintenance of bodily iron levels therefore takes place primarily at the level of iron absorption form the diet (Andrews *et al*., 1999). To date, little is known about the co-ordinated regulation of the plethora of genes coding for proteins involved in iron metabolism, with the exception of the well characterized post-transcriptional IRE/IRP system (Pantopoulos, 2004). This regulatory method is mediated by the interaction of iron regulatory proteins (IRPs) 1 and 2 with iron responsive elements (IREs) in the untranslated regions (UTRs) either upstream (5') or downstream (3') of specific coding mRNAs. These IREs are *cis*-acting consensus recognition sites which result in mRNA degradation or stabilization upon IRP binding in response to different cellular iron levels (Johansson and Theil, 2002). Thus far, IREs have only been

identified in five mammalian iron encoding mRNAs (*DMT1*, *FTH1*, *FTL*, *SLC40A1* and *TFR1*) implying that there may be other regulatory factors responsible for control of transcriptional regulation in this pathway (Hentze and Kühn, 1996; Gunshin *et al*., 1997; Beaumont, 2010). However, this particular mechanism does serve to illustrate the importance of UTRs and *cis*-motifs in the regulation of iron homeostasis.

In the context of overall gene expression, the human genome is a system regulated at numerous different levels. Transcriptional regulation is the first and arguably the most important step in this process. Initiation of transcription involves the occupancy of specific *cis*-regulatory modules (CRMs) by *trans*-acting transcription factors (Tfs), often by the recognition of binding site motifs, recruitment of co-activators and co-repressors and modifications to chromatin structure (Dröge and Müller-Hill, 2001). Determination of these regulatory features has proved a challenging endeavour and the exact mechanisms underlying activation and repression of transcription remain to be elucidated in the vast majority of physiological pathways (Wasserman and Sandelin, 2004). The rapid development of numerous bioinformatic tools for the analysis of regulatory architecture has paralleled the exponential rise in modern sequencing technologies. These tools are available as open source software in many instances and have the potential to be exploited by the biologist for the identification of specific regulatory elements in a cost- and time-effective manner (van Helden, 2003).

Due to the pervasive nature of disorders of iron metabolism, there is a requirement for the development of effective treatment strategies that would benefit those individuals afflicted. The elucidation of specific *cis*-motifs for the co-ordinated regulation of this biochemical pathway could potentially provide targets for pharmaceutical intervention. The molecular control of iron homeostasis in response to physiological iron levels, as a direct consequence of an absence of excretory methods for excess iron, means that this pathway is an ideal model system in which to apply various bioinformatic analyses in order to identify novel transcriptional control elements.

3.3 MATERIALS AND METHODS

The reader is referred to Chapter 2 (Sections 2.1.1 to 2.1.5) for further information regarding the detailed methodology utilized.

3.3.1 Sequence Retrieval and Promoter Prediction

2000 bp (upstream of the initiating ATG, refer to Appendix 2 for the respective genes) of non-coding sequence for 18 of the genes (*ACO1*, *CP*, *CYBRD1*, *FTH1*, *FTL*, *HAMP*, *HEPH*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF*, *SLC11A2*, *SLC40A1*, *STEAP3*, *TF*, *TFRC*, *TFR2*) implicated in the iron metabolic pathway were retrieved from the Ensembl database (Flicek *et al.*, 2012).

Core promoter prediction was conducted on the 5'UTR sequence of each gene using the Promoter 2.0 program (Knudsen, 1999, the Neural Network Promoter Prediction program (Reese, 2001) and the Eukaryotic Promoter Database (Schmid *et al.*, 2005).

3.3.2 *In silico* Detection of Promoter Regulatory Targets

Pattern Discovery

The gene sequences were submitted to mVISTA (Frazer *et al.*, 2004) and YASS (Noe and Kucherov, 2005) in order to search for areas of sequence similarity between the genes.

Pattern Matching

ClustalW (Thompson *et al.*, 1994) was utilized to generate multiple alignments of the regions of similarity in order to prepare a consensus sequence, which was employed in a BLASTn search against the nucleotide (nr/nt) collections and Human ALU repeat elements databases. The regions identified were further analyzed using the suite of programs, MEME (Bailey *et al.*, 2006) to discover repeated sequence motifs.

Pattern Recognition

In order to identify putative transcription factor binding sites (TfBSs) within each of the identified motifs several databases were used: the Motif Comparison Tool, TOMTOM, from the MEME suite was used to compare the identified MEME motifs to the constituents of a database of known motifs (Gupta *et al.*, 2007), JASPAR CORE (Sandelin *et al.*, 2004),

TRANSFAC®7 (Wingender *et al.*, 2001) and rVISTA (Loots and Ovcharenko, 2004). From TRANSFAC®7, two programs were used, PATCH and MATCH™ (v1.0) (Kel *et al.*, 2005).

3.4 RESULTS

3.4.1 Sequence Retrieval and Promoter Prediction

Upstream non-coding sequences were obtained from the human Ensembl database for each of the 18 genes of interest in this study, and are illustrated in Appendix 2. In order to confirm that these sequences contained elements of promoter architecture essential for transcription initiation they were each analyzed with three different promoter prediction *in silico* tools. The results of the analyses from the Neural Network Promoter Prediction (NNPP) software, Promoter 2.0 and the Eukaryotic Promoter Database (EPD) are listed in Table 3.1. Both NNPP and Promoter 2.0 search for binding factors associated with transcription initiation, specifically vertebrate PolII binding sites. NNPP generates scores between 0 and 1, with user-specified cut-off thresholds. For this analysis, 0.8 was selected as the threshold prediction score for maximum stringency, with scores closer to 1 indicating a greater likelihood for the prediction. Predicted transcription start sites (TSS) are depicted in bold, red text in Table 3.1. For Promoter 2.0, a likelihood score is calculated; by default, scores < 0.5 are ignored, those between 0.5 and 0.8 indicate marginal predictions and scores > 0.8 are indicative of highly likely predictions. EPD contains eukaryotic PolII promoters for which the TSS has been experimentally validated.

In Table 3.1, the genomic position corresponds with the sequence for each gene listed in Appendix 2, where 1 was randomly assigned as the first base pair in the sequence.

Table 3.1 *In Silico* Promoter predictions.

	Neural Network Promoter Prediction				Promoter 2.0			Eukaryotic Promoter Database	
Gene	**Start**	**End**	**Score**	**TSS**	**Position**	**Score**	**Likelihood**	**Start**	**End**
ACO1	1175	1225	0.88	attagtact**g**gctattgtt	1800	1.102	High	1534	2116
CP	2192	2242	0.97	TTTTACTTC**A**TTTCTTCTC	400	0.55	Marginal	1732	2266
CYBRD1	2047	2097	0.93	tctttgagc**a**actaactag	2200	1.237	High	1835	2402
FTH1	2190	2240	1.00	CGCAGGGCC**A**GACGTTCTT	2300	1.05	High	1775	2374
FTL	1961	2011	1.00	cgtcccctc**G**CAGTTCGGC	700	0.54	Marginal	1503	2102
HAMP	2182	2232	0.97	GTCCCAGAC**A**CCAGAGCAA	1800	1.02	High	1724	2323
HEPH	1392	1442	0.92	gtaaatgga**a**gtgtgtgca	1100	0.60	Marginal	1503	2102
HFE	2080	2130	0.90	ggatcacct**a**gtgtttcac	1200	0.64	Marginal	1665	2264
HFE2	2054	2104	0.96	agccacctc**c**tccctgct	1500	1.12	High	1668	2267
HMOX1	2422	2472	0.99	CCGGCAGTC**A**ACGCCTGCC	2100	0.626	Marginal	1963	2542
IREB2	1316	1366	1.00	gaggtctgc**a**gaggagcct	2000	0.70	Marginal	1690	2268
LTF	2417	2467	1.00	GCGCAAGTG**G**CAGAGCCTT	1800	0.62	Marginal	1960	2539
SLC11A2	2283	2333	0.99	ATTGGGAAA**G**TCAAACTAG	1900	0.591	Marginal	1009	2390
SLC40A1	2083	2133	0.82	GGCGGAGAG**A**GCTGGCTCA	2200	0.654	Marginal	1647	2246
STEAP3	1893	1943	0.91	agggcggcc**a**cctcccctg	1800	0.67	Marginal	1503	2523
TF	2400	2450	0.98	AGGCTGCAC**A**GAAGCGAGT	1700	0.63	Marginal	1503	2102
TFRC	2257	2307	1.00	CGCCATCCC**C**TCAGAGCGT	1000	0.64	Marginal	1800	2377
TFR2	2814	2864	0.95	ggggcaccc**a**tgtctctgc	1600	0.658	Marginal	2707	3214

Abbreviations: *ACO1*, Aconitase 1 gene; *CP*, Ceruloplasmin gene; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *FTL*, Ferritin light polypeptide gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HEPH*, Hephaestin gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *SLC11A2*, Solute carrier family 11 member 2 gene; *SLC40A1*, Solute carrier family 40 member 1 gene; *STEAP3*, Six-transmembrane epithelial antigen of the prostate 3 gene; *TF*, Transferrin gene; *TFRC*, Transferrin receptor 1 gene; *TFR2*, Transferrin receptor 2 gene.

3.4.2 Pattern Discovery

Submission of the 2 kb UTR sequences to mVISTA demonstrated a region of sequence similarity between nine of the 18 (50%) iron metabolism genes (*CYBRD1*, *FTH1*, *HAMP*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF* and *TFRC*). The other nine genes of interest (*ACO1*, *CP*, *FTL*, *HEPH*, *SLC11A2*, *SLC40A1*, *STEAP3*, *TF* and *TFR2*) did not show any defined regions of sequence similarity when compared to each other (Figure 3.1). mVISTA requires a single inputted sequence to be used as a so-called "base genome"; therefore each of the genes in turn was utilized as the base genome for comparison to each of the others in order to generate an accurate representation of the alignment. Figure 3.1 illustrates the alignment of all the sequences against *HFE* which was utilized as the base genome in this instance.

In order to confirm the accuracy of the predictions by mVISTA, the genomic similarity search tool, YASS, was utilized. The 2 kb sequences of all 18 genes were submitted and YASS results indicated that an area of sequence conservation was detected in the same nine genes (*CYBRD1*, *FTH1*, *HAMP*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF* and *TFRC*) as in those predicted by mVISTA (Table 3.2). In Table 3.2, the numbers underneath the gene names correspond to the genomic position of the region of sequence similarity in the 5'UTR of each gene, with 1 arbitrarily assigned as the first base pair in the sequence for each gene listed in Appendix 2. The positions of the sequence conservation as predicted by mVISTA have been included for comparison, as well as the percentage similarity for the relevant alignment. Percentage similarities ≥ 75.0% are highlighted in red. In addition, YASS generated an *E*-value for each alignment pair: the smaller the *E*-value, the greater the significance of the proposed sequence similarity. *E*-values are a parameter used to describe the number of hits expected when searching a database of a particular size.

The design of the strategy employed in this study (to search for and identify conserved motifs in the promoters of the genes involved in iron metabolism) was based on an assertion of some degree of detectable genomic sequence similarity between the genes. As a consequence, further *in silico* analyses were restricted to those nine genes containing the region of sequence homology discovered in the initial step of pattern discovery with mVISTA and YASS.

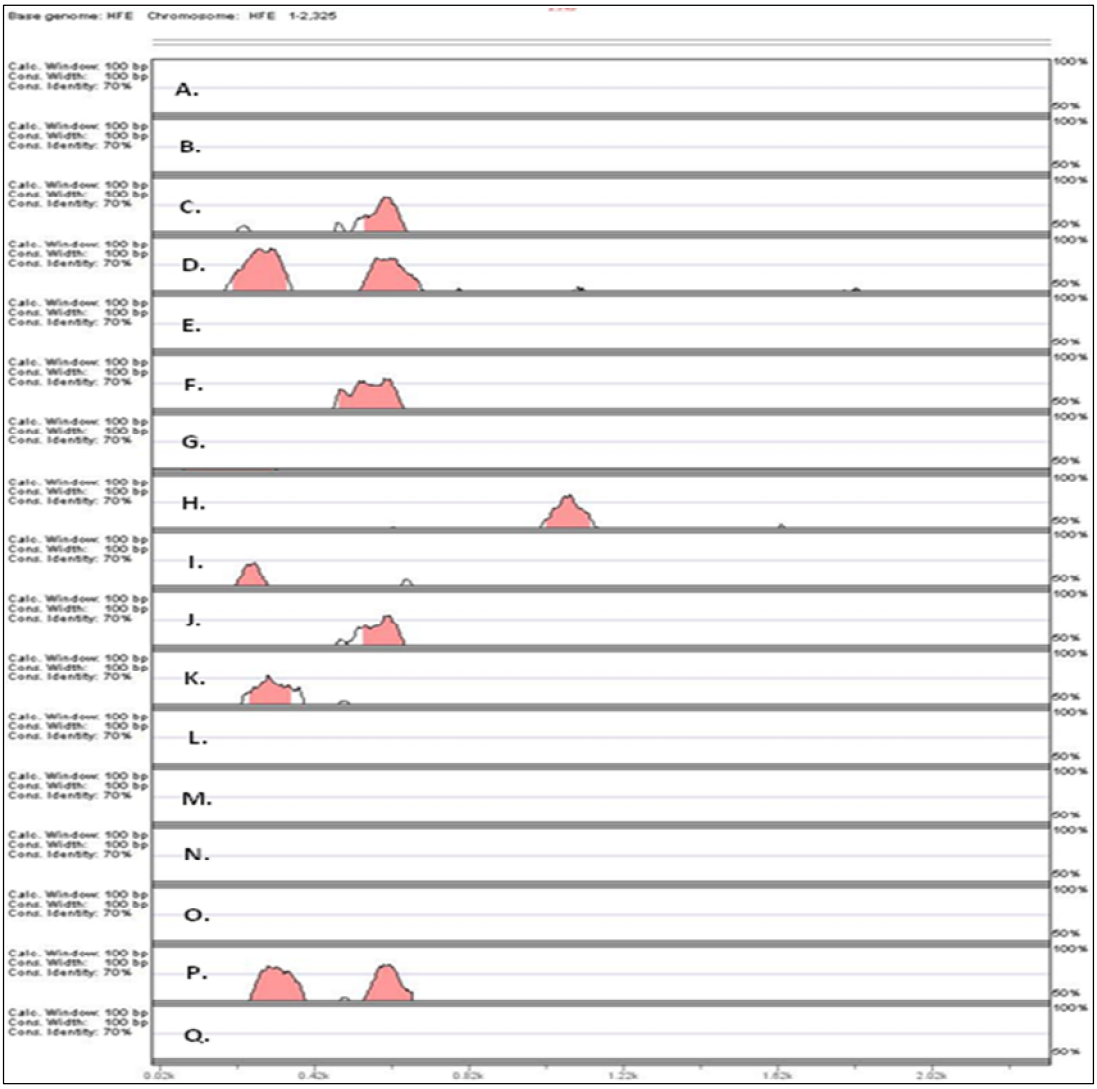


Figure 3.1 mVISTA sequence alignments of each of the 18 iron metabolism genes.

Pink coloured areas indicate regions of sequence similarity. In this example *HFE* was utilized as the base genome for visualization of the alignments against: **A**.) *ACO1*; **B**.) *CP*; **C**.) *CYBRD1*; **D**.) *FTH1*; **E**.) *FTL*; **F**.) *HAMP*; **G**.) *HEPH*; **H**.) *HFE2*; **I**.) *HMOX1*; **J**.) *IREB2*; **K**.) *LTF*; **L**.) *SLC11A2*; **M**.) *SLC40A1*; **N**.) *STEAP3*; **O**.) *TF*; **P**.) *TFRC* and **Q**.) *TFR2*.

Abbreviations: *ACO1*, Aconitase 1 gene; *CP*, Ceruloplasmin gene; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *FTL*, Ferritin light polypeptide gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HEPH*, Hephaestin gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *SLC11A2*, Solute carrier family 11 member 2 gene; *SLC40A1*, Solute carrier family 40 member 1 gene; *STEAP3*, Six-transmembrane epithelial antigen of the prostate 3 gene; *TF*, Transferrin gene; *TFRC*, Transferrin receptor 1 gene; *TFR2*, Transferrin receptor 2 gene.

Table 3.2 mVISTA and YASS sequence alignments.

mVISTA			YASS			
Gene 1	**Gene 2**	**Percentage Similarity**	**Gene 1**	**Gene 2**	**Size (bp) (Gene1/Gene2)**	***E*-value**
CYBRD1	***FTH1***		***CYBRD1***	***FTH1***		
669-813	358-503	**75.9%**	649-945	339-654	297/316	2.0114e-52
CYBRD1	***HAMP***		***CYBRD1***	***HAMP***		
669-813	1302-1442	**77.2%**	659-954	1293-1608	296/316	1.01832e-45
CYBRD1	***HFE***		***CYBRD1***	***HFE***		
669-813	382-527	74.0%	550-975	294-706	426/413	4.84393e-54
CYBRD1	***HFE2***		***CYBRD1***	***HFE2***		
686-781	984-1074	**78.8%**	644-798	937-1090	155/154	9.61974e-17
CYBRD1	***HMOX1***		***CYBRD1***	***HMOX1***		
689-813	1098-1222	71.2%	649-976	1057-1425	328/369	4.21095e-45
CYBRD1	***IREB2***		***CYBRD1***	***IREB2***		
672-813	1020-1161	**81.0%**	668-940	1016-1307	273/292	1.2211e-44
CYBRD1	***LTF***		***CYBRD1***	***LTF***		
698-806	1142-1250	69.7%	636-946	1083-1408	311/326	1.01832e-45
CYBRD1	***TFRC***		***CYBRD1***	***TFRC***		
669-813	980-1124	**75.9%**	661-965	972-1297	305/326	2.03663e-49
FTH1	***HAMP***		***FTH1***	***HAMP***		
358-503	1302-1442	72.4%	343-654	1285-1597	312/313	3.84786e-24
FTH1	***HFE***		***FTH1***	***HFE***		
358-503	382-527	**83.6%**	359-654	382-676	296/295	6.49006e-69
FTH1	***HFE2***		***FTH1***	***HFE2***		
394-491	997-1093	**75.5%**	354-491	948-1093	138/146	3.35897e-14
FTH1	***HMOX1***		***FTH1***	***HMOX1***		
358-503	1078-1222	71.7%	312-654	1033-1394	343/362	6.0802e-42
FTH1	***IREB2***		***FTH1***	***IREB2***		
362-503	1020-1161	**75.4%**	350-653	1011-1311	304/301	7.05255e-49
FTH1	***LTF***		***FTH1***	***LTF***		
377-501	1131-1256	69.4%	338-654	1089-1407	317/319	6.87887e-55
FTH1	***TFRC***		***FTH1***	***TFRC***		
358-503	980-1124	**87.6%**	333-653	957-1276	321/320	1.10518e-68
HAMP	***HFE***		***HAMP***	***HFE***		
1302-1442	375-681	**75.3%**	1296-1602	375-681	307/307	4.02278e-56
HAMP	***HFE2***		***HAMP***	***HFE2***		
1320-1440	982-1104	73.0%	1320-1430	982-1093	111/112	3.33117e-16
HAMP	***HMOX1***		***HAMP***	***HMOX1***		
1329-1442	1105-1222	72.9%	1276-1597	1050-1394	322/345	1.20097e-48
HAMP	***IREB2***		***HAMP***	***IREB2***		
1309-1442	1024-1161	**76.1%**	1300-1591	1018-1306	292/289	4.74432e-59
HAMP	***LTF***		***HAMP***	***LTF***		
1322-1429	1133-1244	**75.0%**	1317-1609	1128-1419	293/292	9.76868e-56
HAMP	***TFRC***		***HAMP***	***TFRC***		
1302-1442	980-1124	**77.2%**	1284-1598	965-1278	315/314	8.04553e-60
HFE	***HFE2***		***HFE***	***HFE2***		
402-515	984-1093	72.8%	141-353	874-1093	213/220	2.78955e-16

HFE	***HMOX1***		***HFE***	***HMOX1***		
384-513	1078-1208	69.2%	182-540	1035-1393	359/359	6.02988e-44
HFE	***IREB2***		***HFE***	***IREB2***		
385-528	1020-1161	72.0%	378-671	1012-1307	294/296	4.15877e-48
HFE	***LTF***		***HFE***	***LTF***		
401-513	1132-1243	72.6%	378-681	1106-1412	304/307	2.0114e-52
HFE	***TFRC***		***HFE***	***TFRC***		
382-527	980-1124	**85.6%**	335-695	930-1297	361/368	2.63952e-72
HFE2	***HMOX1***		***HFE2***	***HMOX1***		
991-1091	1107-1208	**77.5%**	948-1091	1071-1208	144/138	1.94808e-17
HFE2	***IREB2***		***HFE2***	***IREB2***		
959-1092	1023-1147	72.1%	982-1093	1035-1149	112/115	2.30706e-19
HFE2	***LTF***		***HFE2***	***LTF***		
959-1090	1112-1240	69.2%	946-1091	1099-1243	146/145	2.77798e-17
HFE2	***TFRC***		***HFE2***	***TFRC***		
991-1090	1009-1112	**77.8%**	988-1093	1006-1112	106/107	3.96145e-17
HMOX1	***IREB2***		***HMOX1***	***IREB2***		
1079-1222	1018-1161	73.6%	1075-1279	1016-1219	205/204	1.26764e-35
HMOX1	***LTF***		***HMOX1***	***LTF***		
1098-1218	1133-1253	**76.0%**	1057-1395	1085-1408	339/324	4.80384e-56
HMOX1	***TFRC***		***HMOX1***	***TFRC***		
1078-1222	980-1124	73.1%	1074-1335	978-1245	262/268	7.26073e-42
IREB2	***LTF***		***IREB2***	***LTF***		
1037-1154	1133-1250	**76.3%**	1002-1308	1094-1403	307/310	3.98948e-58
IREB2	***TFRC***		***IREB2***	***TFRC***		
1020-1161	982-1124	**78.9%**	1016-1430	979-1391	415/413	3.31318e-60
LTF	***TFRC***		***LTF***	***TFRC***		
1131-1251	998-1118	72.7%	1107-1406	981-1276	300/296	2.89222e-50

Abbreviations: bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor 1 gene.

The conserved regions (CRs) predicted by mVISTA and YASS were inspected visually and narrowed down to the most representative sequence from each of the genes, in order to include the entire region of conservation that was common to all of the genes examined. A 140 bp area was selected from each gene (illustrated in green text in Appendix 2) and these sequences were resubmitted to mVISTA for alignment of the CR. A representative example of the mVISTA results of the CR alignment is illustrated in Figure 3.2. In this illustration *CYBRD1* was utilized as the base genome and compared to the eight remaining genes. The mVISTA results for each of the other genes in turn are depicted in Appendix 4: Figure S1, page 179).

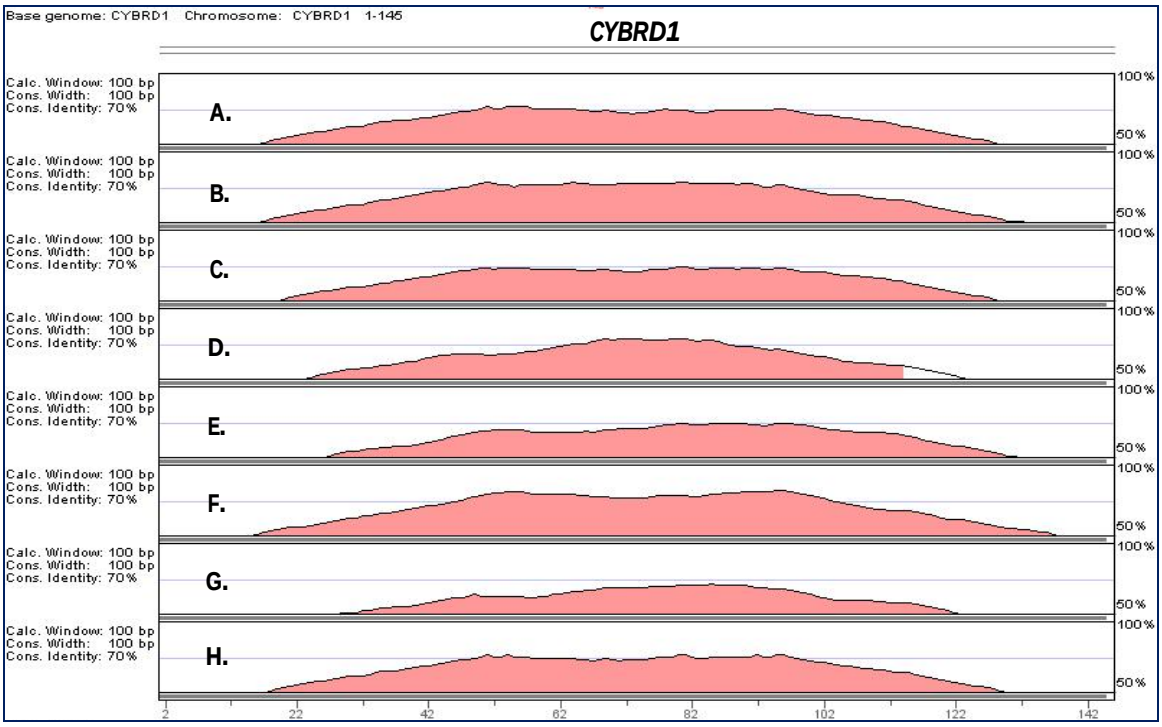


Figure 3.2 mVISTA sequence alignments of the CR in each of the nine iron metabolism genes.

Pink coloured areas indicate regions of sequence similarity. In this example *CYBRD1* was utilized as the base genome for visualization of the alignments against: **A**.) *FTH1*; **B**.) *HAMP*; **C**.) *HFE*; **D**.) *HFE2*; **E**.) *HMOX1*; **F**.) *IREB2*; **G**.) *LTF* and **H**.) *TFRC*.

Abbreviations: *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor 1 gene.

3.4.3 Pattern Matching

In order to generate a consensus CR sequence from each of the genes of interest, ClustalW (Thompson *et al.*, 1994) was utilized to produce a multiple alignment (Figure 3.3). The consensus sequence was prepared by hand due to the fact that the majority of commercially available resources require sequences of equal length, and the CR in this case differs by a few base pairs between the different genes. Base pairs that were responsible for the introduction of gaps in the multiple alignment (particularly in the case of *HFE2*), and therefore did not align with nucleotides in any of the other genes, were excluded from the consensus sequence. The consensus CR sequence was subjected to a nucleotide Basic Local Alignment Search (BLASTn) against the nucleotide (nr/nt) collection as well as the Human ALU repeat elements databases. The BLASTn search result with the lowest *E*-value (7×10^{-52}), the highest maximum identity (93%) and a total coverage of 92% was to the human *Alu*-J subfamily consensus sequence (Figure 3.4 A). This 290 bp sequence was retrieved from the nucleotide database on GenBank (accession number U14567) and utilized in a ClustalW Multiple alignment with the CR consensus sequence (Figure 3.4 B).

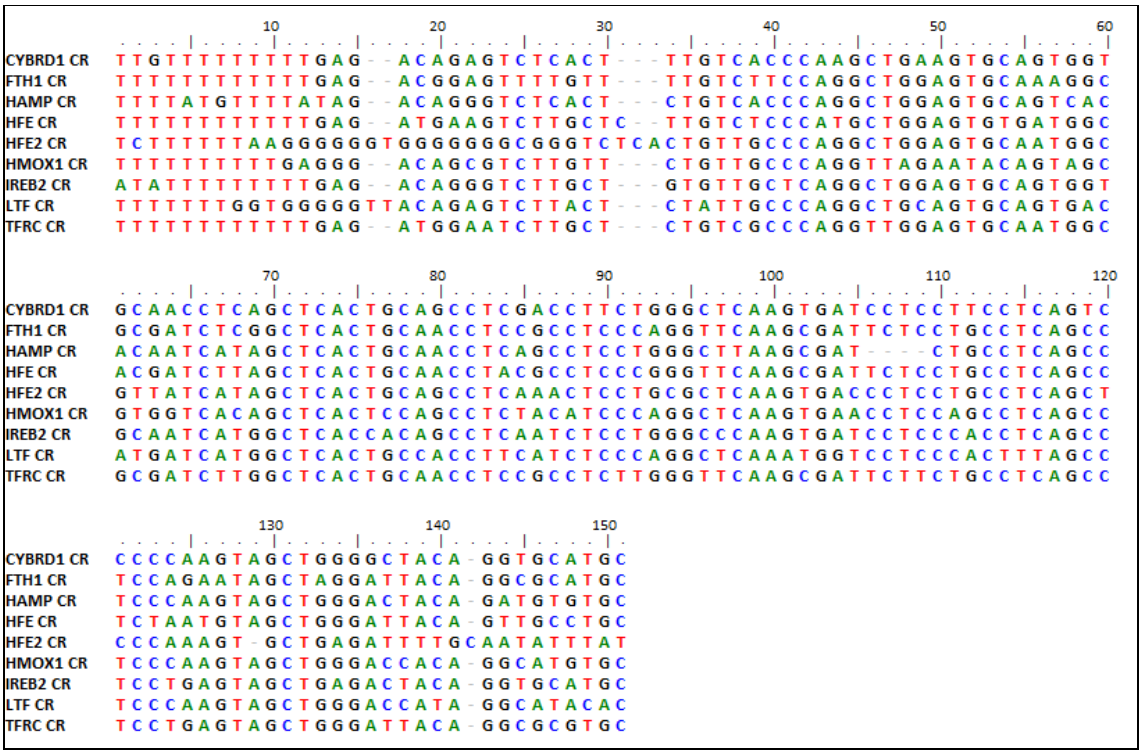


Figure 3.3 ClustalW multiple alignment of the CR of each of the nine iron genes.

Abbreviations: CR, conserved region; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

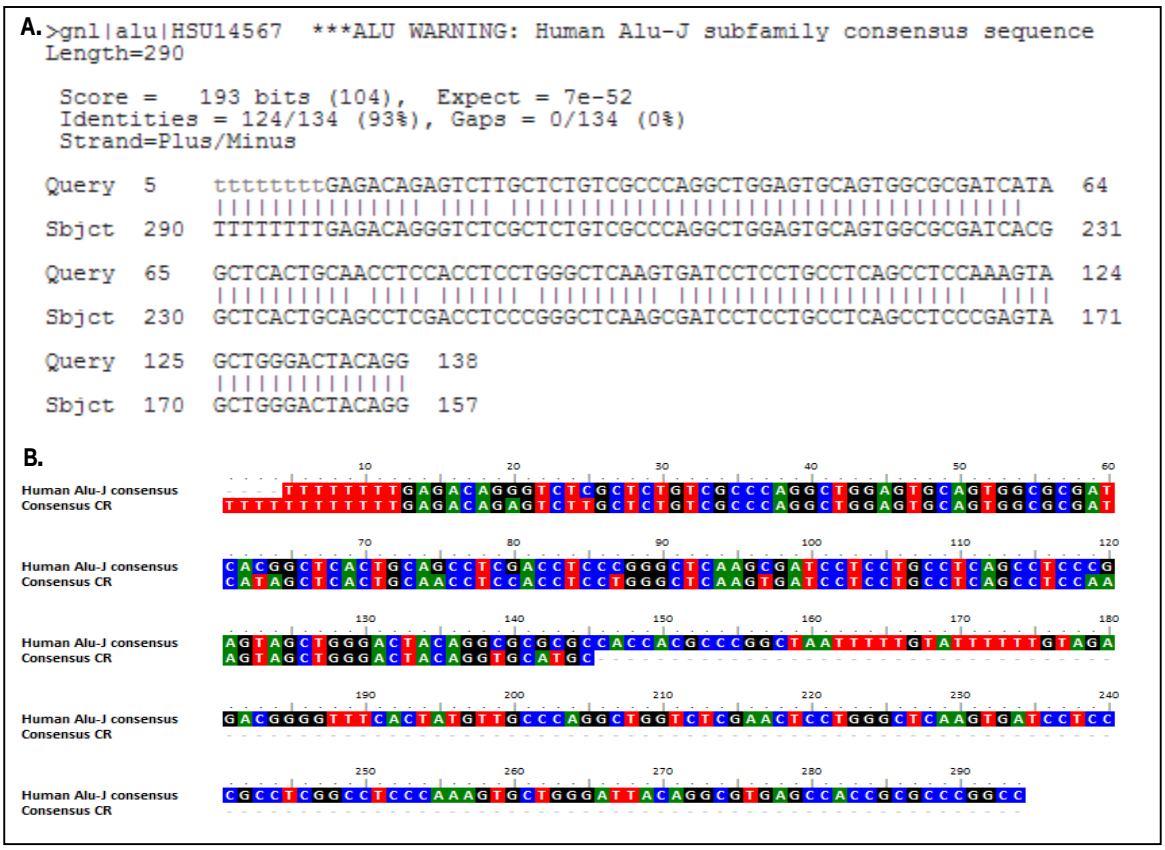


Figure 3.4 Alignment of the consensus CR with the human *Alu*-J consensus sequence. **A**.) BLASTn output of CR consensus query; **B**.) ClustalW alignment of CR consensus and full human *Alu*-J sequence retrieved from Genbank.

Abbreviations: CR, conserved region.

The CR region from each of the nine genes was submitted to the *in silico* program MEME (Multiple EM for Motif Elicitation) for identification of statistically significant, novel sequence signals or *cis*-regulatory modules (CRMs) residing within this area. MEME discovered the presence of four CRMs (Motif 1 to Motif 4) that were shown to be present in the same order, position and orientation in each of the nine CR sequences submitted. However, Motif 3 (red block) was not predicted to occur in the CR of *HFE2* or *LTF* (Figure 3.5). Combined *p*-values for the four CRMs were highly statistically significant for all nine CRs, and are illustrated in Figure 3.5. In addition, individual *p*-values for each motif in the nine CRs were calculated and were found to be statistically significant (Appendix 4: Figure S2, page 183).

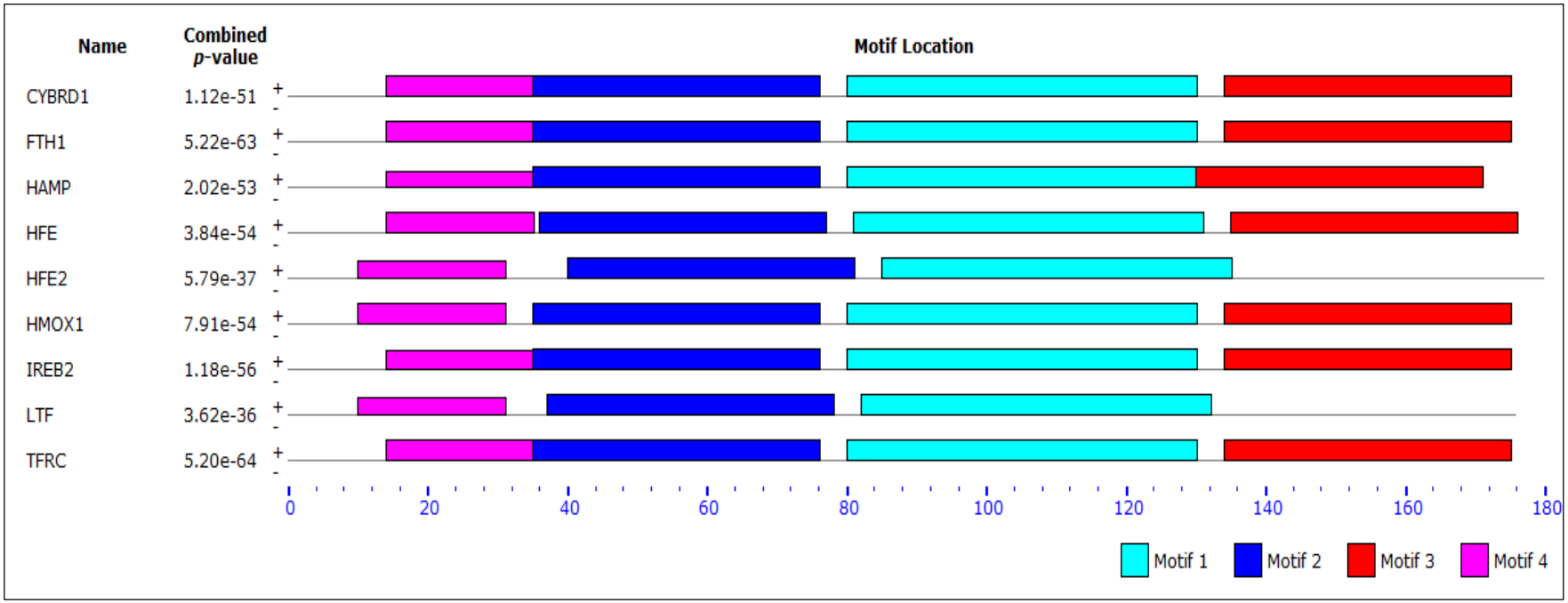


Figure 3.5 CRM discovery using MEME.

The CR of each gene is shown with the identified CRMs (Motif 1 to Motif 4) indicated by different coloured blocks. The genomic order and location of each motif is the same for each of the nine genes investigated. Combined *p*-value scores for the four motifs are listed.

Abbreviations: CR, conserved region; CRM, *cis*-regulatory module; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor 1 gene.

3.4.4 Pattern Recognition

The TOMTOM Motif Comparison Tool, within the MEME suite, was utilized to further analyze each motif individually in order to identify putative transcription factor binding sites (TfBSs). The JASPAR CORE vertebrates database was chosen as the target database for the identification of overrepresented TfBSs. Sequence logos corresponding to each of the four motifs, or CRMs, were generated by TOMTOM and used to align predicted TfBSs from JASPAR in the pertinent positions (Figure 3.6). Only TfBSs with a statistically significant *p*-value < 0.05 and an *E*-value < 10 were considered relevant and are illustrated by their own sequence logos, arranged in ascending order of significance, in Figure 3.6.

Due to the different computational algorithms employed by individual prediction software, identification of the same TfBS by more than one program increases the likelihood that the motif exists. The 140 bp CR from each of the nine genes was therefore analyzed with the PATCH and MATCH™ programs within the TRANSFAC®7 database, as well as rVISTA to compare the putative TfBSs resulting from the TOMTOM search of the JASPAR CORE database. For the PATCH and MATCH™ programs, the default settings and parameters were employed. For rVISTA, a subset of motifs were selected based on the results of the TOMTOM search, as well as known liver-specific TfBSs, and used in the analysis of the CR. The default core similarity values of 0.75 and matrix similarity values of 0.70 were utilized.

The results obtained through *in silico* analysis of the nine CRs are shown in Table 3.3. An rVISTA graphical representation of the results of the TfBS alignments with the consensus sequence generated from the nine CRs is shown in Appendix 4: Figure S3, page 184. The majority of TfBSs predictions were confirmed by at least two of the different programs utilized. The results from both MATCH™ and PATCH appeared to correspond more accurately with the results from rVISTA than with each other. MATCH™ predicted the presence of consensus binding sites for NKX2-5 (in the CR of *CYBRD1*, *HFE2* and *HMOX1*) and PAX4 (in the CR of *HFE* and *TFRC*), which was confirmed by the rVISTA program (Table 3.3).

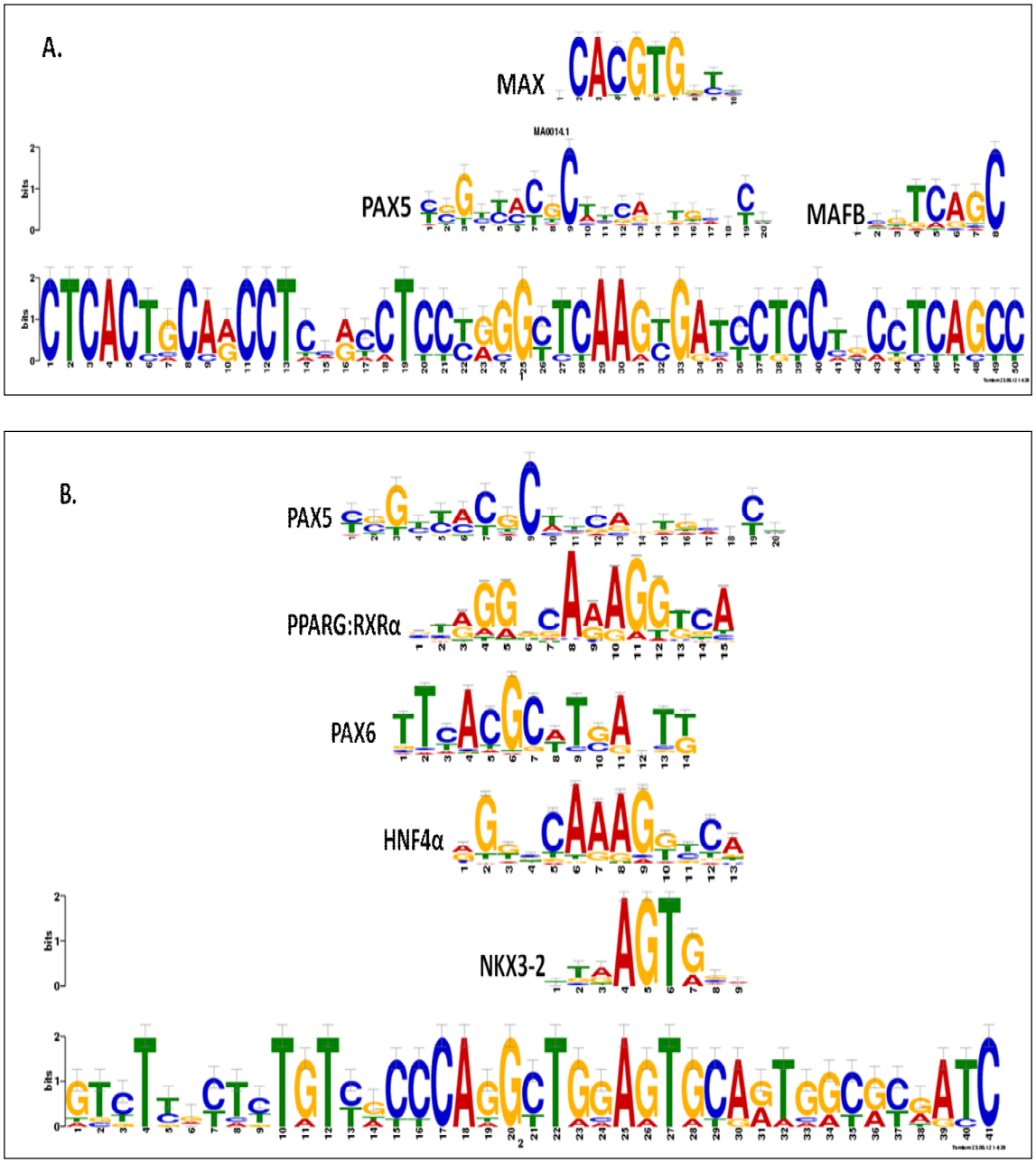


Figure 3.6 TOMTOM TfBS prediction for each CRM.

For each CRM, a sequence logo representing the individual motif sequence was aligned to sequence logos produced for the most statistically significant TfBSs. TfBSs names are indicated next to the relevant sequence logo, arranged in ascending order of significance. **A**.) Motif 1 and **B**.) Motif 2.

Abbreviations: CRM, *cis*-regulatory module; TfBSs, transcription factor binding sites.

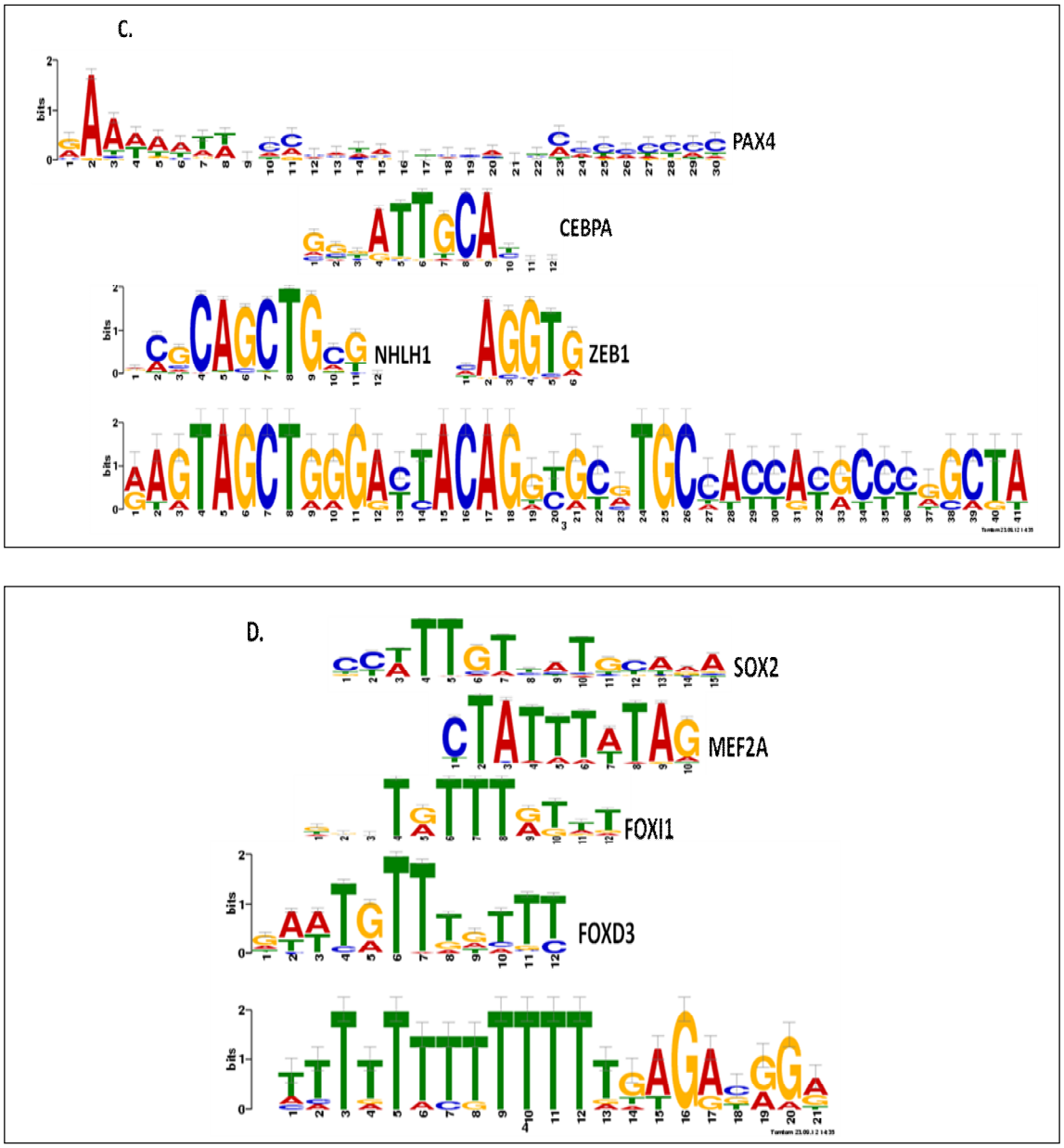


Figure 3.6 Cont. TOMTOM TfBS prediction for each CRM.

For each CRM, a sequence logo representing the individual motif sequence was aligned to sequence logos produced for the most statistically significant TfBSs. TfBSs names are indicated next to the relevant sequence logo, arranged in ascending order of significance. **C.**) Motif 3 and **D.**) Motif 4.

Abbreviations: CRM, *cis*-regulatory module; TfBSs, transcription factor binding sites.

Table 3.3 Predicted TfBSs in the CR of the iron metabolism genes.

Gene	Putative Transcription Factor Binding Sites		
	PATCH	**MATCH™**	**rVISTA**
CYBRD1	AP-1, c-MYC, GR, HNF-3α, HNF-3β, PAX2, PAX5, PAX8, RXR-α, RXR-γ, SP1, USF2	NKX2-5	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1
FTH1	AP-2, FOXO3a, HNF-3α, HNF-3β, MTF-1, RXR-α, RXR-β, RXR-γ, SP1, STAT5A, STAT5B	-	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF-1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1
HAMP	AP-1, AP-2, GR, HNF-3α, HNF-3β, MTF-1, PAX2, PAX5, PAX8, RXR-α, RXR-β, RXR-γ, SP1, STAT5A, STAT5B	-	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF-1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1
HFE	AP-1, AP-2, c-MYC, GATA-1, GR, HNF-1α, HNF-3α, HNF-3β, RXR-α, RXR-β, RXR-γ, STAT5A, STAT5B	PAX4	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF-1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1
HFE2	AP-2, GATA-1, GATA-2, GR, HNF-3α, HNF-3β, MTF-1, PAX2, PAX5, PAX8, RXR-γ, SMAD-3, SMAD-4, SP1, USF2	NKX2-5	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF-1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1
HMOX1	AP-1, AP-2, c-MYC, HNF-1α, HNF-1β, PAX2, PAX5, PAX8, RXR-α, RXR-β, RXR-γ, SP1, STAT5A, STAT5B	NKX2-5	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF-1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1
IREB2	AP-1, c-MYC, GATA-1, GR, HNF-3α, HNF-3β, MTF-1, PAX2, PAX5, PAX8, RXR-γ, SP1, USF2	-	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF-1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1
LTF	AP-1, AP-2, PAX2, PAX5, PAX8, RXR-γ, SP1, STAT5A, STAT5B, USF2	-	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF-1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1
TFRC	AP-1, GR, MTF-1, RXR-α, RXR-β, RXR-γ, SP1, STAT5A, STAT5B	PAX4	AP-1, AP-2, C/EBP, FOXD3, FOXM1, GATA-1, GATA-2, GATA-3, GR, HNF-1, HNF-3, HNF-4, MEF2A, MTF-1, NKX2-5, NKX3-2, PAX4, PAX6, PPAR, SMAD-1, SMAD-3, SMAD-4, SP1, STAT5A, USF, YY1

Abbreviations: AP-1, Activator protein 1; AP-2, Activator protein 2; C/EBP, CCAAT enhancer binding protein; c-MYC, Myelocytomatosis viral oncogene c; FOXM1, Forkhead box protein M1; FOXO3a, Forkhead box protein O3a; GATA-1, GATA-binding protein 1; GATA-2, GATA-binding protein 2; GATA-3, GATA-binding protein 3; GR, Glucocorticoid receptor; HNF-1α, Hepatocyte nuclear factor 1 alpha; HNF-1β, Hepatocyte nuclear factor 1 beta; HNF-3α, Hepatocyte nuclear factor 3 alpha; HNF-3β, Hepatocyte nuclear factor 3 beta; MEF2A, Myocyte-specific enhancer factor 2A; MTF-1, Metal transcription factor 1; NKX2-5, NK 2 homeobox 5; NKX3-2, NK 3 homeobox 2; PAX2, Paired box 2; PAX4, Paired box 4; PAX5, Paired box 5; PAX6, Paired box 6; PAX8, Paired box 8; PPAR, Peroxisome proliferator-activated receptor; RXR-α, Retinoid X receptor alpha; RXR-β, Retinoid X receptor beta; RXR-γ, Retinoid X receptor gamma; SMAD-1, Mothers against decapentaplegic homolog 1; SMAD-3, Mothers against decapentaplegic homolog 3; SMAD-4, Mothers against decapentaplegic homolog 4; SP1, Specificity protein 1; STAT5A, Signal transducer and activator of transcription 5A; STAT5B, Signal transducer and activator of transcription 5B; USF, Upstream transcription factor ; USF2, Upstream transcription factor 2; YY1, Ying Yang 1.

3.5 DISCUSSION

Promoter prediction software was utilized to assess the upstream sequences retrieved from the Ensembl database for 18 of the iron metabolism genes of interest in this study. Conventionally, promoter prediction has proved a stumbling block in the search for regulatory elements in non-coding eukaryotic sequences due to difficulties presented by the complexity of these regions and the limited power and accuracy of software programs to precisely identify promoter signals (Fickett and Hatzigeorgiou, 1997). However, experimental data from large scale ChIP- and microarray studies for the identification of TSS in human and mouse genomes has led to the availability of precise promoter sequences for numerous genes (deposited into databases such as EPD) and a subsequent improvement of the algorithms underlying promoter prediction software (Down and Hubbard, 2002; Okazaki, *et al.*, 2002). In this study, experimentally determined promoter sequences were obtained from EPD (Schmid *et al.*, 2005) and compared to predictions performed by Promoter 2.0 and NNPP, two of the most popular and frequently used programs for the identification of vertebrate PolII binding sites (Table 3.1). All of the 2 kb upstream sequences for the genes of the iron metabolism pathway were found to contain promoter elements essential for basal transcription levels and thus were considered to be true promoters. From the results, it is evident that the NNPP program performed more accurately than Promoter 2.0 when compared to the data from EPD (Table 3.1). Both prediction programs were able to identify regions that overlapped those contained within EPD, however those identified by NNPP were more exact in terms of the regions identified.

The promoter sequences from all of the genes were then submitted to mVISTA in order to search for regions of sequences similarity between the genes (pattern identification) which could be further utilized to search for conserved regulatory elements common to the iron metabolism pathway. mVISTA forms part of the VISTA family of comparative genomic tools and is a constantly maintained and updated resource (Frazer *et al.*, 2004). Traditionally, it is designed for the comparison and alignment of multi-species sequences, however, the AVID alignment engine used to generate the pairwise global alignments can be utilized for the evaluation of any subset of user-submitted sequences (Bray *et al.*, 2003). RepeatMasker is automatically used by the mVISTA server to mask repetitive sequences in the sequence

selected as the reference or "base genome". The results of the mVISTA alignment of the promoter sequences demonstrated a region of sequence conservation in nine of the 18 genes (*CYBRD1*, *FTH1*, *HAMP*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF* and *TFRC*) (pink region in Figure 3.1). Examination of the alignments between each of the genes showed that sequence similarity was between ~70.0 and 85.0% (Table 3.2, sequence similarities ≥ 75.0% are highlighted in red), indicating a high magnitude of sequence conservation of these areas.

In order to validate the mVISTA prediction, all 18 sequences were submitted to another alignment program, YASS (Noe and Kucherov, 2005). Comparison of the regions of sequence similarity predicted by mVISTA and YASS (Table 3.2) illustrated that both programs were able to identify the same genomic area when each of the genes were aligned to each other in turn. The *E*-values generated by YASS for each alignment were highly statistically significant (Table 3.2) although the alignments were slightly longer than those obtained from mVISTA.

The conserved region, predicted by mVISTA and YASS, from each of the nine genes was examined visually in order to obtain a genomic region which was representative for each of the genes and excluded erroneous flanking sequences that may be common to some of the genes, but not all. This conserved region (termed the CR) includes 140 bp of sequence from each gene (highlighted in green in Appendix 2) representing a high level of sequence conservation which has implications for regulatory control mechanisms of these genes. The nine CRs were again submitted to mVISTA for alignment of the sequences and a visualization of the results (utilizing *CYBRD1* as the reference) is presented in Figure 3.2. A clear conservation (pink area) of the CR is visible in the results, once again confirming the strength of similarity of the identified CR.

The results of the pattern discovery portion of the *in silico* analysis represent a novel finding in the 5'-UTR of the iron metabolism pathway which is worthy of further investigation on the premise that genes which share similar promoter features, and are involved in a common pathway, may be co-regulated (Wasserman and Sandelin, 2004).

A BLAST search of the consensus sequence generated from the CR of each gene (Figure 3.4) demonstrated a significant (*E* value = 7×10^{-52}) hit to the GenBank human *Alu*-J subfamily consensus sequence. *Alu* elements are repetitive elements that arose from the fusion of the 7SL RNA gene approximately 65 million years ago (Batzer and Deininger, 2002). Although

not yet fully elucidated, it is believed that *Alu* elements amplified across primate genomes through a retrotransposition mechanism *via* RNA intermediates (Rowold and Herrera, 2000). *Alu* elements are ~300 bp in length and are located primarily in AT-rich areas of the genome. They are the most abundant member of the SINE (Short Interspersed Nuclear Elements) family of repetitive elements and more than a million copies are present throughout the human genome, comprising 10% of total genomic mass (Häsler and Strub, 2006). These mobile elements are grouped into two subfamilies based on their relative ages, the *Alu*-J and *Alu*-S subfamilies (Jurka and Smith, 1988). Due to their profusion within human and primate genomes, *Alu* elements are thought to have played a significant role from an evolutionary perspective, both in a positive and negative manner. Due to their conserved sequence homology, *Alus* have been proposed to provide regulatory elements to neighbouring genes. They have been identified in numerous 5'- and 3'-UTRs gene transcripts and are transcribed by RNA polymerase II to form part of the resulting mRNAs. They act as enhancers, repressors and may present novel promoter elements to modulate gene transcription and mRNA translation (Brosius, 1999; Tomilin, 1999). Several studies have implicated UTR (either 5' or 3') *Alu* repeats in the regulation of transcription in specific gene transcripts; human growth hormone receptor (hGHR) (Goodyer *et al.*, 2001), manganese superoxide dismutase (MnSOD) (Stuart *et al.*, 2000) and the best described example, *BRCA1*, in which a 5'-UTR *Alu* element in the normal *BRCA1* transcript results in translational inhibition (Sobczak and Krzyzosiak, 2002). In addition, *Alu* elements have been shown to contain binding sites for various Tfs and are able to interact with Tfs *in vivo* to regulate the transcription of genes (Polak and Domany, 2006). Polak and Domany identified numerous Tfs, able to bind to *Alu* elements, with particular relevance to this study including, NKX2-5, HNF-4, MEF2A, HNF-1, PAX4, PAX6 and YY1. Binding sites for these Tfs were all predicted to be present in the CR of the nine genes investigated in this study, therefore adding credibility to the proposition that the CR may potentially be comprised of an *Alu* repeat. Although *Alu* elements have not yet been described for the iron metabolism pathway, the level of sequence conservation between the CR consensus sequence and the *Alu*-J subfamily consensus identified in this study has potential implications for a possible mechanism of co-ordinated regulation. It is probable that the *Alu* element identified here may represent a reservoir of TfBSs for specific Tfs, which could act to modulate the transcriptional activation

or repression of this subset of the iron metabolism genes, therefore representing a novel regulatory mechanism.

Investigation of the CR of the nine iron genes utilizing MEME (pattern matching) demonstrated that four different sequence motifs (Motif 1 – Motif 4) were present in this 140 bp sequence (Figure 3.5). Each of these CRMs was statistically significant on an individual basis, as well as when combined together (Figure 3.5 and Appendix 4: Figure S2, page 183). Of interest, was the fact that each of these four CRMs was located in the same genomic location, orientation and order in each of the nine genes, with the exception of *HFE2* and *LTF* which were found not to possess Motif 3 (red block, Figure 3.5) at the 3' end of the CR. This CRM conservation strongly suggests a functional role for the CR in each of the genes examined. It is however imperative to assess the *in vitro* functionality of this region in order to substantiate the bioinformatic predictions.

In the context of gene regulation CRMs represent multiple binding sites for numerous Tfs thereby enabling a superior level of transcriptional control (Blanchette *et al.*, 2006). CRMs are responsible for the activation and repression of transcriptional initiation by mediating the binding of specific Tfs in different combinations depending on the needs of a particular cell in response to a variety of stimuli (so-called "molecular switches"). Multiple bound Tfs therefore have the ability to act in a combinatorial manner to alter transcriptional activity and genes regulated by a common set of Tfs tend to be co-expressed (Gupta and Liu, 2005).

The characteristic features of CRMs described can be exploited in the search for regulatory elements for specific Tfs (Bulyk, 2003). *In silico* analysis of the CR identified in this study demonstrated the presence of numerous putative TfBSs (pattern recognition) (Table 3.3). Only Tfs that were predicted to be conserved across all CRs examined or that are biologically relevant (involved in iron metabolism or liver-specific) will be discussed further. The program TOMTOM, part of the MEME suite, predicted potential TfBSs for each of the four identified CRMs comprising the CR (Figure 3.6). TOMTOM utilizes the JASPAR CORE vertebrates database to search for conserved binding sites, and only statistically significant ($p < 0.05$) Tfs were selected from the results generated from the investigation. In addition, other bioinformatic analyses for the identification of putative TfBSs were performed with the aim to increase the likelihood of the predictions and therefore the potential functional

relevance. Results from a TRANSFAC®7 (utilizing the PATCH and MATCH™ programs) and rVISTA search (Table 3.3) were compared to the results obtained from TOMTOM.

With respect to Motif 1, TOMTOM predicted the following TfBSs with *p*-values indicated: MAFB ($p = 0.01$), PAX5 ($p = 0.03$) and MAX ($p = 0.04$) (Figure 3.6 A). After analysis with the other TfBS prediction programs, members of the PAX family of Tfs were the only TfBS to be identified by PATCH, MATCH and rVISTA (Table 3.3). PAX (paired box) factors are a class of basic helix-loop-helix (bHLH) Tfs that are highly conserved in vertebrates (Wang *et al.*, 2010). There are nine members of the PAX family assembled into four distinct groups: PAX1 and PAX9 (PAX group I); PAX2, PAX5 and PAX8 (PAX group II); PAX3 and PAX7 (PAX group III); PAX4 and PAX6 (PAX group IV). PAX5 encodes a B-cell specific activator protein (BSAP) and is implicated in the differentiation of B-cells during early stage delineation (Adams *et al.*, 1992). The PAX Tfs are known to be involved in cell development and organogenesis, although little is known about their specific target genes (Lang *et al.*, 2007). Due to the ubiquitous distribution of the PAX factors in vertebrate organisms, they are considered to be universal master gene regulators with the capability to bind to the upstream regions of a multitude of genes (Buckingham and Relaix, 2007). PAX Tfs have not been extensively described with regard to regulation of genes involved in iron metabolism however a study by Chaudhary *et al.* (1997) revealed the presence of E-box motifs in the *transferrin* (*TF*) promoter in Sertoli cells, capable of binding bHLH Tfs resulting in elevated expression of *TF*. Expression was abrogated upon inhibition of bHLH factors indicating the importance of these Tfs in the regulation of *TF* expressed in this cell type. Although *TF* was not found to contain the CR of interest in this study, the fact that bHLH TfBSs have been identified in this gene and are actively involved in regulation of expression indicates that PAX Tfs may have a defined role in regulation of other genes involved in iron metabolism. PAX (PAX2, PAX4, PAX5, PAX6 and PAX8) Tfs were consistently identified by all *in silico* tools utilized in this study (Table 3.3), across the different motifs comprising the CR, and therefore warrant further investigation in this metabolic pathway. In addition, it remains to be elucidated if PAX and other bHLH Tfs are actively involved in gene regulation in cells of the liver, intestine and macrophages which are the primary sites of expression of iron metabolizing genes. Traditionally PAX factors have been found to be involved in gene regulation primarily in cells of the thyroid, kidney (Poleev *et al.*, 1995) and pancreas (Brun *et al.*, 2005) and appear to exert their functions during early stage cell differentiation and development, with a less

defined role in adult tissues (Calame *et al.*, 2003). Due to a lack of substantial evidence for the involvement of PAX factors in the regulation of iron gene expression, functional investigation would be required to validate the *in silico* predictions from this study.

A TOMTOM search of Motif 2 revealed putative TfBSs for NKX3-2 (p = 0.02), HNF-4α (p = 0.03), PAX6 (p = 0.03), PPARγ-RXRα (p = 0.04) and PAX5 (p = 0.05) (Figure 3.6 B).

In conjunction with the PAX factors discussed above, a HNF-4α binding site was predicted by TOMTOM as well as rVISTA (Table 3.3). The HNF (hepatocyte nuclear factor) Tfs are expressed predominately in the liver in large quantities and are known to regulate liver-specific expression of target genes in a synergistic manner by binding to enhancer or repressor elements in the promoter region. Members of the HNF family are phylogenetically unrelated and are grouped into four subfamilies based on their functional characteristics and interactions (Cereghini, 1996). HNF-4α is a member of the hepatocyte nuclear family 4 (HNF-4) of Tfs that are characterized as nuclear hormone receptors. The HNF-4 nuclear hormone receptors exert their function *via* DNA-binding as a homodimer or as a heterodimer with members of the RXR (retinoid X receptor) family of Tfs (Li *et al.*, 2000). Interestingly, binding sites for RXR-α, RXR-β, RXR-γ in the CR were predicted by the PATCH program in this study (Table 3.3), and may therefore indicate a functional role for HNF-4α in this pathway. Furthermore, functionally confirmed binding sites for HNF-4α have been identified in the hepcidin promoter (Courselaud *et al.*, 2002) and were associated with negative regulation of gene expression during *in vitro* assays. Furthermore, PATCH and rVISTA predicted binding sites for other HNF Tfs, specifically HNF-1α, HNF-1β, HNF-3α and HNF-3β (Table 3.3) in the CR, therefore increasing the likelihood of functional regulation by these Tfs. Genes of the iron metabolism pathway, particularly hepcidin, are expressed largely in hepatocytes and therefore gene regulation by the HNF family of Tfs is not unexpected and requires further examination.

In addition, TOMTOM predicted the presence of a PPARγ TfBS in Motif 2 of the CR. Conservation of this TfBS across the CR of the nine genes was confirmed by analysis with rVISTA (Appendix 4: Figure S3, page 184). PPARs (peroxisome proliferator-activated receptors) are a group of nuclear receptor proteins that are involved in the regulation of cellular metabolism and differentiation in eukaryotic organisms (Feige *et al.*, 2006). PPAR

proteins form heterodimers specifically with RXRs facilitating DNA-binding to PPREs (peroxisome proliferator hormone response elements) with the consensus sequence -AGGTCANAGGTCA- (Michalik *et al.*, 2006). Numerous studies have demonstrated an association with iron and PPAR, particularly in reference to hepatic insulin resistance and liver disease (Silva *et al.*, 2010; Kell, 2009; Valenti *et al.*, 2007). Excess iron has been shown to result in an increase in triglyceride levels mediated by a decrease in PPAR interaction (Silva *et al.*, 2008), therefore illustrating the relationship between iron, PPAR and hepatic lipid metabolism. In addition, a number of pharmacological agents are utilized for the activation of PPAR (particularly PPARα) which result in lower levels of hepatic triglycerides and are used in the treatment of chronic obesity and liver disease (Rakhshandehroo *et al.*, 2007; Ip *et al.*, 2004). As in the case of HNF-4α, the presence of numerous RXR TfBSs predicted by PATCH (Table 3.3) strengthens the potential association of PPAR with regulation of iron-metabolizing genes.

TfBSs predicted for Motif 3 included NHLH1 (p = 0.01), ZEB1 (p = 0.02), C/EBPα (p = 0.03) and PAX4 (p = 0.05) (Figure 3.6 C). TOMTOM predictions of C/EBPα TfBSs were confirmed by rVISTA (Table 3.3). C/EBPα is a bZIP (basic leucine zipper) Tf that binds to gene promoters as a homodimer or as a heterodimer in association with other C/EBP proteins. Sequence-specific DNA-binding of C/EBPα, followed by the recruitment of co-activators such as CBP (CREB1-binding protein) and basal Tfs, activates transcription of a variety of liver-specific genes implicated in cell proliferation, differentiation and metabolism (Kovács *et al.*, 2003). C/EBP binding sites are characteristic of CRM architecture as a result of their innate ability to recruit additional Tfs, specifically the leucine zipper domain-containing family (eg. c-FOS and JUN) (Shaulian and Karin 2002). C/EBP Tfs have been associated with the regulation of hepatic iron, nitrogen and lipid metabolism (Pedersen *et al.*, 2007) and it is therefore feasible that potential C/EBPα TfBSs reside within the CR of the genes of interest in this study. The hepcidin promoter has been experimentally shown to contain binding sites for C/EBPα that are associated with enhanced hepatocyte-specific promoter activity in both human and mouse *via* direct interaction of C/EBPα with recognition sites in the hepcidin promoter (Courselaud *et al.*, 2002).

The *in silico* investigation of Motif 4 revealed putative binding sites for FOXD3 (p = 0.00), FOXI1 (p = 0.01), MEF2A (p = 0.03) and SOX2 (p = 0.04) (Figure 3.6 D). Within this motif, FOX (forkhead box) Tfs were overrepresented due to their affinity for recognizing T-rich stretches of DNA characteristic of Motif 4 at the 5'-end of the CR (see Appendix 2 for CR sequences or Appendix 4: Figure S2 D, page 183). The FOX family are a large group of functionally diverse Tfs that are classified phylogenetically according to structure into 15 distinct subfamilies (Coffer and Burgering, 2004). They form part of the bHLH class of Tfs and are composed of a winged helix domain that comprises the distinctive forkhead motif of these proteins (Katoh and Katoh, 2004). TOMTOM analysis of Motif 4 revealed the presence of two FOX family members, FOXD3 and FOXI1, which were substantiated by rVISTA (Table 3.3). Although literature pertaining to the relationship between FOX factors and iron metabolism is scarce, a large-scale microarray study of differentially expressed genes under conditions of iron deficiency resulted in the identification of conserved SP1 (specificity protein 1) and FOX TfBSs in the promoters of known iron-metabolizing genes (Collins and Hu, 2007). With particular relevance to this study was the discovery that *CYBRD1*, *HMOX1*, *SLC11A2* and *TFRC* were upregulated under conditions of iron deficiency in rat intestinal tissue and were enriched for SP1 and FOX Tfs. The authors hypothesize that SP1 and FOX may synergistically regulate transcription of a subset of the genes induced by iron deficiency in rat, mouse and human. In this study, SP1 TfBSs were identified in the CR of all nine iron-metabolizing genes following *in silico* analysis with PATCH and rVISTA (Table 3.3), indicating a potential functional regulatory role, in combination with the FOX TfBSs predicted in Motif 4, as suggested by Collins and Hu. In humans, SP1 is part of a large family of zinc finger Tfs involved in a wide variety of cellular processes and is among the most potent activators of transcription identified to date.

In addition to the TfBSs already discussed, *in silico* analyses predicted putative binding sites for MTF-1 (metal transcription factor 1) in the CR of the nine genes of interest (Table 3.3). This Tf is of particular interest in this subset of genes due to the recent discovery of MTF-1 TfBSs in the hepcidin promoter (Balesaria *et al.*, 2010). MTF-1 proteins are divalent metal ion sensitive Tfs that mediate their regulatory effect by binding to cognate MREs (metal responsive elements) in the promoter region of target genes (Günther *et al.*, 2012). The study by Balesaria and colleagues identified four MRE motifs in the hepcidin promoter, one

of which lies within the CR element identified in this study. Hepcidin expression levels were shown to increase in response to MTF-1 binding activated by the addition of zinc. It has been shown that zinc activates DNA binding of MTF-1 by reversible interactions with specific zinc finger domains (Dalton *et al.*, 1997; Bittel *et al.*, 1998). In addition to hepcidin, *SLC40A1* expression is directly induced by MTF-1 binding (Troadec *et al.*, 2010). The previously identified hepcidin MRE corresponds to the recognition sequence of -GAGTGCAG- in the CR identified in this study, which is in the reverse orientation to the consensus MRE sequence, -TGCRCNC- (where R stands for A or G and N for any of the four bases). The other eight CR-containing genes (Appendix 2) were examined and identical recognition sites were found in reverse orientation in the CR (corresponding to Motif 2) of *FTH1*, *HFE2*, *IREB2* and *TFRC*. Despite an in-depth literature search, information regarding MTF-1 associations with these genes could not be found, indicating the uniqueness of this discovery.

Tfs such as USF1 (upstream regulatory factor 1) (Andrews *et al.*, 2001) and C/EBPα (Datta *et al.*, 2007) have been previously reported to interact with MTF-1 in order to mediate transcriptional regulation. Binding sites for both of these Tfs were identified with the PATCH and rVISTA computational tools (Table 3.3) and it is therefore feasible to consider that regulation of the CR-containing genes may be controlled by MTF-1 in conjunction with USF1 and C/EBPα.

3.6 CONCLUSION

The search for regulatory elements within non-coding regions of the human genome has traditionally proved an immensely difficult task. The expanding field of bioinformatics has lead to a vast improvement in the *in silico* tools available for investigation of potential functionally relevant regulatory regions. These software tools present a cost- and time effective method to explore the genome for transcriptional control elements, and have the potential to serve as a starting point for the design of experimental methodologies.

In silico analysis of the promoter region of 18 of the genes implicated in iron metabolism resulted in the identification of a novel region of sequence conservation in nine of the genes examined. Furthermore, the identified CR is proposed to be comprised of an *Alu*-J transposable element that may harbour numerous TfBSs thereby contributing toward co-ordinated transcriptional regulation of this subset of genes. Numerous TfBSs were

identified upon *in silico* analyses which could be credibly associated with transcriptional regulation in a tissue- and pathway-specific manner. Further functional investigation is warranted to determine if the CR element contributes toward co-ordinated transcriptional regulation of the genes implicated in iron metabolism *in vitro*, as suggested by the unexpectedly high level of sequence similarity.

Taken together, the results presented here represent a comprehensive and novel computational investigation of iron gene promoter regions with the identification of *cis*-regulatory motifs that have the potential to serve as targets for future experimental investigations.

CHAPTER FOUR

Functional analyses of promoter regulatory targets in the iron metabolism pathway

4.1 ABSTRACT

Previously, computational analysis of the upstream non-coding region of genes involved in the iron metabolism pathway resulted in the identification of a novel region of sequence similarity in a subset of the genes. Validation of the bioinformatic predictions is essential in order to fully assess the relevance of the results in an *in vitro* setting. The nine CR-containing genes were functionally investigated following the design of luciferase reporter constructs containing: 1) the 2 kb promoter, 2) a 1.86 kb promoter with the CR removed and 3) the 140 bp CR element. These three reporter gene constructs were transfected into HepG2 and COS-1 cell lines and expression levels were monitored with a dual-luciferase reporter assay under standard culture conditions and simulated iron overload conditions. Results of the luciferase assays indicate that the 1.86 kb CR-removed promoter constructs displayed variances in expression values when compared to the untreated control construct. Further, the CR appears to mediate transcriptional regulatory effects *via* an iron-independent mechanism. With a few exceptions, the trends in expression observed for each of the constructs for the respective genes were consistent in both of the cell lines utilized and represent repeatable results across the different experiments performed. It is therefore apparent that the bioinformatic predictions were shown to be functionally relevant in this study and warrant further investigation.

4.2 INTRODUCTION

The wealth of sequence information available from high-throughput sequencing projects of numerous eukaryotic genomes have resulted in projects intent on annotating the functional elements responsible for gene expression. The majority of these groups employ a combinatorial bioinformatic and experimental approach to investigate regulatory components (Pilpel *et al.*, 2001; Pedro, 2002; Wilson *et al.*, 2010). One of the largest is the ENCODE (ENCyclopedia of DNA Elements) Project launched by the National Human Genome Research Institute (NHGRI) in 2003, following the sequencing of the human genome (The ENCODE Project Consortium, 2004). Recent results have shown that over 80% of the human genome is comprised of these functional elements; an exponentially higher amount than was earlier predicted (The ENCODE Project Consortium, 2012). In addition, it was found that numerous SNPs identified in previous genome-wide association studies (GWAS) on different

human diseases mapped to these areas regulating gene expression or transcription (Maurano *et al.*, 2012; Schaub *et al.*, 2012). The results are representative of a shift from traditional investigations of disease-causing variation and are indicative of the complexity of regulation of gene expression.

Utilizing a similar combinatorial approach, comprehensive computational analyses of the promoter and flanking regions of genes implicated in iron metabolism revealed the presence of a region of sequence conservation (termed the CR) in nine of the genes investigated in this study (Chapter 3). This region demonstrated greater than 75% sequence identity between the genes of interest and, based on sequence homology, was proposed to be an *Alu* element comprising four CRMs. *Alu* elements have been previously been shown to contain binding sites for numerous Tfs (Polak and Domany, 2006) and have been associated with transcriptional regulation in a number of instances (Goodyer *et al.*, 2001; Sobczak and Krzyzosiak, 2002; Stuart *et al.*, 2000). The CR element was therefore hypothesized to act as a novel target for Tf binding resulting in a level of co-ordinated gene expression *in vivo*. The use of bioinformatic software for the prediction of regulatory elements in eukaryotic organisms has the potential to be an effective tool for the study of transcriptional regulation. However, it is imperative to validate *in silico* results in an experimental manner as computational analyses are by nature predictive and do not serve to characterize the presence or function of regulatory targets *in vivo*. Numerous experimental assays are available for functional validation, the choice of which depends on the nature of the research to be conducted. Examples include: electrophoretic mobility shift assays (EMSA) (Fried and Crothers, 1981), DNase I footprinting (Brenowitz *et al.*, 1986), real-time PCR (qPCR) (VanGuilder *et al.*, 2008), chromatin immunoprecipitation (ChIP) (Collas, 2010) and reporter genes assays (Wasserman and Sandelin, 2004). In an attempt to elucidate the potential effect of the CR with regard to differential gene expression, functional analyses were performed in this study. Expression levels of each of the nine CR-containing iron genes were investigated under normal culture conditions and iron overload conditions *in vitro* utilizing a reporter gene assay system. Reporter gene assays are amongst the most versatile techniques available and are frequently utilized to assess gene expression under a variety of conditions and in different cell types (Naylor, 1999). Reporter genes themselves are numerous and can be tailored to suit the type of assay to be performed (Alam and Cook,

1990). In this study, the *firefly* luciferase reporter gene was chosen to quantitatively measure expression of the different promoter constructs that were assembled.

4.3 MATERIALS AND METHODS

The reader is referred to Chapter 2 (Sections 2.2 and 2.3) for further information regarding the detailed methodology utilized.

4.3.1 Construct Preparation

For each of the nine genes (*CYBRD1*, *FTH1*, *HAMP*, *HFE*, *HFE2*, *HMOX1*, *IREB2*, *LTF* and *TFRC*) examined, three different reporter gene constructs were prepared. The full 2 kb promoter region as well as the 140 bp CR element were subjected to PCR utilizing flanking primers designed to incorporate appropriate restriction sites at the 5'-end (Chapter 2, Tables 2.1 and 2.2). A promoter, in which the 140 bp CR was deleted from the full 2 kb 5'UTR, was constructed utilizing an adaptation of the PCR-driven overlap extension technique (Heckman and Pease, 2007). The 1.86 kb CR-removed fragments were generated from a nested PCR of the 2 kb promoter for each gene utilizing different primer combinations (Chapter 2, Tables 2.3 and 2.4). The PCR products were sent for semi-automated bidirectional sequencing in order to confirm the sequence of interest.

Following sequencing, the 2 kb and 1.86 kb PCR products were cloned into a promoterless *firefly* luciferase reporter vector, pGL4.10[*luc2*], and the 140 bp CR PCR fragments were cloned into the minimal promoter pGL4.23[*luc2*/minP] vector (Promega) (Appendix 3) using standard methodologies. The pGL4.10[*luc2*] vector was selected due to the fact that it does not contain any basal promoter or enhancer elements and the genomic sequence of interest is cloned immediately upstream of the luciferase gene. The 2 kb and 1.86 kb CR-removed promoter constructs were shown to contain the core promoter for each gene, based on the *in silico* investigation, and therefore do not require any additional promoter elements to drive transcription. In contrast, the 140 bp CR element for each of the genes was shown to lie upstream of the core promoter and it is unlikely that this relatively small genomic region contains basal transcriptional factors. For this reason, the pGL4.23[*luc2*/minP] vector was chosen in order to clone these fragments. The

pGL4.23[*luc2*/minP] vector includes a minimal promoter region with a TATA-box capable of initiating basal transcription of the luciferase gene.

After competent cell transformation and plasmid extraction, positive colonies were selected and again subjected to sequencing analysis to verify insert inclusion and orientation.

4.3.2 Reporter Gene Assays

The prepared luciferase reporter vectors were analyzed using the Dual-Luciferase® Reporter Assay System (Promega) in order to determine their respective effects on gene expression. Each reporter construct was transfected into two mammalian cells lines – human hepatocarcinoma liver (HepG2) cells and African green monkey kidney (COS-1) cells (ECACC, Wiltshire, England) – cultured under standard conditions. Co-transfection with a control reporter vector (pGL4.73[*hRluc*/SV40], Promega Corporation) was conducted for normalization of luciferase activity to correct for differences in transformation efficiency. Conditions of iron overload were simulated by the addition of ferric ammonium citrate (FAC) as an exogenous stimulus to assess the effect of high iron levels on expression levels of the relevant promoter constructs. For each cell line, assays were performed in triplicate for each individual construct and experiments were repeated on three independent occasions.

4.3.3 Statistical Analyses

For each individual experiment, the *firefly* luciferase values obtained for each construct were divided by the *Renilla* control luciferase readings to obtain normalized values, which were then averaged. The normalized averaged values were divided by the average normalized value of the construct acting as a control (grey shaded boxes, Appendix 5: Table S1.1 to S9.4, page 190) to determine the fold change in expression of the respective constructs for each experiment.

GraphPad Prism® version 5.10 (GraphPad Software) was utilized to analyze the data and to generate graphical representations of the results. One-way ANOVA and Dunnett's multiple comparison post-test were used for statistical analysis. In instances where statistically significant ($p < 0.05$) differences were detected by ANOVA, paired *t* tests were performed in

order to determine variance. For all experiments error bars represent the SEM of three independent experiments.

4.4 RESULTS

4.4.1 PCR Amplification

PCR amplification of the 2 kb promoter fragments, the 140 bp CR element and the 1.86 kb CR-removed promoter were performed as described in Chapter 2, Section 2.2.1. Results of the agarose gel electrophoresis of the PCR amplicons for the 2 kb promoter fragments and the 140 bp CR element are illustrated in Appendix 5, Figure S5, page 186. Appropriate fragment sizes are listed in Tables 2.1 and 2.2 respectively. The results of the PCR-driven overlap extension technique utilized to generate the 1.86 kb CR-removed promoter amplicons are illustrated in Figures 4.1 and 4.2. Respective amplicon sizes are listed in Table 2.4. Due to difficulties experienced when attempting to optimize the PCR-driven overlap extension, fragments of the correct size were excised from the agarose gel and used in a Wizard® SV Gel and PCR Clean-up (as per Chapter 2, Section 2.2.3). This was followed by semi-automated DNA sequencing analysis as described in Chapter 2, Section 2.2.5.

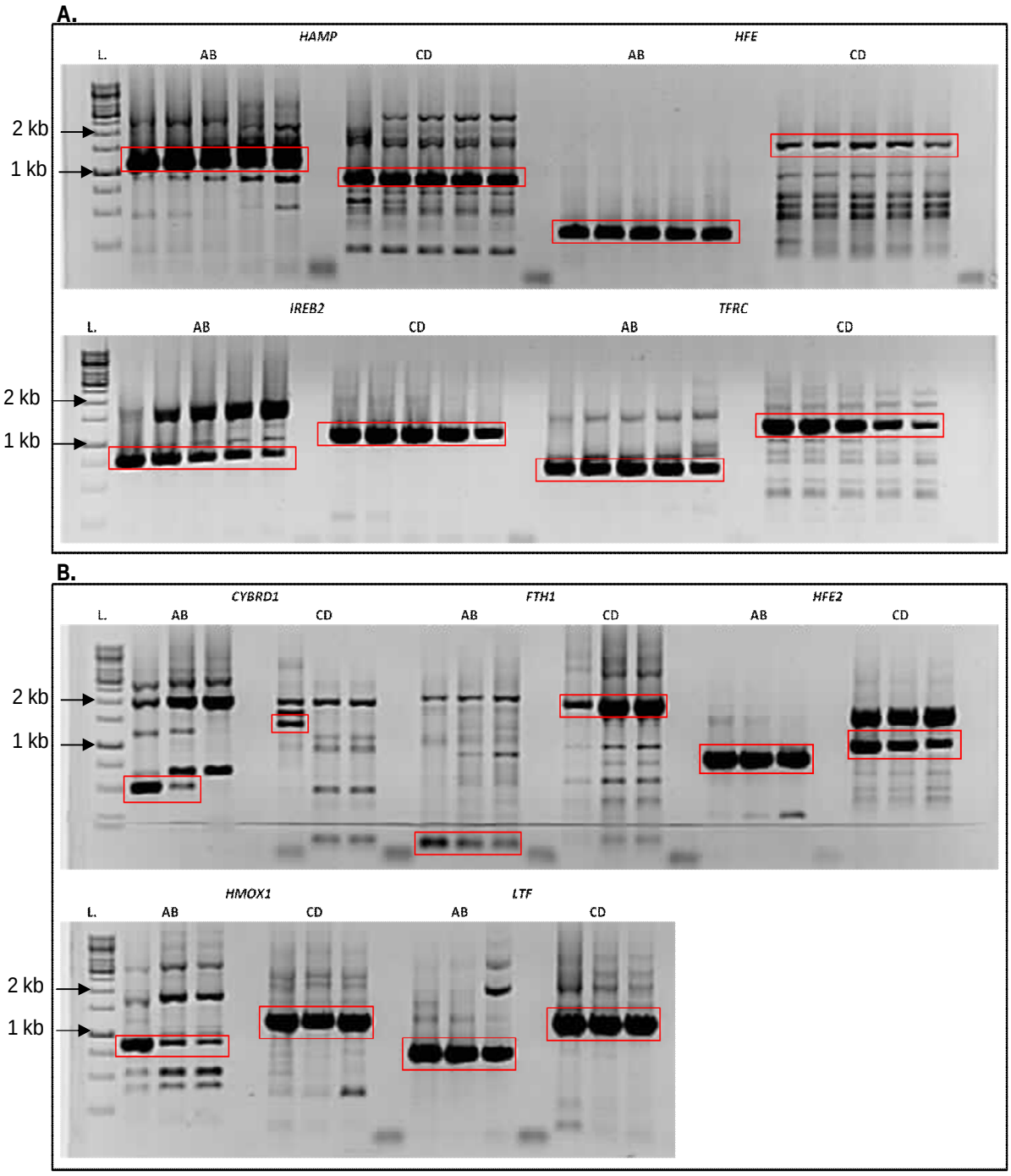


Figure 4.1 Agarose gel electrophoresis of intermediate PCR-driven overlap extension products (AB and CD). PCR products highlighted in red were excised from the agarose gel and used in a Wizard® SV Gel and PCR Clean-up. For each gene, **A.**) five or **B.**) three samples were amplified for AB and CD and electrophoresed adjacent to each other.

Abbreviations: bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; kb, kilobase pair; L., Ladder; *LTF*, Lactotransferrin gene; PCR, polymerase chain reaction; *TFRC*, Transferrin receptor protein 1 gene.

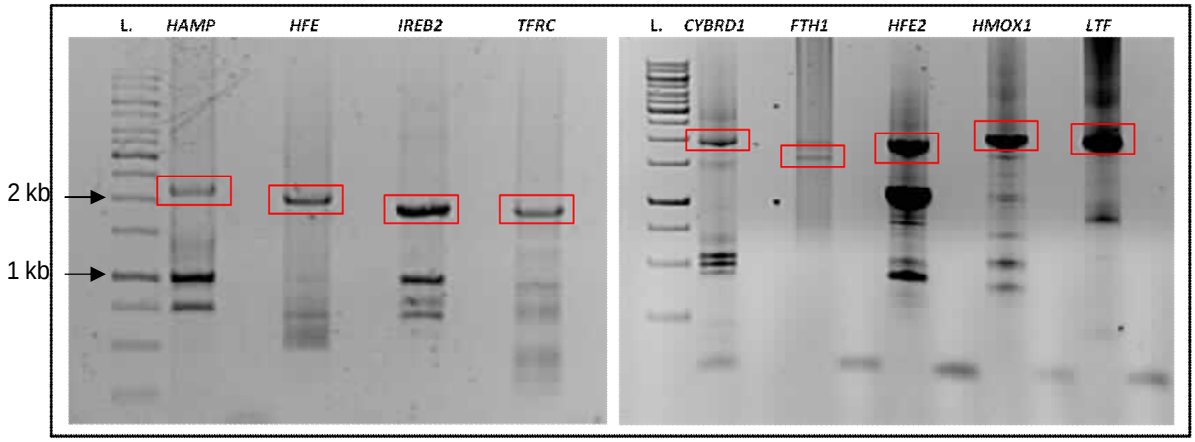


Figure 4.2 Agarose gel electrophoresis of PCR-driven overlap extension products (AD).

PCR products highlighted in red were excised from the agarose gel and used in a Wizard® SV Gel and PCR Clean-up.

Abbreviations: bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; kb, kilobase pair; L., Ladder; *LTF*, Lactotransferrin gene; PCR, polymerase chain reaction; *TFRC*, Transferrin receptor protein 1 gene.

4.4.2 Clone selection

Following preparation of the pGL4 constructs (as explained in Chapter 2, Section 2.2.6 to 2.2.9) and transformation of chemically competent cells (Chapter 2, Section 2.2.10), colonies were selected from the LB agar ampicillin plates and screened for positive transformants as described in Chapter 2, Section 2.2.11. The results of the plasmid minipreps performed for the 2 kb constructs and the 1.86 kb CR-removed promoter constructs, as well as the colony PCR results for the 140 bp CR elements are illustrated in Appendix 5: Figure S6, page 187. (bands of the correct insert size are surrounded by red boxes). Positive colonies were subjected to DNA sequence analysis to verify correct insert orientation and nucleotide sequence (Chapter 2, Section 2.2.12). Alignments of the sequencing results with the reference promoter sequence of the respective genes were performed with CLC Sequence Viewer 6.7.1. Graphical representations of the 1.86 kb CR-removed promoter sequence alignments are shown in Appendix 5: Figure S6, page 187.

4.4.3 Reporter Gene Assays

For each of the genes under investigation, normalized expression values as well as fold changes for each construct are indicated in Appendix 5 (Tables S1.1 to S9.4). These values were indicative of consistent and repeatable expression during all of the repetitive experiments performed. In order to calculate fold change values, the untreated 2 kb construct (with respect to the pGL4.10 constructs) and the untreated pGL4.23 vector (with respect to the 140 bp CR constructs) were used as reference (control constructs) and assigned a value of 1. The untreated 2 kb pGL4.10 construct was selected as a control due to the fact that the full, untreated 2 kb promoter is representative of the "wildtype" state when compared to the FAC-treated 2 kb construct and the 1.86 kb CR-removed plasmids. The empty, untreated pGL4.23[*luc2*/minP] was chosen as the control plasmid for comparison to the 140 bp CR constructs (untreated and FAC-treated) as a result of the inclusion of a minimal promoter in this vector. Basal transcription of the luciferase gene, driven by the minimal promoter, is therefore considered as "wildtype" and any differences in expression observed can be attributed to the cloned CR of interest.

Normalized values for the other constructs were expressed as a value relative to that of the untreated 2 kb construct or untreated pGL4.23 vector, respectively. The empty promoterless pGL4.10 vector was included as a negative control in those experiments in which the constructs were cloned into this plasmid. In all cases, the expression levels for this vector were negligible, as was expected, due to the lack of functional promoter elements with which to drive luciferase expression. In contrast, the empty pGL4.23 vectors demonstrated a basal level of transcription driven by the minimal promoter upstream of the luciferase reporter gene. The results obtained for the different constructs will be presented with respect to the individual genes investigated.

4.4.3.1 *Cytochrome b reductase 1* (*CYBRD1*)

Examination of the 140 bp CR *CYBRD1* constructs illustrated that no significant change in expression occurred (right panel, Figure 4.3). A marginal increase in expression was observed in the FAC-treated constructs in the HepG2 cells. However, the opposite was seen in the COS-1 cell line. With respect to the pGL4.10 *CYBRD1* constructs, comparison of the values obtained for both the HepG2 and COS-1 cell lines demonstrated the same trend in expression (left panel, Figure 4.3), although the fold increases were lower in the COS-1 cells. In the HepG2 cells, a statistically significant increase in expression ($p = 0.03$) was exhibited with the untreated 1.86 kb construct. This increase was 10-fold higher than that of the untreated 2 kb control. FAC treatment of the 1.86 kb CR-removed promoter construct resulted in a slight decrease in expression, although it was still significantly higher ($p = 0.02$) than that of the control. The values obtained for the untreated and FAC-treated 1.86 kb construct in the COS-1 cells were again elevated above the control ($p = 0.03$ and $p = 0.01$ respectively).

4.4.3.2 *Ferritin heavy polypeptide 1* (*FTH1*)

In both cell lines, the 140 bp CR *FTH1* constructs demonstrated a significant increase in expression when compared to the untreated pGL4.23 control (right panel, Figure 4.4). For the untreated 140 bp constructs statistically significant *p* values of 0.00 and 0.01 were observed in the HepG2 and COS-1 cells respectively. Expression was further increased upon addition of FAC [$p = 0.00$ (HepG2) and $p = 0.00$ (COS-1)].

Overall, addition of FAC resulted in higher expression values for the *FTH1* constructs. This increase was statistically significant in the HepG2 cells with the FAC-treated 2 kb promoter ($p = 0.00$) (left panel, Figure 4.4 A). With respect to the CR-removed *FTH1* constructs (left panel, Figure 4.4), removal of the CR resulted in a 66% decrease in expression of the untreated 1.86 kb construct when compared with the control [$p = 0.01$ (HepG2) and $p = 0.00$ (COS-1)]. Addition of FAC increased the levels slightly, although the values for this construct were still significantly decreased ($p = 0.00$) in the COS-1 cell line (left panel, Figure 4.4 B).

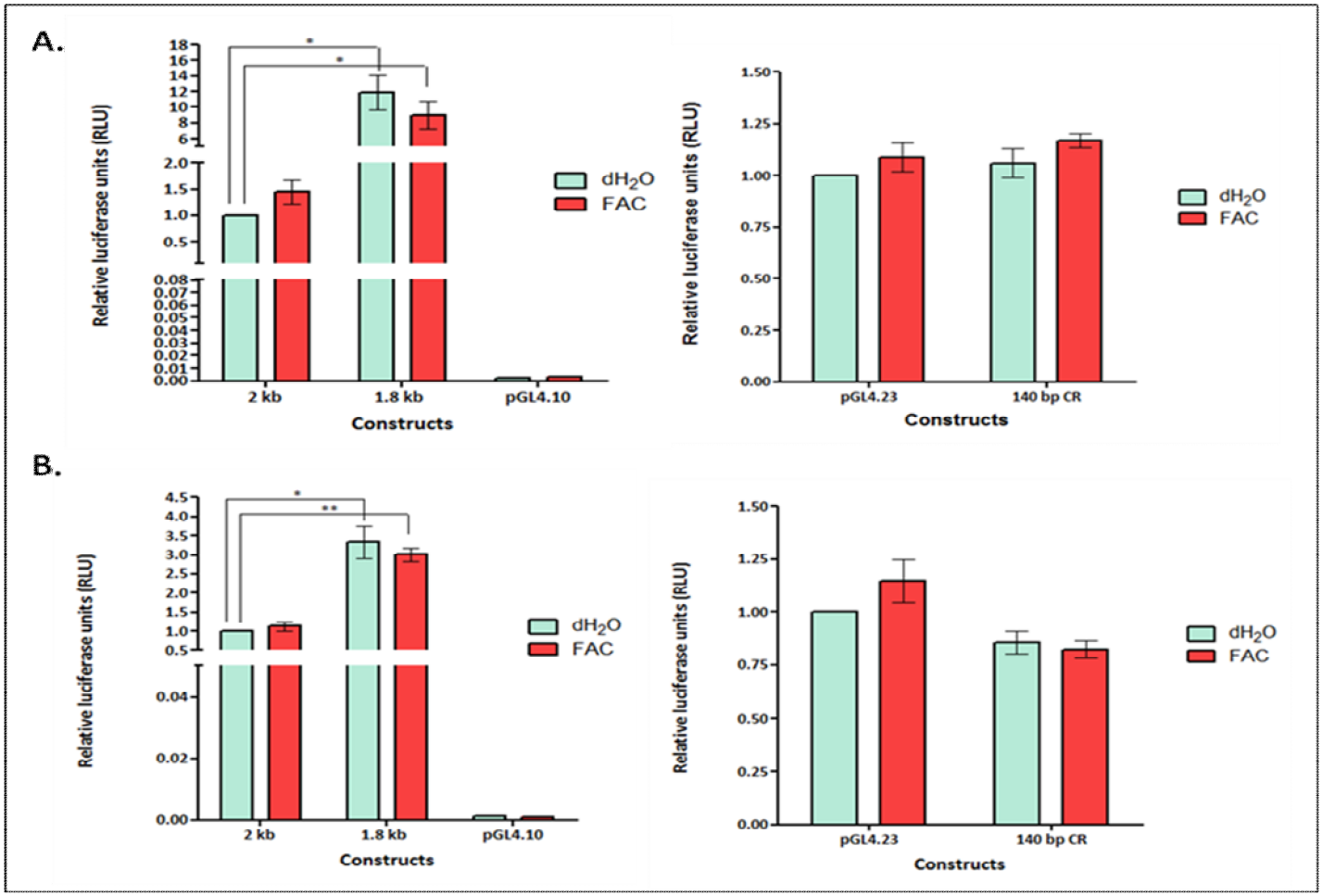


Figure 4.3 *Cytochrome b reductase 1* (*CYBRD1*) expression.

The cloned human *CYBRD1* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$.

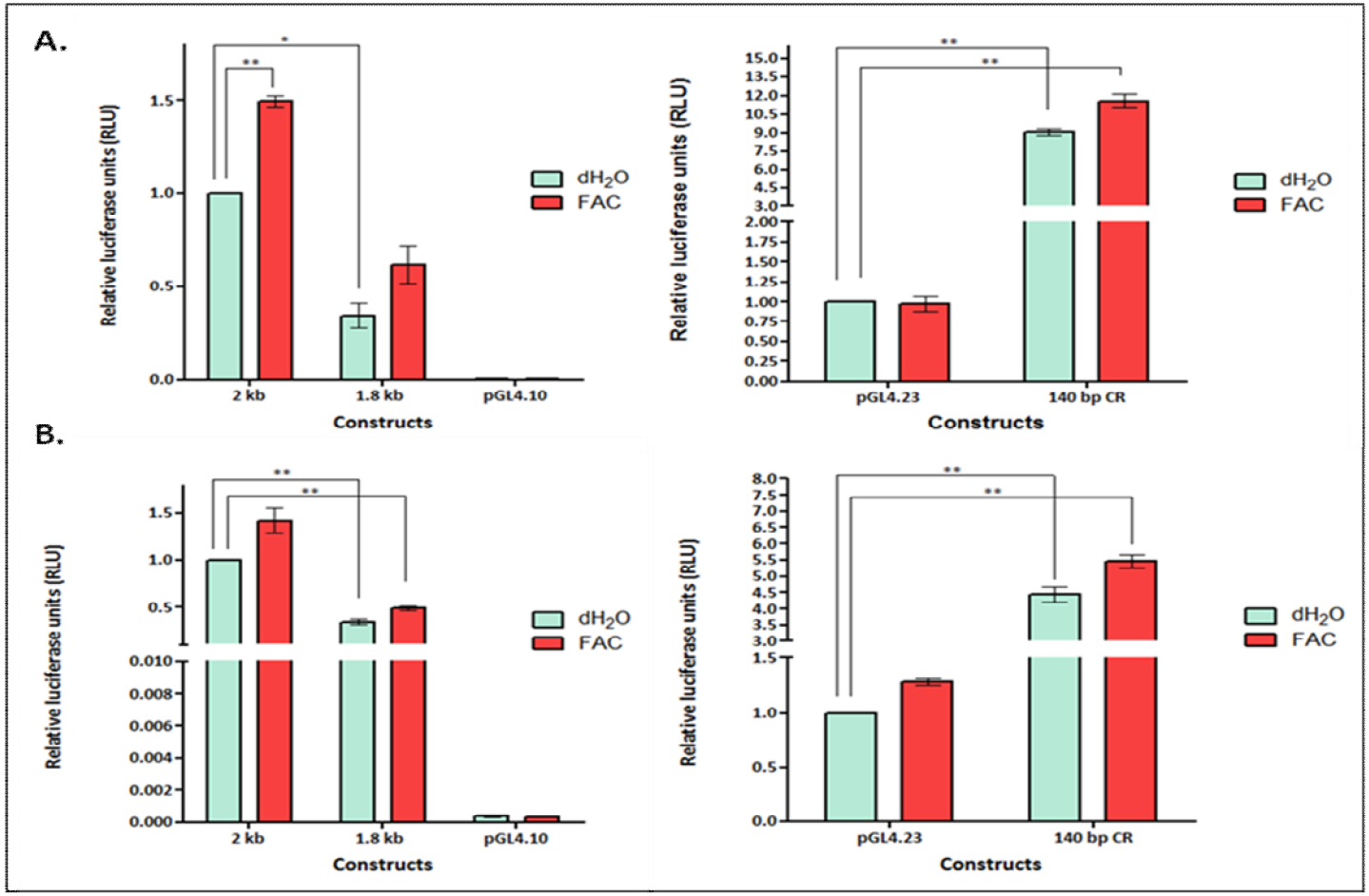


Figure 4.4 *Ferritin heavy polypeptide 1* (*FTH1*) expression.

The cloned human *FTH1* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$.

4.4.3.3. *Hepcidin* (*HAMP*)

Although not determined to be statistically significant, a modest increase in expression upon FAC treatment was observed with each of the *HAMP* constructs tested in both cell lines (Figure 4.5). Results indicated a significant decrease in expression, with a *p* value of 0.00, for the untreated 140 bp construct when compared to the pGL4.23 control in the HepG2 cells (right panel, Figure 4.5 A). A similar decrease of 20% was demonstrated in the COS-1 cell line however, this was not significant following statistical analysis. Addition of FAC increased expression levels slightly in both cell lines although the values were still below those of the untreated control and were only significant in the HepG2 cells ($p = 0.01$) (right panel, Figure 4.5 A).

Of interest was the observation that removal of the CR element from the *HAMP* promoter had little effect on expression levels in the HepG2 cells (right panel, Figure 4.5 A). In contrast, transcript levels in the COS-1 cell line for the 1.86 kb CR-removed promoter were significantly decreased in both untreated (41% decrease; $p = 0.01$) and FAC-treated (26% decrease; $p = 0.01$) constructs (left panel, Figure 4.5 B).

4.4.3.4 *Haemochromatosis* (*HFE*)

Expression levels of the *HFE* 140 bp CR constructs were dramatically decreased in both cell lines (right panel, Figure 4.6). In the HepG2 cells, the untreated CR construct showed significantly reduced luciferase levels ($p < 0.00$). FAC-treatment increased these levels slightly but did not abrogate the observed decrease in expression when compared to the control vector ($p < 0.00$). The same trend was observed in the COS-1 cells with *p* values for both the untreated and FAC-treated constructs below 0.00 (right panel, Figure 4.6 B).

FAC treatment of the 2 kb *HFE* constructs resulted in a 30% increase in expression in both cells lines relative to the untreated 2 kb control (left panel, Figure 4.6), although this was deemed statistically significant in only the COS-1 cells ($p = 0.05$). Although not significant, deletion of the CR from the *HFE* promoter led to an 84% expression increase in the HepG2 cells and a 14% increase in the COS-1 cells when compared to the control construct. FAC treatment of the 1.86 kb constructs had a negligible effect on these levels.

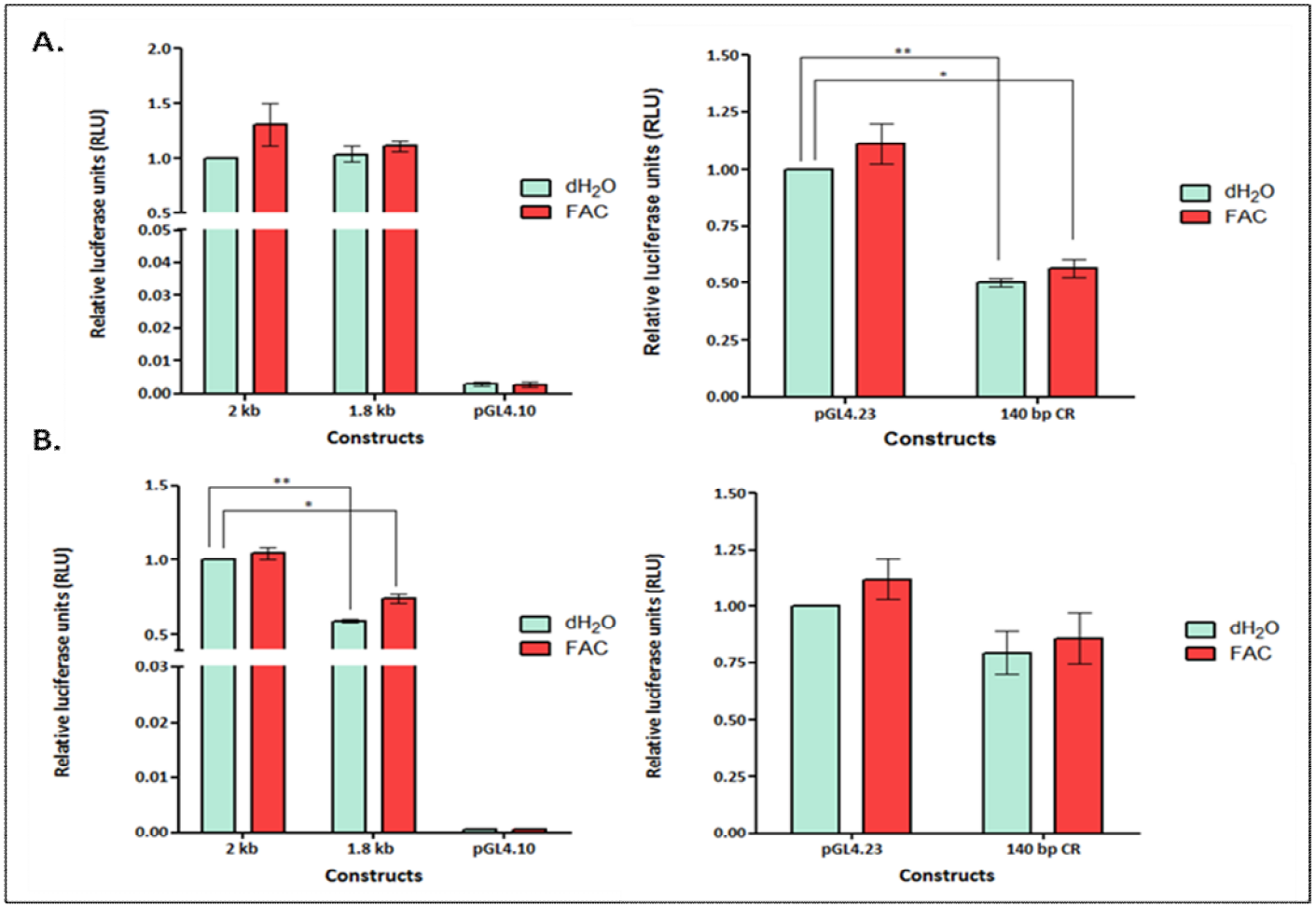


Figure 4.5 *Hepcidin* (*HAMP*) expression.

The cloned human *HAMP* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$.

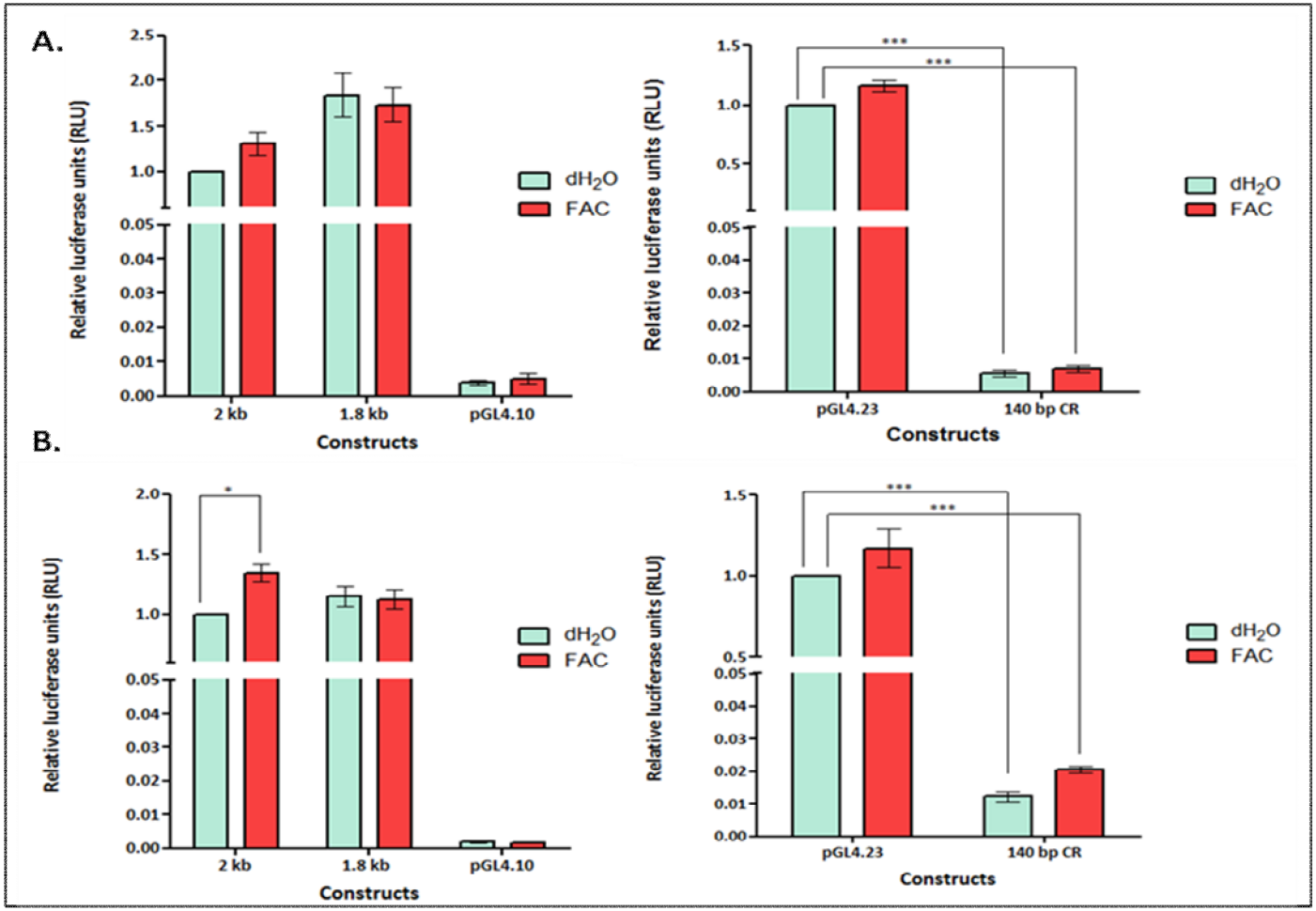


Figure 4.6 *Haemochromatosis (HFE)* expression.

The cloned human *HFE* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$.

4.4.3.5 *Hemojuvelin* (*HFE2*)

The untreated 140 bp CR *HFE2* construct resulted in a 329% increase ($p = 0.00$) in expression levels in the HepG2 cell line and a 146% increase ($p = 0.02$) in the COS-1 cells (right panel, Figure 4.7). Expression was further increased ($p = 0.02$) in the HepG2 cells upon FAC-treatment (right panel, Figure 4.7 A) however, a decrease in expression relative to the untreated 140 bp CR construct was seen in the COS-1 cells (right panel, Figure 4.7 B). The overall fold change in expression of the FAC-treated 140 bp CR construct in the COS-1 was still significantly higher than the untreated pGL4.23 control ($p = 0.04$) although the trend differed from observations in the HepG2 cells.

Simulation of iron overload with addition of FAC led to a modest increase in luciferase expression of between 12 and 34% in the 2 kb and 1.86 kb constructs in both cell lines. This increase was not determined to be statistically significant. In addition, the CR-removed promoter constructs demonstrated insignificant changes from the untreated 2 kb control plasmid (left panel, Figure 4.7).

4.4.3.6 *Haem oxygenase 1* (*HMOX1*)

With regard to the *HMOX1* 140 bp CR constructs, a significant decrease was observed in both the untreated (54% decrease, $p = 0.01$) and FAC-treated constructs (45% decrease, $p = 0.01$) in the HepG2 cells (right panel, Figure 4.8 A). The same trend could be seen in the COS-1 cells, however the differences in expression were marginal and not statistically significant (right panel, Figure 4.8 B).

Removal of the CR resulted in a 6-fold and 2-fold increase in expression of the untreated 1.86 kb constructs in the HepG2 ($p = 0.01$) and COS-1 ($p = 0.03$) cell lines respectively (left panel, Figure 4.8). FAC treatment further enhanced this effect with statistically significant increases [$p = 0.01$ (HepG2) and $p = 0.01$ (COS-1)] when compared with the untreated 2 kb control.

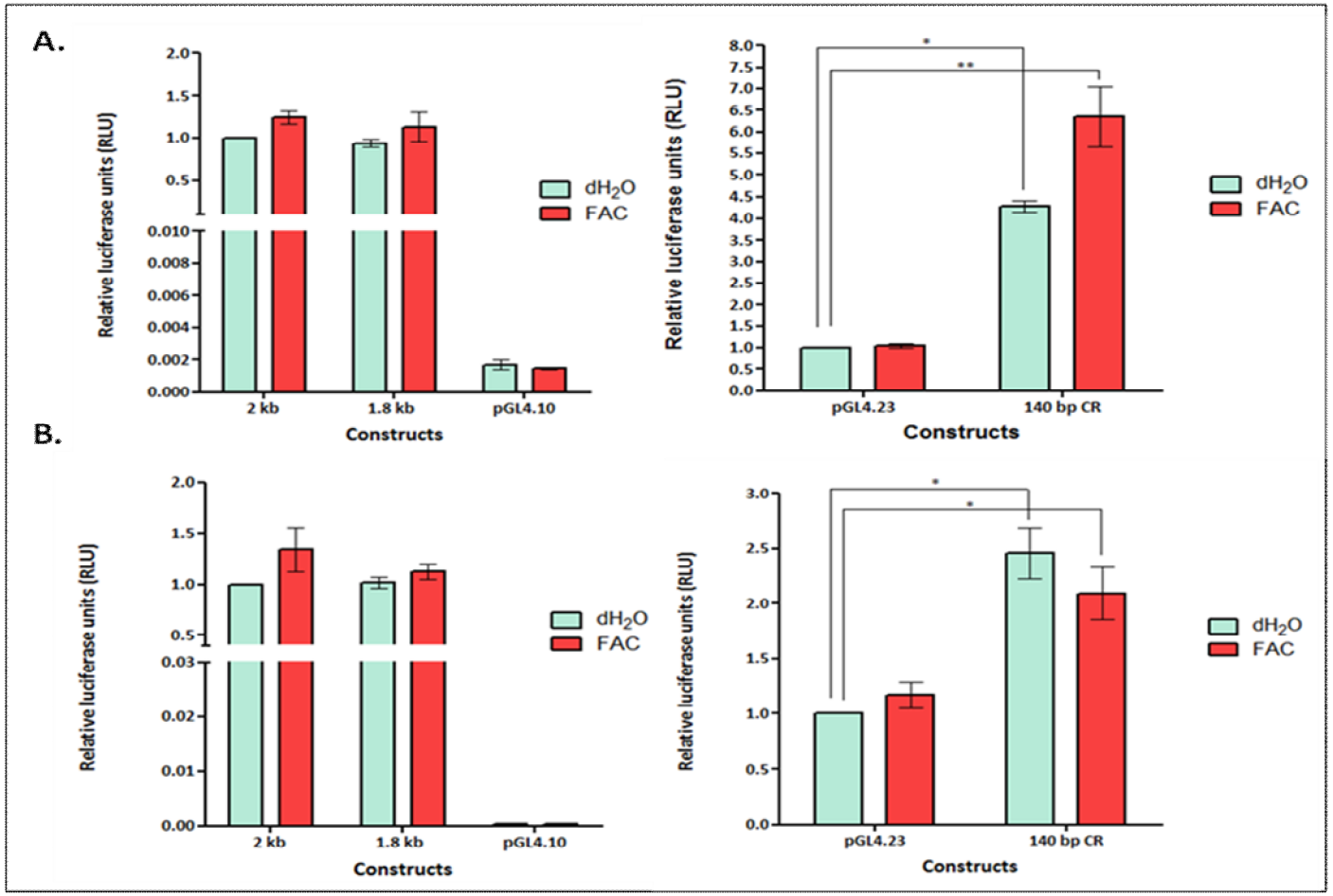


Figure 4.7 *Hemojuvelin (HFE2)* expression.

The cloned human *HFE2* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$.

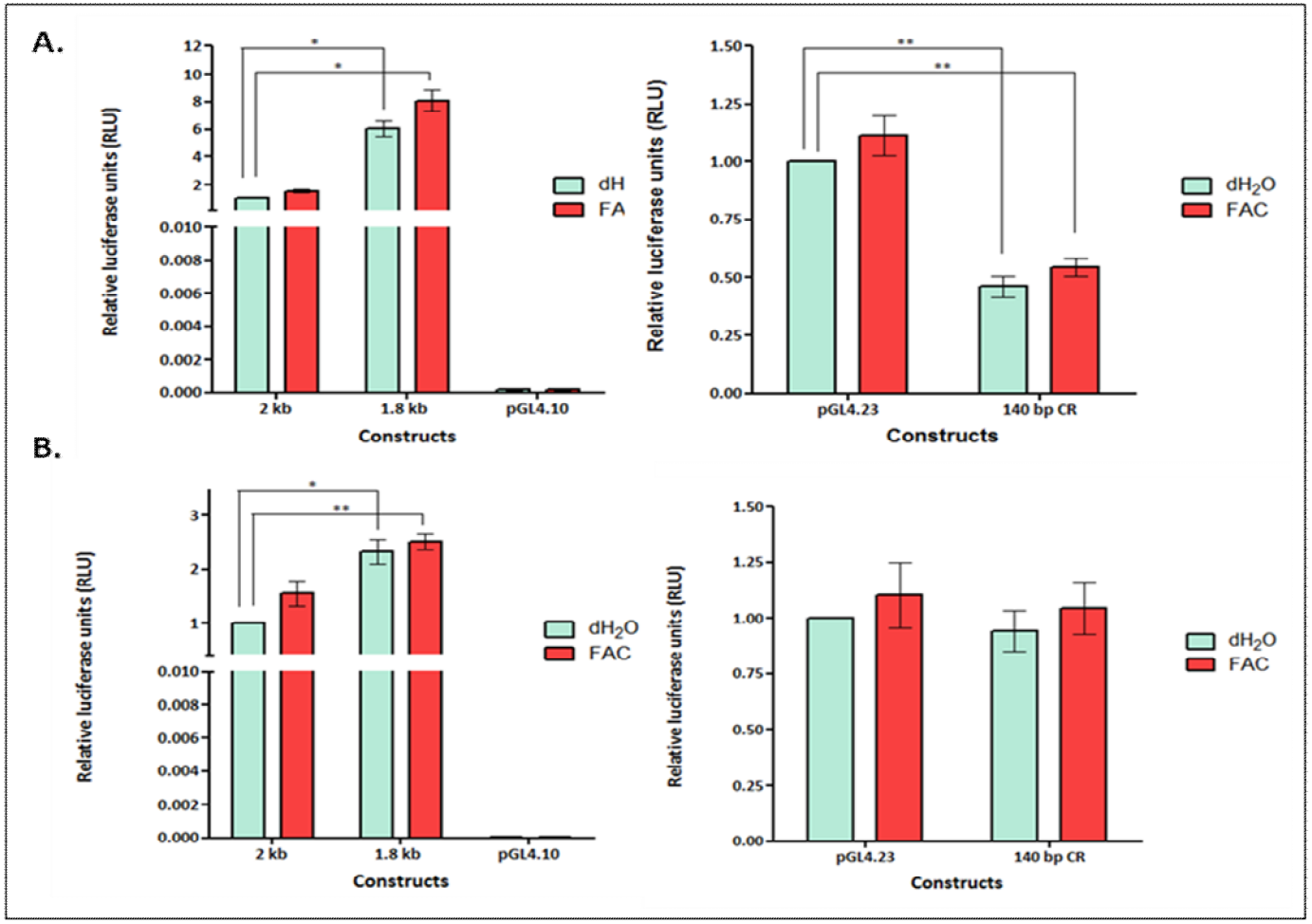


Figure 4.8 *Haem oxygenase 1 (HMOX1)* expression.

The cloned human *HMOX1* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$.

4.4.3.7 *Iron regulatory binding protein 2* (*IREB2*)

Results for the untreated 140 bp CR *IREB2* constructs indicated a decrease in luciferase expression of 67% in the HepG2 cells and 41% in the COS-1 cells (p = 0.00 and p = 0.00 respectively) (right panel, Figure 4.9). Expression was modestly increased following addition of FAC, although the overall decrease was still significant with respect to the untreated pGL4.23 control [p = 0.00 (HepG2) and p = 0.02 (COS-1)] (right panel, Figure 4.9).

The 1.86 kb CR-removed promoter constructs in the HepG2 cells did not demonstrate any significant deviations from the control vector, although a small increase in expression was demonstrated upon FAC treatment (left panel, Figure 4.9 A). In contrast, the untreated 1.86 kb construct showed a significant decrease (p = 0.00) relative to the control in the COS-1 cell line. FAC treatment of this construct led to a further decrease in luciferase expression (p = 0.00) relative to the untreated control (left panel, Figure 4.9 B).

4.4.3.8 *Lactotransferrin* (*LTF*)

With the exception of the 2 kb construct transfected into the HepG2 cells, an overall increase in expression was observed upon FAC treatment with all of the *LTF* constructs (Figure 4.10). The FAC-treated 2 kb construct resulted in decreased levels of expression (p = 0.01) in the HepG2 cells when compared to the untreated control (left panel, Figure 4.10 A). A 72% decrease in expression of the untreated 140 bp CR construct (p = 0.00) and a 65% decrease with the FAC-treated CR construct (p = 0.00) were observed in the HepG2 cells (right panel, Figure 4.10 A). The same trend was observed in the COS-1 cells although the expression levels were not as low as those observed in the HepG2 cells [p = 0.01 (untreated) and p = 0.02 (FAC-treated)] (right panel, Figure 4.10 B).

In both cell lines, targeted deletion of the CR resulted in decreased expression levels relative to the untreated 2 kb control (left panel, Figure 4.10). Statistical analyses revealed that these observations were only significant in the HepG2 cells; p = 0.04 for the untreated 1.86 kb construct and p = 0.01 for the FAC-treated equivalent.

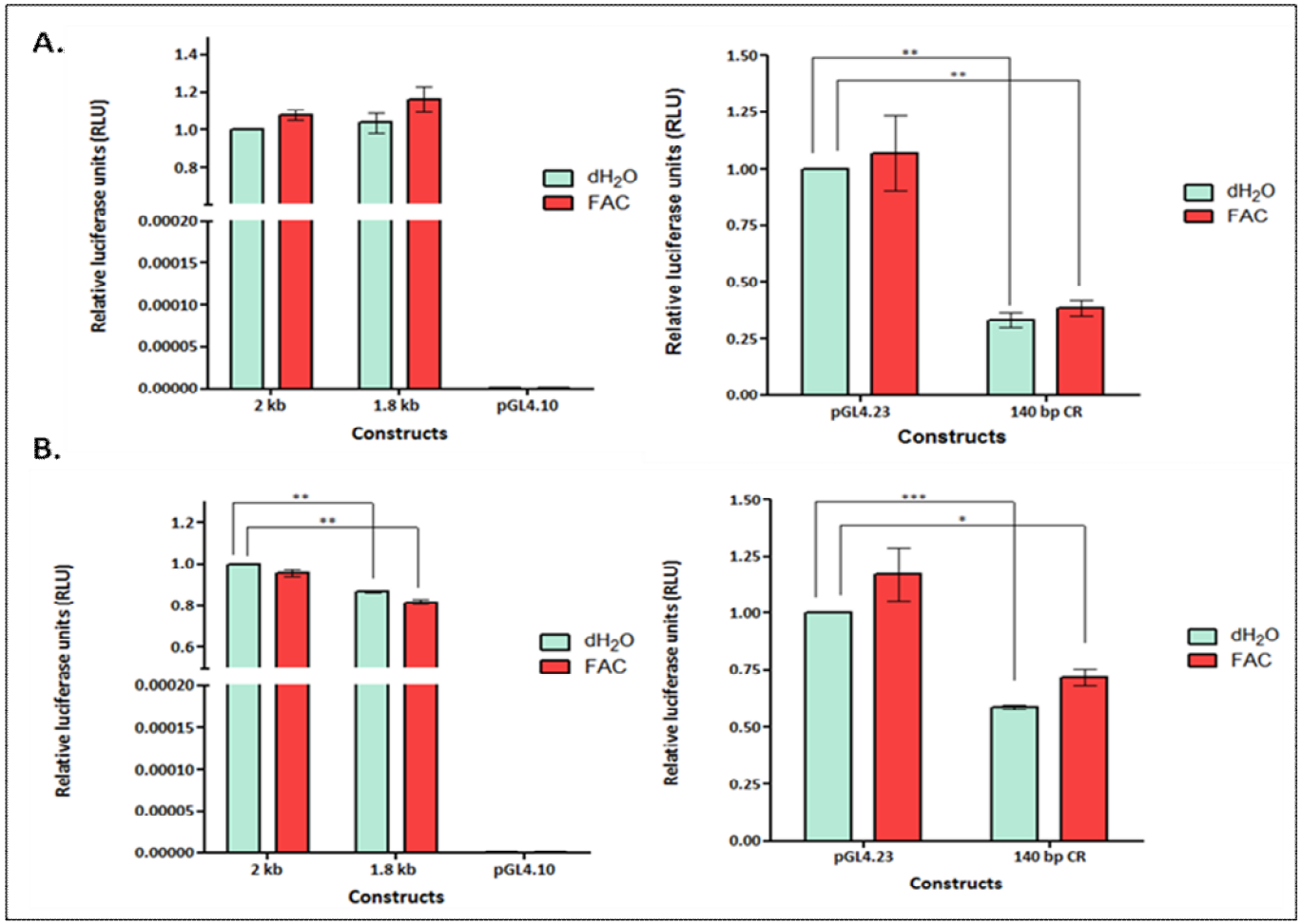


Figure 4.9 *Iron regulatory binding protein 2 (IREB2)* expression.

The cloned human *IREB2* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$.

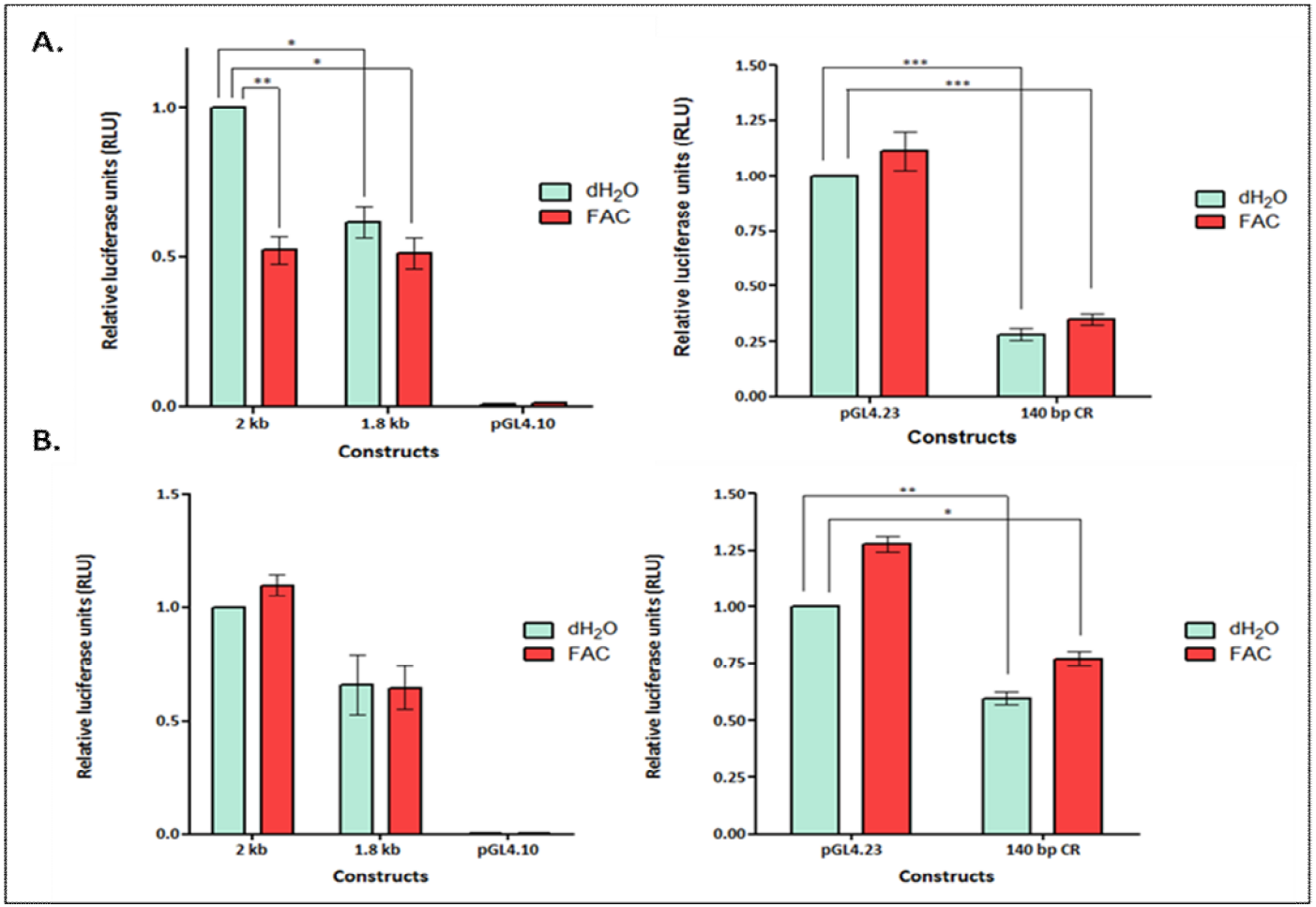


Figure 4.10 *Lactotransferrin (LTF)* expression.

The cloned human *LTF* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$.

4.4.3.9 *Transferrin receptor 1* (*TFRC*)

Overall, FAC treatment of the *TFRC* constructs in both the HepG2 and COS-1 cells resulted in modest expression increases (Figure 4.11). In the HepG2 cells, a statistically significant decrease (p = 0.00) of 56% with the untreated 140 bp CR construct and a 41% decrease (p = 0.02) with the FAC-treated CR construct was observed relative to the control pGL4.23 vector (right panel, Figure 4.11 A). In the COS-1 cell line, an insignificant decrease in expression resulted from transfection of the untreated 140 bp CR construct. Following FAC treatment, expression increased to levels above those of the untreated control (p = 0.05), an elevation of 17% (right panel, Figure 4.11 B).

In both cell lines, the untreated CR-removed promoter constructs demonstrated discernable expression increases relative to the untreated 2 kb control although these were not determined to be statistically significant (left panel, Figure 4.11). The addition of FAC to these constructs resulted in further small expression increases (between 49% and 52%) which were not statistically significant when compared to the control.

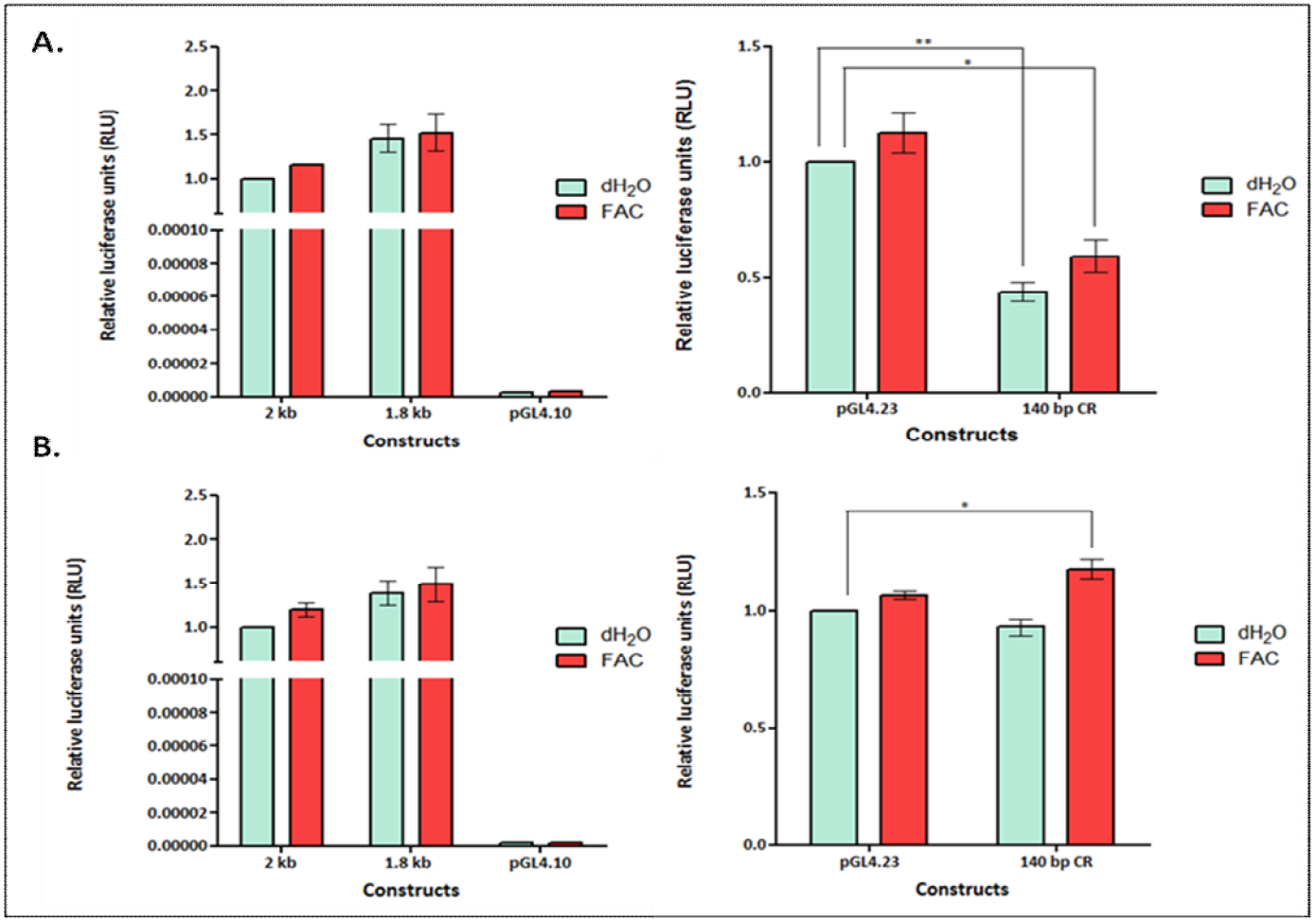


Figure 4.11 *Transferrin receptor 1 (TFRC)* expression.

The cloned human *TFRC* promoter constructs were transfected into **A.**) HepG2 cells or **B.**) COS-1 cells. Transfected cells were subsequently treated with ferric ammonium citrate (FAC). Fold activation was calculated with respect to the activity of the untreated 2 kb construct (left panels) and the untreated pGL4.23 vector (right panels). All transfections were performed in triplicate and included pGL4.73[*hRluc*/SV40] as internal control; luciferase levels were normalized to *Renilla* activity. Data are representative of three independent experiments and are presented as means ± SEM. * $p < 0.05$, ** $p < 0.01$.

4.5 DISCUSSION

PCR amplification of the different promoter elements of interest in this study was performed successfully. The PCR-driven overlap extension technique that was employed for the generation of the 1.86 kb CR-removed promoters proved effective (Figure 4.1 and 4.2). Standard mutagenesis-based techniques to introduce restriction sites in order to remove sequences by enzymatic digestion often introduce erroneous base pairs into the sequence. These sites have the potential to act as binding sites for new Tfs, thus inadvertently influencing transcriptional regulation of the genes of interest. In this study, the CR was deleted from the whole promoter amplicons of the respective genes in a manner that did not affect the native sequence, as evidenced by DNA sequence analysis (Appendix 5: Figure S6, page 187). Subsequent to PCR amplification and molecular cloning, positive colonies for each of the different promoter constructs for all nine genes were successfully identified (Appendix 5: Figure S5, page 186). Due to difficulties experienced during the optimization of colony PCR conditions for the larger constructs (2 kb and 1.86 kb), double digestion reactions of the extracted plasmids were performed. This departure from customary colony PCR methodology proved useful in this study for the identification of positive transformants (Appendix 5: Figure S5 A and B, page 186).

In general, *in vitro* functional data for the upstream coding regions of human genes implicated in iron metabolism is scarce. Exceptions include the *HAMP* promoter region, which has been extensively characterized (Falzacappa *et al.*, 2008; Parajes *et al.*, 2010), and selected cases of individual disease-associated mutations which have been assessed in some of the genes studied here (Constantine *et al.*, 2009; Gao *et al.*, 2009). The data presented here therefore represent an overview of the activity of the promoters of the genes of interest under standard culture conditions, and in response to simulated conditions of iron overload, in two different mammalian cell lines. In this study, HepG2 cells were chosen for the *in vitro* assays due to the fact that the majority of iron sequestration and metabolism takes place in the hepatocyte cells of the liver. The genes under investigation here, with the possible exception of *LTF* (Levay and Viljoen, 1995), are all ubiquitously expressed in hepatocytes. In contrast, COS-1 cells, which are derived from the kidney of the African green monkey, represent an alternate *in vitro* environment for the determination of differential

gene expression due to the tissue-specific nature of particular Tfs which may interact with the gene promoters. The HNF family are one group of Tfs that demonstrate this tissue-specificity and are known to regulate transcription of genes expressed in the liver (Costa *et al*., 2003). Consequently, differences observed in the expression values for the different constructs, between the two cell lines, could potentially be attributed to the absence/presence of tissue-specific Tfs.

With respect to FAC treatment of the promoter constructs, a summary of the graphical results previously presented is illustrated in Table 4.1. The results of three independent transfection experiments are shown as a percentage change in luciferase expression for each of the constructs, for all nine genes, and in both cell lines. Statistically significant differences are shown in bold text and are indicated by the relevant asterisks. Constructs that demonstrated over expression (↑) or under expression (↓) relative to the respective constructs utilized as a control are also indicated in Table 4.1.

With particular reference to transfections of the full 2 kb promoter constructs for each gene, the majority demonstrated moderate fold change increases upon FAC treatment when compared to the untreated 2 kb construct (Table 4.1). The observations for the *CYBRD1*, *HMOX1* and *TFRC* constructs were unexpected in a biological context where expression of these genes would be likely to decrease in response to high iron conditions (Hentze and Kühn, 1996; Gunshin *et al.*, 2005). However, none of the observations were statistically significant and it is possible that a longer exposure time to FAC is required for these effects to become noticeable. In addition, previously performed functional studies, in which these genes were shown to be upregulated in response to iron deprivation, were conducted on mice (Gunshin *et al.*, 2005) and rat intestinal cells (Collins, 2006) and slightly different expression profiles may be revealed *in vivo* in other mammalian species. With regard to *TFRC*, five tandem IREs have been identified in the 3'-UTR of the gene which are responsible for stabilizing the mRNA under low iron conditions resulting in increased expression (Hentze and Kühn, 1996). The focus of this study was on gene promoters and thus only the 5'-UTR of *TFRC* was utilized for construct preparation. This region may not be sufficient for appropriate *TFRC* response to iron as would be expected in an *in vivo* context.

Table 4.1 Percentage change in luciferase activity for the FAC-treated constructs.

	Treated (FAC)											
	HepG2						COS-1					
Gene	**140 bp CR pGL4.23**		**1.8 kb pGL4.10**		**2 kb pGL4.10**		**140 bp CR pGL4.23**		**1.8 kb pGL4.10**		**2 kb pGL4.10**	
CYBRD1	↑	17	**↑**	***798**	↑	45	↓	18	**↑**	****200**	↑	13
FTH1	**↑**	****1056**	↓	38	**↑**	****49**	**↑**	****445**	**↓**	****51**	↑	42
HAMP	**↓**	***43**	↑	11	↑	31	↓	14	**↓**	***26**	↑	5
HFE	**↓**	*****99**	↑	73	↑	30	**↓**	*****98**	↑	12	**↑**	***34**
HFE2	**↑**	****534**	↑	13	↑	25	**↑**	***109**	↑	12	↑	34
HMOX1	**↓**	****45**	**↑**	***708**	**↑**	53	↑	4	**↑**	****152**	↑	55
IREB2	**↓**	****62**	↑	16	↑	8	**↓**	***28**	**↓**	****18**	↓	5
LTF	**↓**	*****65**	**↓**	***49**	**↓**	****48**	**↓**	***23**	↓	36	↑	10
TFRC	**↓**	***41**	↑	52	↑	15	↑	17	↑	49	↑	20

Abbreviations: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.0001$; ↑, over expression; ↓, under expression; bp, base pair; COS-1, African green monkey kidney cells; CR, conserved region; *CYBRD1*, Cytochrome b reductase 1 gene; FAC, ferric ammonium citrate; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; HepG2, human hepatocarcinoma liver cells; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; kb, kilobase pair; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

Statistically significant increases in expression were seen in the FAC-treated 2 kb constructs for *FTH1* in HepG2 cells (p = 0.00) and *HFE* in COS-1 cells (p = 0.05) (Table 4.1). The raised values obtained for the FAC-treated *FTH1* promoter construct were expected in this study because of the inclusion of the *FTH1* IRE in the 2 kb construct. *FTH1* expression is known to be elevated in iron sufficient circumstances as a consequence of the inability of iron-bound IRP1 or IRP2 to interact with the 5'-UTR IRE, therefore permitting translation of the *FTH1* mRNA to proceed unhindered (Thomson *et al.*, 1999). When cells are iron replete, synthesis of both *FTH1* and *FTL* is enhanced to allow for sequestration of the excess iron within the ferritin molecule, thereby reducing the bioavailability of potentially toxic free iron (Torti and Torti, 2002). Although not determined to be statistically significant, a 42% increase in expression of the FAC-treated *FTH1* 2 kb construct was also observed in the COS-1 cells. In the case of *HFE*, increased expression levels in response to iron overload can be explained by examining the interaction between the HFE protein and TFR1 *in vivo*. Under iron overload conditions, competitive binding of HFE to TFR1 leads to a decrease in the affinity of TFR1 for diferric-TF and a subsequent decline in cellular iron uptake, minimizing intracellular iron concentrations (Feder *et al.*, 1998). It is therefore feasible that FAC-treatment would result in heightened transcription of the luciferase reporter driven by the 2 kb *HFE* promoter.

Deviations from the increased expression trend were observed with the FAC-treated 2 kb constructs for *LTF* and *IREB2* (Table 4.1). In HepG2 cells, *LTF* showed a statistically significant decrease in expression of 48% (p = 0.01) and in COS-1 cells, *IREB2* transfection resulted in a 5% decrease. The decline observed in *LTF*-driven expression was unexpected as, like *TF*, *LTF* levels are normally elevated in response to high iron conditions (Kawakami and Lönnerdal, 1991). It is possible that due to the defined role of LTF in the immune system, a longer period of FAC treatment may be required to be able to visualize the appropriate increase in expression. In addition, hepatocytes are not the primary tissue type for *LTF* expression and may therefore not be the ideal *in vitro* environment in which to assess transcriptional regulation of this gene. Corroboration of this theory is evidenced by the 10% increase in expression of the FAC-treated *LTF* 2 kb construct in the COS-1 cell line (Table 4.1). The decrease in expression observed for the FAC-treated *IREB2* construct in COS-1 cells was expected. *IREB2* mRNA has previously been demonstrated to undergo proteosomal degradation when exposed to high cellular iron levels (Wang *et al.*, 2008). However, in the

HepG2 cells a marginal increase in *IREB2*-driven expression of 7% relative to the untreated control was observed. This finding was not statistically significant and may have been as a consequence of a single experimental repeat with fold change values slightly above those of the other experiments.

For the 1.86 kb CR-removed promoter constructs, supplementation with FAC resulted in luciferase activity differences that were consistent with what was observed in the 2 kb untreated *versus* FAC-treated constructs (Table 4.1). Any FAC-related expression deviations in the 2 kb constructs were maintained in almost every instance in the 1.86 kb CR-removed constructs, even in those cases where the effect of targeted deletion of the CR was dramatic (*FTH1*, *HMOX1*, *LTF*). This finding is in accordance with observations from the FAC-treated 140 bp CR constructs and it is therefore unlikely that the CR element alone exerts a transcriptional effect in an iron-dependent manner. This assumption is not unexpected, as the previously described IRE/IRP system clearly fulfils the iron-sensory role necessary for transcriptional regulation of a subset of the genes of interest. It is probable that the CR element functions more as a general transcriptional regulator, which from the results of this study, appears to be unrelated to conditions of iron overload in an *in vitro* setting.

Although of interest and scientific value, the functional activity of the gene promoters in response to iron was of secondary importance in this study. The primary goal of the luciferase reporter assays performed was to assess any potential effect that may be mediated by the CR that was identified in the *in silico* portion of this study. A summary of the graphical results previously presented for the untreated 140 bp CR constructs and the 1.86 kb CR-removed constructs, for each gene of interest and in both cell lines, is illustrated in Table 4.2. The results of three independent transfection experiments are shown as a percentage change in luciferase expression for each of the constructs. Statistically significant differences are shown in bold text and are indicated by the relevant asterisks. Those constructs that demonstrated over expression (↑) or under expression (↓) relative to the respective constructs utilized as a control are also indicated in Table 4.2.

Table 4.2 Percentage change in luciferase activity for the untreated constructs.

	Untreated (dH_2O)							
	HepG2				COS-1			
Gene	**140 bp CR pGL4.23**		**1.8 kb pGL4.10**		**140 bp CR pGL4.23**		**1.8 kb pGL4.10**	
CYBRD1	↑	6	↑	***1090**	↓	15	↑	***233**
FTH1	↑	****800**	↓	***66**	↑	****344**	↓	****66**
HAMP	↓	****50**	↑	4	↓	20	↓	***41**
HFE	↓	*****99**	↑	84	↓	*****99**	↑	14
HFE2	↑	***329**	↓	7	↑	146	↑	2
HMOX1	↓	**** 54**	↑	***503**	↓	6	↑	***133**
IREB2	↓	****67**	↑	4	↓	*****41**	↓	****13**
LTF	↓	*****72**	↓	***38**	↓	****40**	↓	34
TFRC	↓	****56**	↑	45	↓	7	↑	39

Abbreviations: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.0001$; ↑, over expression; ↓, under expression; bp, base pair; COS-1, African green monkey kidney cells; CR, conserved region; *CYBRD1*, Cytochrome b reductase 1 gene; dH_2O, distilled water; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; HepG2, human hepatocarcinoma liver cells; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; kb, kilobase pair; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

The CR-removed promoter constructs that displayed statistically significant increases in expression in both cell lines when compared to the untreated 2 kb control were *CYBRD1* (Figure 4.3) and *HMOX1* (Figure 4.8). Both *CYBRD1* and *HMOX1* demonstrated an increase in expression with *p*-values, for the untreated 1.86 kb construct, of 0.0305 and 0.0112 respectively in the HepG2 cells. This significant increase was maintained in the COS-1 cell line (Table 4.2). In addition, expression of the 1.86 kb untreated construct for *HFE* was also increased above the 2 kb control, although this was not determined to be statistically significant (Figure 4.6). An increase of 84% was observed in the HepG2 cells and a 14% expression increase in the COS-1 cells (Table 4.2). For *CYBRD1* and *HFE*, FAC-treatment of the 1.86 kb CR-removed promoter construct resulted in an almost negligible decrease in expression relative to treated construct (Table 4.1). For *HMOX1*, a marginal increase was observed following FAC treatment of the 1.86 kb promoter (Table 4.1). However, for each of these genes, the changes were insignificant and consistent with what was observed between the untreated and FAC-treated 2 kb constructs and are therefore unlikely to be related to removal of the CR from the respective gene promoters.

The results obtained from the luciferase assays of the 140 bp CR cloned into the pGL4.23 vector indicate that this promoter element may have an inhibitory effect on transcription of *CYBRD1*, *HFE* and *HMOX1*. Luciferase expression values for all three of these

constructs were reduced in one or both of the cell lines tested, and the opposite effect was observed upon removal of the CR with the 1.86 kb promoters (Table 4.2). The untreated *HFE* CR appeared to result in the most dramatic decrease in expression with *p*-values of <0.0001 for both the HepG2 and COS-1 cells (right panel, Figure 4.6). The untreated *HMOX1* 140 bp CR construct also exhibited a significant decrease in luciferase expression in the HepG2 cells ($p = 0.01$) and a slight decrease in the COS-1 cells (right panel, Figure 4.8). As with the larger CR-removed promoter constructs, FAC treatment did not result in any significant differences in expression, over and above what was observed between the empty pGL4.23 vectors themselves (Table 4.1).

With respect to the 1.86 kb CR-removed promoter constructs for *HAMP*, *HFE2*, *IREB2* and *TFRC*, no significant change in expression was observed relative to the wildtype 2 kb construct in the HepG2 cells (Table 4.2). For *HAMP* (Figure 4.5) and *IREB2* (Figure 4.9) however, a significant decrease in expression was exhibited in the COS-1 cells ($p = 0.01$ and $p = 0.00$, respectively) with the untreated constructs (Table 4.2). It is interesting to note that each of these genes was found to contain an identical binding site (MRE) for the MTF-1 Tf originally identified in the *HAMP* CR (Chapter 3, Table 3.3). Balesaria and colleagues (2010) proved that induction of *HAMP* expression by MTF-1 binding in an *in vitro* environment is dependent on the metal zinc. Other metals such as copper and cadmium were also able to positively influence luciferase activity in a dose-dependent manner. It is postulated that the CR identified in these four genes may require the presence of another metal to manipulate transcriptional activity *via* MTF-1 binding in hepatocyte cells. In this study, cells were only treated with FAC to simulate iron overload conditions which may explain the lack of significant change in luciferase activity observed with the CR-removed promoter constructs in the HepG2 cells. Further experimentation would be required to prove this hypothesis.

Significant decreases in luciferase activity resulted from transfection of both cell lines with the 140 bp CR for *HAMP*, *IREB2* and *TFRC* (Figures 4.5, 4.9 and 4.11; Table 4.2). It therefore appears that this promoter element elicits an inhibitory effect on gene expression in these genes *in vitro*. Despite the significant change in activity exhibited, it should be noted that due to the MRE discussed above, the CR element may have a more positive effect *in vivo*, and in the presence of zinc, than was demonstrated by the reporter gene assays. Transfection of the *HFE2* 140 bp CR however, resulted in a significant increase in

luciferase activity [p = 0.00 (HepG2) and p = 0.02 (COS-1)] (Table 4.2). The reason for this deviation from the trend seen in the other genes is unclear. However, results of the *in silico* analysis of the *HFE2* CR indicated that Motif 3 (Chapter 3, Figure 3.5, red block) was missing from this gene when compared to the others. TfBSs located in the CRM represented by Motif 3 may be responsible for the repression of transcriptional activity. The lack thereof could be responsible for the observed increase in luciferase expression with the *HFE2* CR.

Transfection of *FTH1* (Figure 4.4) and *LTF* (Figure 4.10) CR-removed promoter constructs resulted in luciferase activity changes that followed a different overall trend in comparison to the genes discussed above (Table 4.2). *FTH1* was found to contain the MRE in the CR element, however the 1.86 kb CR-removed promoter appeared to decrease luciferase activity in both cell lines tested [p = 0.01 (HepG2) and p = 0.00 (COS-1)]. It is plausible that the removal of the MRE from the *FTH1* promoter resulted in the decrease in expression. However, the reason for the much larger reduction in luciferase activity relative to the other MRE-containing genes, also in the absence of zinc, cannot be satisfactorily explained. A similar decrease was observed with the *LTF* 1.86 kb construct in both cell lines, although this was only significant in the HepG2 cells (p = 0.04) (Table 4.2). As in the case of *HFE2* discussed above, Motif 3 was predicted to be absent from the *LTF* CR following bioinformatic investigation. One of the Tfs identified in Motif 3 was C/EBPα. This particular Tf has been shown to act a positive regulator of transcription in conjunction with other molecules that form part of the pre-initiation complex required for basal transcriptional activation (Nerlov, 2007). With regard to the role of C/EBPα in iron metabolism, studies on the *HAMP* promoter have confirmed that C/EBPα binding is capable of activating *HAMP* expression in a hepatocyte-specific manner (Courselaud *et al.*, 2002). Absence of the TfBS for C/EBPα could explain the difference in luciferase activity that was observed for *LTF* in comparison to the other genes that were investigated.

4.6 CONCLUSION

The functional influence of the CR identified in this study warrants further investigation. Overall, it seems as if the CR may be mediating an inhibitory response (with the exception of *FTH1* and *HFE2*), decreasing the tempo of transcription, *via* an iron-independent mechanism. It is evident that deletion of the CR did not abrogate or alter the response of the respective promoter regions to FAC treatment. This region may therefore have a role in the regulation of transcription under physiologically normal conditions by effecting the co-ordinated regulation of this gene network. In the majority of the genes of interest (*CYBRD1*, *HFE*, *HMOX1*, *IREB2*, *TFRC*), targeted deletion of the CR led to higher expression values when compared to the control. Examination of the findings of the *in vitro* cell culture assays demonstrate that this promoter element results in significant differences in promoter activity, especially when examined individually in the 140 bp CR constructs. Although the effect of the CR appeared less statistically significant when removed from the full promoter, as in the case of the 1.86 kb CR-removed constructs, discernible variations in luciferase activity were still visible. Expression variations were stable and repeatable over the triplicate experiments performed, indicating the validity of the functional importance of this region. These results are crucial indicators that the methodology utilized in this study has merit for future investigations of gene regulatory features. The bioinformatic predictions were validated by the reporter assays conducted thus establishing a platform for future work in this critical pathway. Further studies should focus on examining the individual motifs predicted within the CR, possibly utilizing a deletion construct-type approach. In addition, it will be of paramount importance to elucidate the authentic functional interactions that take place between the CR (or individual motifs) and Tfs of interest. Techniques such as EMSA or ChIP would be of benefit in this regard (Euskirchen *et al.*, 2007; Gurevich *et al.*, 2010; Hu *et al.*, 2010).

CHAPTER FIVE

Conclusions and Future Directions

In the era following the publication of the human genome, genetic research of eukaryotic genomes is experiencing a departure from traditional methods. The technologies available to study DNA sequences, with particular reference to next-generation sequencing, have improved rapidly over the years and are widely available and utilized. Computational tools required for the analysis of the vast quantities of data produced by these techniques have been developed in parallel and are numerous in their number and potential applications. A large proportion of these software tools are available as freeware, the use of which has great potential to assist the bench biologist with limited bioinformatic expertise. It would be negligent to ignore this new avenue of research, particularly with regard to the study of functional regulatory elements in the genomes of complex organisms. Evidence of the emerging study of regulatory features of the human genome is illustrated by consortiums such as the ENCODE project. Recently, the researchers involved simultaneously published a series of 30 papers in different scientific journals containing the results of their study (see http://www.encodeproject.org/ENCODE/pubs.html for publication collection). The main findings effectively dismiss the theory that the human genome consists primarily of "junk DNA". The ENCODE results have shown that more than 80% of non-coding DNA sequence takes part in biochemical- or chromatin-associated events and most lie close to areas of protein-DNA interaction (The ENCODE Project Consortium, 2012). Results of this type reiterate the importance of research into identifying promoter regulatory elements that may help to explain the complex process of transcription.

The overall aim of the current proof-of-concept study was to investigate the upstream untranslated region of specific genes utilizing a combinatorial computational and functional genetic approach to identify potential transcriptional regulatory targets. With regard to the bioinformatic portion of the strategy, the first objective involved the selection of candidate genes. The iron metabolism pathway was chosen as a model system due to the inherent level of physiological control required to maintain homeostatic iron levels within the body. Despite intensive investigation, the transcriptional mechanisms responsible for regulation of gene expression in this pathway are only beginning to be unravelled. In addition, disorders of iron metabolism are among the most common diseases affecting individuals worldwide and warrant further research. In order to satisfy the first objective of this study, eighteen genes encoding the most central role players of the iron metabolism pathway were chosen.

The decision was made not to select those genes known to overlap extensively with alternate biochemical pathways, such as the haem biosynthesis pathway, and mitochondrial genes. Examples of the protein products of these genes include: ABCB7, a membrane transporter required for [Fe-S] cluster transport in the mitochondria (Dunn *et al.*, 2006); frataxin (FXN), which aids in [Fe-S] cluster formation (Schmucker *et al.*, 2011) and δ-aminolevulinic acid synthase (ALAS), the first rate-limiting enzyme in the synthesis of haem (Nilsson *et al.*, 2009). Genes encoding proteins with multiple functions in different pathways are likely to be regulated *via* diverse mechanisms, thus complicating the process of motif discovery unnecessarily at this stage. This represents a potential limitation of this study which would be remedied in future investigations by applying the strategy designed in this study to these genes. Expanding analyses of this type to these genes, as well as to others that have yet to be discovered, would supplement the current understanding of this critical pathway.

The second and third objectives for the bioinformatic section of this study included the *in silico* analyses of the non-coding sequences of the respective genes with the aim to potentially identify novel regulatory targets (discussed in Chapter 3). This was achieved through the use of a variety of software tools that were selected based on their relative popularity and ease of use. The main findings included a previously uncharacterized region of unexpectedly high sequence conservation in nine of the genes. This region is proposed as a novel *Alu-J* element based on sequence alignments. These mobile elements have been associated with TfBSs in numerous cases thus substantiating the idea that this region may contribute to regulation of the genes of this pathway (Polak and Domany, 2006). *In silico* prediction of TfBSs resulted in the identification of numerous putative Tfs that will serve as the basis for future work based on the outcomes of this study. The results of the *in silico* analyses of the promoter region of the genes of interest represent a novel contribution to the field of iron research.

Functional promoter assays were designed in an attempt to validate the *in silico* predictions and are discussed in Chapter 4. The objectives for this subset of the study were effectively achieved in each instance. Three different promoter constructs were prepared for each of the genes found to contain the region of sequence conservation. These were successfully

transfected into mammalian cell lines in an *in vitro* culture system and promoter activity was assessed with a dual-luciferase reporter assay system. For each of the genes, statistically significant variations in luciferase expression were observed with the constructs representing the CR element. Although the trends in expression were not always consistent with what was expected for each gene, it has been shown in previous studies that interactions with specific Tfs can result in markedly different effects between different genes. Wu and colleagues (1999) demonstrated that the interaction of the Tf, c-myc, with *FTH1* resulted in repression of expression. The same interaction with c-myc was responsible for the stimulation of *IREB2* expression in an *in vitro* environment (Wu *et al.*, 1999). Deletion of the CR may therefore have unexpected consequences with regard to transcriptional regulation of gene expression.

Functional validation confirmed the bioinformatic predictions with regard to the postulated effect of the CR on transcriptional regulation. However, relative to the constructs designated as controls, these variations were not notably different upon treatment of the cells with FAC indicating that the CR functions in an iron-independent manner under the conditions tested. As previously eluded to, it is more likely that the CR is exerting functional effects in a manner that is related transcriptional regulation under physiologically normal conditions. Although not a new concept, the hypothesis of co-ordinated regulation of a network of genes is supported by the findings of this study. Based on the bioinformatic investigation conducted, the possibility that the CR element may actually be an *Alu* element substantiates this idea. From an evolutionary perspective, transposable elements are more likely to have been conserved in the non-coding regions of genes if they confer some type of advantage (Häsler and Strub, 2006).

Despite the vital data that resulted from the functional assays, certain limitations exist which could be addressed in further studies. The cell lines utilized for transfection of the constructs should be expanded to include different cell types, particularly an intestinal cell line such as Caco-2 (Alvarez-Hernandez *et al.*, 1991; Han *et al.*, 1995; Tallkvist *et al.*, 2000; Tandy *et al.*, 2000), which was unfortunately unavailable for the current study. Although the majority of the genes investigated have been shown to demonstrate expression in the cell types employed, it is possible that divergent results with regard to the CR may be observed

in alternate tissue types. It should also be emphasized that the suitability of the particular cell type to be utilized should be evaluated extensively. Recent work has shown that HepG2 cells cultured with relatively high glucose concentrations (as is standard for this cell type) may have significantly lower levels of endogenous HNF4-α (Selva *et al.*, 2007; Simó *et al.*, 2012). The down regulation of this important liver-specific Tf could have severe consequences with respect to the evaluation of expression levels in a HepG2 *in vitro* experiment. The use of exogenous stimuli, other than FAC, is recommended for further work of this nature. It will be important is assess the potential effects of zinc on the CR and CR-removed promoters to confirm if the MTF-1 TfBS identified in five of the genes is relevant in the context of transcriptional regulation as suggested by the results presented here. Exogenous stimuli such as bacterial lipopolysaccharide (LPS) and interferon gamma (INF-γ) should also be tested to determine if the CR responds to these immunomodulators. LPS and INF-γ are typically added to culture medium to induce an immune response in the growing cells (Lewis *et al.*, 1990). Iron metabolism genes, and *HAMP* in particular, are known to be regulated by conditions of inflammation (Ganz and Nemeth, 2005).

Although not within the scope of this study, the most crucial aspect of future work should involve the determination of the physical interactions between the CR DNA sequence and Tfs selected on the basis of the bioinformatic results. Antibody-based techniques such as Western blots (Brunet *et al.*, 2004), EMSA (Gurevich *et al.*, 2010) and ChIP (Hu *et al.*, 2010) would be ideal to carry out this task. It would also be imperative to confirm that the cell lines selected for this type of work express the Tfs of interest. Determination of Tf mRNA concentration in the relevant cells could be easily performed by means of qRT-PCR prior to the instigation of experimental protocols (Czechowski *et al.*, 2004). Co-transfection of cell lines with vectors expressing the Tfs of interest together with the promoter constructs prepared in this study would also allow for determination of the effect of the potential interaction(s) on gene expression (Trujillo *et al.*, 1991). Deletion constructs of the 140 bp CR should also be prepared to identify if the individual motifs within this region are contributing significantly to the changes observed in luciferase activity. It may be a simpler undertaking to pursue candidate Tf-interaction investigations on smaller genomic regions (represented by the conserved motifs) than the whole span of 140 bp. Identification of a core subset of Tfs that are responsible for the induction or repression of the respective

genes, by binding to the CR, would provide tantalizing possibilities for the development of Tf-specific agonists or antagonists. Drugs of this nature have proved effective for the control of diabetes mellitus and other conditions of insulin resistance. In these examples, the Tf, PPARγ, is targeted and activated by the drug class thiazolidinedione. Activated PPARγ is able to bind to DNA, when complexed to RXR, and modulate the expression of a variety of lipid metabolism genes (Krentz and Friedmann, 2006)

Standard molecular genetic screening approaches for the identification of DNA sequence variants that could confer risk for disease development would benefit from the type of study performed here. Putative regulatory regions identified by computational means would represent an ideal starting position within which to search for mutations in modifier genes for HH. One of the significant findings from the ENCODE publications was related to the re-analysis of previous GWAS conducted on various human disorders (Butter *et al.*, 2012). It was demonstrated than an abundance of the non-coding SNPs, which were the focus of the GWAS were associated with ENCODE-annotated functional regions (Schaub *et al.*, 2012). SNPs located within TfBSs have the ability to influence cognate Tf-binding and can potentially result in aberrant gene expression, particularly if the Tf in question is responsible for recruiting transcriptional co-activators and repressors (Kasowski *et al.*, 2010). Numerous human diseases result from mutations in TfBSs (reviewed in Maston *et al.*, 2006) illustrating the importance of regulatory SNPs. Variations identified within functionally important regions, during mutational screens of non-coding DNA sequences, may therefore have more significant clinical consequences than has generally been acknowledged until recently.

To our knowledge, this is the first study to employ a strategy of this nature in an investigation of the iron metabolic pathway. Results indicate that the design of experiments can be greatly facilitated by prior computational examination of genomic sequences, and that these software programs have the capability to predict functionally relevant DNA regions. A novel regulatory element was identified in this study that has potential implications for the co-ordinated regulation of genes of this pathway. The further studies described above will be required to fully elucidate the physiological mechanisms underlying the observations from this work. It is hoped that studies, such as the one presented in this dissertation, will provide invaluable information that will help to unravel the complexities

associated with transcriptional regulation. Although not an overnight process, the development of pharmaceutical interventions for the diseases associated with aberrant iron metabolism will be extremely beneficial to affected individuals, and will be facilitated by research of this nature.

CHAPTER SIX

References

6.1 GENERAL REFERENCES

Abboud S and Haile DJ (2000) A Novel Mammalian Iron-regulated Protein Involved in Intracellular Iron Metabolism. *J Biol Chem* **275**: 19906-19912.

Adams B, Dorfler P, Aguzzi A, Kozmik Z, Urbanek P, Maurer-Fogy I, Busslinger M (1992) Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev* **6**: 1589-607.

Aguila A, Herrera AG, Morrison D, Cosgrove B, Perojo A, Montesinos I, Pérez J, Sierra G, Gemmell CG, Brock JH (2001) Bacteriostatic activity of human lactoferrin against *Staphylococcus aureus* is a function of its iron-binding properties and is not influenced by antibiotic resistance. *FEMS Immunol Med Microbiol* **31**: 145-152.

Aisen P, Enns C, Wessling-Resnick M (2001) Chemistry and biology of eukaryotic iron metabolism. *Int J Biochem Cell Biol* **33**: 940-959.

Alam J and Cook JL (1990) Reporter genes: application to the study of mammalian gene transcription. *Anal Biochem* **188**: 245-254.

Alderton G, Ward WH, Ferold HL (1946) Identification of the bacteria-inhibiting iron-binding protein of egg white as conalbumin. *Arch Biochem Biophys* **11**: 9-12.

Alvarez-Hernandez X, Nichols GM, Glass J (1991) Caco-2 cell line: a system for studying intestinal iron transport across epithelial cell monolayers. *BBA Biomembranes* **1070**: 205-208.

Anderson BF, Baker HM, Norris GE *et al.* (1990) Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins. *Nature* **344**: 784-787.

Anderson GJ, Darshan D, Wilkins SJ, Frazer DM (2007) Regulation of systemic iron homeostasis: how the body responds to changes in iron demand. *Biometals* **20**: 665-674.

Anderson GJ, Frazer DM, McKie AT, Vulpe CD, Smith A (2005) Mechanisms of haem and non-haem iron absorption: lessons from inherited disorders of iron metabolism. *Biometals* **18**: 339-348.

Andrews GK, Lee DK, Ravindra R, Lichtlen P, Sirito M, Sawadogo M, *et al.* (2001) The transcription factors MTF-1 and USF1 cooperate to regulate mouse metallothionein-I expression in response to the essential metal zinc in visceral endoderm cells during early development. *EMBO J* **20**: 1114-1122.

Andrews NC (1999) Disorders of iron metabolism. *N Engl J Med* **341**: 1986-1995.

Andrews NC (2000) Iron homeostasis: insights from genetics and animal models. *Nature Rev Genet* **1**: 208-217.

Andrews NC (2008) Forging a field: the golden age of iron biology. *Blood* **112**: 219-230.

Andrews NC Fleming MD, Levy JE (1999) Molecular insights into mechanisms of iron transport. *Curr Opin Hematol* **6**: 61-64.

Babitt JL, Huang FW, Wrighting DM, Xia Y, Sidis Y, Samad TA, Campagna JA, Chung RT, Schneyer AL, Woolf CJ, Andrews NC, Lin HY (2006) Bone morphogenetic protein signaling by hemojuvelin regulates hepcidin expression. *Nat Genet* **38**: 531-539.

Babitt J, Huang FW, Xia Y, Sidis Y, Andrews NC, Lin H (2007) Modulation of bone morphogenetic protein signalling *in vivo* regulates systemic iron balance. *J Clin Invest* **117**: 1933-1939.

Bailey S, Evans RW, Garratt RC, *et al.* (1988) Molecular structure of serum transferrin at 3.3 Å resolution. *Biochem* **27**: 5804-5812.

Bailey TL, Williams N, Misleh C and Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucl Acids Res* **34**: W369–W373.
available at: http://meme.sdsc.edu/meme/intro.html

Bali PK and Aisen P (1991) Receptor-modulated iron release from transferrin: differential effects on the N- and C-terminal sites. *Biochem* **30**: 9947-9952.

Banerjee S, Farhana A, Ehtesham NZ, Hasnain SE (2011) Iron acquisition, assimilation and regulation in mycobacteria. *Infect Genet Evol* **11**: 825-838.

Batzer MA and Deininger PL (2002) ALU repeats and human genomic diversity. *Nat Rev Genet* **3**: 370-380.

Beard JL, Dawson H, Piñero DJ (1996) Iron Metabolism: A Comprehensive Review. *Nutr Rev* **54**: 295-317.

Beaumont C (2010) Multiple regulatory mechanisms act in concert to control ferroportin expression and heme iron recycling by macrophages. *Haematol* **95**: 1233-1236.

Bishop JKB, Wood TJ, Davis RE, Sherman JT (2004) Robotic observations of enhanced carbon biomass and export at 55 degrees during SOFeX. *Science* **304**: 417-420.

Bittel D, Dalton T, Samson SL, Gedamu L, Andrews GK (1998) The DNA binding activity of metal response element-binding transcription factor-1 is activated in vivo and in vitro by zinc, but not by other transition metals. *J Biol Chem* **273**: 7127-7133.

Blanchette M, Bataille AR, Chen X, *et al*. (2006) Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression *Genome Res* **16**: 656-668.

Bolouri H and Davidson EH (2003) Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics. *Proc Natl Acad Sci USA* **100**: 9371-9376.

Bou-Abdallah F (2012) Does Iron Release from Transferrin Involve a Reductive Process? *Bioenergetics* **1**: 1000e111.

Boussingault JB (1872) Du fer contenu dans le sang et dans les ailments. *CR Acad Sci Paris* **74**: 1353-1359.

Boyd PW, Jickells T, Law CS, Blain S, Boyle EA, Buesseler KO, Coale KH, Cullen JJ, de Baar HJW Follows M, *et al.* (2007) Mesoscale Iron Enrichment Experiments 1993-2005: Synthesis and Future Directions. *Science* **315**: 612-617.

Boyd PW, Watson AJ, Law CS, Abraham ER, Trull T, Murdoch R, Bakker DCE, Bowie AR, Buessler KO, Chang H, *et al.* (2000) A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**: 695-702.

Bray N, Dubchak I, Pachter L (2003) AVID: A Global Alignment Program. *Genome Res* **13(1)**: 97-102.

Brenowitz M, Senear DF, Shea MA, Ackers GK (1986) Quantitative DNase footprint titration: a method for studying protein-DNA interactions. *Method Enzymol* **130**: 132-181.

Bridle KR, Frazer DM, Wilkins SJ, *et al.* (2003) Disrupted hepcidin regulation in HFE-associated haemochromatosis and the liver as a regulator of body iron homoeostasis. *Lancet.* **361**: 669-673.

Brissot P, Bardou-Jacquet E, Jouanolle A-M, Loréal O (2011) Iron disorders of genetic origin: a changing world. *Trends Mol Med* **17**: 707-713.

Brosius J (1999) RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238**: 115-134.

Brun T, Franklin I, St-Onge L, *et al.* (2005) The diabetes-linked transcription factor PAX4 promotes β-cell proliferation and survival in rat and human islets. *J Cell Biol* **167**: 1123-35.

Brunet A, Sweeney LB, Sturgill JF, Chua KF, Greer PL, *et al.* (2004) Stress-Dependent Regulation of FOXO Transcription Factors by the SIRT1 Deacetylase. *Science* **303**: 2011-2015.

Buckingham M and Relaix F (2007) The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu Rev Cell Dev Biol* **23**: 645-673.

Bulyk ML (2003) Computational prediction of transcription-factor binding site locations. *Genome Biol* **5**: 201.

Burrows A (2000) Supernova explosions in the Universe. *Nature* **403**: 727-733.

Butler JEF and Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* **16**: 2583-2592.

Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, Mann M (2012) Proteome-Wide Analysis of Disease-Associated SNPs That Show Allele-Specific Transcription Factor Binding. *PLOS Genet* **8**: e1002982-e1002982.

Cairo G and Pietrangelo A (2000) Iron regulatory proteins in pathobiology. *Biochem J* **352**: 241-250.

Calame KL, Lin KI, Tunyaplin C (2003) Regulatory mechanisms that determine the development and function of plasma cells. *Annu Rev Immunol* **21**: 205-30.

Camaschella C, Roetto A, Cali A, *et al.* (2000) The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22. *Nat Genet* **24**: 14-15.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science* **309**: 1559-63.

Carter SL, Brechbühler CM, Griffin M, Bond AT (2004) Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20**: 2242-2250.

Casanovas G, Mleczko-Sanecka K, Altamura S, Hentza MW, Muckenthaler MU (2009) Bone morphogenetic protein (BMP)-responsive elements located in the proximal and distal hepcidin promoter are critical for its response to HJV/BMP/SMAD. *J Mol Med* **87**: 471-480.

Cassar N, Bender ML, Barnett BA, Fan S, Moxim WJ, Levy H, Tilbrook B (2007) The Southern Ocean Biological Response to Aeolian Iron Deposition. *Science* **317**: 1067-1070.

Cavill I (2002) Erythropoiesis and iron. *Best Pract Res Cl Ha* **15**: 399–409.

Cereghini S (1996) Liver-enriched transcription factors and hepatocyte differentiation. *FASEB J* **10**: 267-82.

Chaudhary J, Cupp AS, Skinner MK (1997) Role of basic-helix-loop-helix transcription factors in Sertoli cell differentiation: Identification of an E-box response element in the transferrin promoter. *Endocrinology* **138**: 667-675.

Chen W, Huang FW, Barientos de Renshaw T, Andrews NC (2011) Skeletal muscle hemojuvelin is dispensable for systemic iron homeostasis. *Blood* **112**: 6319-6325.

Chen WJ, Chang SH, Hudson ME, Kwan W-K, Li J, Estes B, Knoll D, Shi L, Zhu T (2005) Contribution of transcriptional regulation to natural variations in Arabidopsis. *Genome Biol* **6**: R32.

Cheng Y, Zak O, Aisen P, Harrison SC, Walz T (2004) Structure of the human transferrin receptor-transferrin complex. *Cell* **116**: 565-576.

Coale KH, Johnson KS, Chavez FP, *et al.* (2004) Southern Ocean iron enrichment experiment: carbon cycling in high- and low-Si waters. *Science* **304**: 408-414.

Coffer PJ and Burgering BMT (2004) Forkhead box transcription factors and their role in the immune system. *Nature Rev Immunol* **4**: 889-899.

Collas P (2010) The Current State of Chromatin Immunoprecipitation. *Mol Biotechnol* **45**: 87-100.

Collins JF (2006) Gene chip analyses reveal differential genetic responses to iron deficiency in rat duodenum and jejunum. *Biol Res* **39**: 25-37.

Collins JF and Hu Z (2007) Promoter analysis of intestinal genes induced during iron-deprivation reveals enrichment of conserved SP1-like binding sites. *BMC Genomics* **8**: 1-13.

Conrad ME, Umbreit JN, Moore EG (1999) Iron absorption and transport. *Am J Med Sci* **318**: 213-229.

Constantine CC, Anderson GJ, Vulpe CD, McLaren CE, Bahlo M, *et al.* (2009) A novel association between a SNP in *CYBRD1* and serum ferritin levels in a cohort study of *HFE* hereditary haemochromatosis. *Br J Haematol* **147**: 140-149.

Cooper GM and Sidow A (2003) Genomic regulatory regions: insights from comparative sequence analysis. *Curr Opin Genet Dev* **13:** 1-7.

Costa RH, Kalinichenko VV, Holterman AX, Wang X (2003) Transcription factors in liver development, differentiation, and regeneration. *Hepatology* **38**: 1331-1347.

Courselaud B, Pigeon C, Inoue Y, Inoue J, Gonzalez FJ, *et al.* (2002) C/EBPα regulates hepatic transcription of hepcidin, an antimicrobial peptide and regulator of iron metabolism. *J Biol Chem* **277**: 41163-41170.

Crawford GE, Holt IE, Mullikin JC, Tai D, Blakesley R, *et al.* (2004) Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc Natl Acad Sci USA* **101**:992-97.

Crichton R (2009) *Iron Metabolism: From Molecular Mechanisms to Clinical Consequences*. 3rd edn, United Kingdom: John Wiley & Sons Ltd.

Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK (2004) Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J* 38: 366-379.

Daimon M, Yamatani K, Igarashi M, Fukase N, Kawanami T, Kato T, Tominaga M, Sasaki H (1995) Fine Structure of the Human Ceruloplasmin Gene. *Biochem Bioph Res Co* **208**: 1028–1035.

Dalton TP, Bittel D, Andrews GK (1997) Reversible activation of mouse metal response element-binding transcription factor 1 DNA binding involves zinc interaction with the zinc finger domain. *Mol Cell Biol* **17**: 2781-2789.

Das MK and Dai H-K (2007) A survey of DNA motif finding algorithms. *BMC Bioinformatics* **8**: S21.

Datta J, Majumder S, Kutay H, Motiwala T, Frankel W, Costa R, *et al.* (2007) Metallothionein expression is suppressed in primary human hepatocellular carcinomas and is mediated through inactivation of CCAAT/enhancer binding protein alpha by phosphatidylinositol 3-kinase signaling cascade. *Cancer Res* **67**: 2736-2746.

Dautry-Varsat A, Ciechanover A, Lodish HF (1983) pH and the recycling of transferrin during receptor-mediated endocytosis. *Proc Natl Acad Sci USA* **80**: 2258-2262.

Deaglio S, Capobianco A, Calì A, Bellora F, Alberti F, *et al.* (2002) Structural, functional, and tissue distribution analysis of human transferrin receptor-2 by murine monoclonal antibodies and a polyclonal antiserum. *Blood* **100**: 3782-3789.

De Domenico I, McVey Ward D, Langelier C, Vaughn MB, Nemeth E, Sundquist WI, Ganz T, Musci G, Kaplan J (2007) The molecular mechanism of hepcidin-mediated ferroportin down-regulation. *Mol Biol Cell* **18**: 2569-2578.

De Domenico I, McVey Ward D, Kaplan J (2008) Regulation of iron acquisition and storage: consequences for iron-linked disorders. *Nature Rev* **9**: 72-81.

Deiss A (1983) Iron metabolism in reticuloendothelial cells. *Semin Hematol* **20**: 81-90.

Deugnier Y, Brissot P, Loréal O (2008) Iron and the liver. Update 2008. *J Hepatol* **48**: S113-123.

Donovan A, Brownlie A, Zhou Y, Shepard J, Pratt SJ, Moynihan J, Paw BH, Drejer A, Barut B, Zapata Z, Law TC, Brugnara C, Lux SE, Pinkus GS, Pinkus JL, Kingsley PD, Palis J, Fleming MD, Andrews MC, Zon LI (2000) Positional cloning of zebrafish ferroportin 1 identifies a conserved vertebrate iron exporter. *Nature* **403**: 776-781.

Down TA and Hubbard TJ (2002) Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Res* **12**: 458-461.

Dröge P and Müller-Hill B (2001) High local protein concentrations at promoters: strategies in prokaryotic and eukaryotic cells. *Bioessays* **23(2):** 179-83.

Dunn LL, Rahmanto YS, Richardson DR (2006) Iron uptake and metabolism in the new millennium. *Trends Cell Biol* **17**: 93-100.

Eckenroth BE, Steere AN, Chasteen ND, Everse SJ (2011) How the binding of human transferrin primes the transferrin receptor potentiating iron release at endosomal pH. *Proc Natl Acad Sci USA* **108**: 13089-13094.

Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucl Acids Res* **30**: 207-210.

Elgar G, Vavouri T (2008) Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet* **24**: 344-52.

Emerit J, Beaumont C, Trivin F (2001) Iron metabolism, free radicals and oxidative injury. *Biomed Pharmacother* **55**: 333-339.

Euskirchen GM, Rozowsky JS, Wei C-L, Lee WH, Zhang ZD, *et al.* (2007) Mapping of transcription factor binding regions in mammalian cells by ChIP: Comparison of array- and sequencing-based technologies. *Genome Res* **17**: 898-909.

Evans P and Kemp J (1997) Exon/intron structure of the human transferrin receptor gene. *Gene* **199**: 123-131.

Falzacappa MVV, Casanovas G, Hentze MW, Muckenthaler MU (2008) A bone morphogenetic protein (BMP)-responsive element in the hepcidin promoter controls HFE2-mediated hepatic hepcidin expression and its response to IL-6 in cultured cells. *J Mol Med* **86**:531-540.

Falzacappa MVV and Muckenthaler MU (2005) Hepcidin: iron-hormone and antimicrobial peptide. *Gene* **364**: 37-44.

Feder JN, Gnirke A, Thomas W, *et al.* (1996) A novel MHC class I-like gene is mutated in patients with hereditary hemochromatosis. *Nat Genet* **13**: 399-408.

Feder JN, Tsuchihashi Z, Irrinki A, *et al.* (1997) The hemochromatosis founder mutations in HLA-H disrupts beta2-microglobulin interaction and cell surface expression. *J Biol Chem* **272**: 14025-14028.

Feder JN, Penny DM, Irrinki A, *et al.* (1998) The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. *Proc Natl Acad Sci USA* **95**: 1472-1477.

Feige JN, Gelman L, Michalik L, Desvergne B, Wahli W (2006) From molecular action to physiological outputs: peroxisome proliferator-activated receptors are nuclear receptors at the crossroads of key cellular functions. *Prog Lipid Res* **45**: 120-59.

Fenton HJH (1894) The oxidation of tartaric acid in the presence of iron. *J Chem Soc Trans* **10**: 157-58.

Fickett JW and Hatzigeorgiou AG (1997) Eukaryotic promoter recognition. *Genome Res* **7**: 861-878.

Fleming MD, Trenor CC III, Su MA, Foernzler D, Beier DR, Dietrich WF, Andrews NC (1997) Microcytic anaemia mice have a mutation in Nramp2, a candidate iron transporter gene. *Nat Genet* **16**: 383-386.

Fleming RE, Ahmann JR, Migas MC, Waheed A, Koeffler HP, *et al.* (2002) Targeted mutagenesis of the murine transferrin receptor-2 gene produces hemochromatosis. *Proc Natl Acad Sci USA* **99**: 10653-10658.

Fleming RE (2005) Advances in understanding the molecular basis for the regulation of dietary iron absorption. *Curr Opin Gastroenterol* **21**: 201-206.

Fleming RE, Migas MC, Holden CC, Waheed A, Britton RS, Tomatsu S, Bacon BR, Sly WS (2000) Transferrin receptor 2: continued expression in mouse liver in the face of iron overload and in hereditary hemochromatosis. *Proc Natl Acad Sci USA* **97**: 2214-2219.

Flicek P, Ridwan Amode M, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, *et al.* (2012) Ensembl 2012. *Nucl Acids Res* **40**: D84-D90.
available at: www.ensembl.org

Fourel G, Magdinier F, Gilson E (2004) Insulator dynamics and the setting of chromatin domains. *Bioessays* **26**: 523-32.

Frazer DM, Wilkins SJ, Vulpe CD, Anderson GJ (2005) The role of duodenal cytochrome b in intestinal iron absorption remains unclear. *Blood* **106**: 4413.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comaritive genomics. *Nucl Acids Res* **32**: W273-279.
available at: http://genome.lbl.gov/vista/index.shtml

Fried M and Crothers DM (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucl Acids Res* **9**: 6505-6525.

Galaris D and Pantopoulos K (2008) Oxidative stress and iron homeostasis: mechanistic and health aspects. *Crit Rev Clin Lab Sci* **45**: 1-23.

Ganz T (1999) Defensins and host defense. *Science* **286**: 420-421.

Ganz T (2005) Hepcidin - a regulator of intestinal iron absorption and iron recycling by macrophages. *Best Pract Res Clin Haematol* **18**: 171-182.

Ganz T and Nemeth E (2005) Hepcidin and regulation of body iron metabolism. *Am J Physiol Gastrointest Liver Physiol* **290**: G199-G203.

Gao J, Chen J, Kramer M, Tsukamoto H, Zhang AS, Enns CA (2009) Interaction of the hereditary hemochromatosis protein HFE with transferrin receptor 2 is required for transferrin-induced hepcidin expression. *Cell Metab* **9**: 217-227.

Garrick MD, Dolan KG, Horbinski C, Ghio AJ, Higgins D, *et al.* (2003) DMT1: A mammalian transporter for multiple metals. *Biometals* **16**: 41-54.

Gaston K and Jayaraman PS (2003) Transcriptional repression in eukaryotes: repressors and repression mechanisms. *Cell Mol Life Sci* **60**: 721-741.

Gkouvatsos K, Papanikolaou G, Pantopoulos K (2012) Regulation of iron transport and the role of transferrin. *Biochim Biophys Acta* **1820**: 188-202.

Goodyer CG, Zheng H, Hendy GN (2001) *Alu* elements in human growth hormone receptor gene 5'-untranslated region exons. *J Mol Endocrinol* **27**: 357-366.

Goswami T and Andrews NC (2006) Hereditary hemochromatosis protein, HFE, interaction with transferrin receptor 2 suggests a molecular mechanism for mammalian iron sensing. *J Biol Chem* **281**: 2494-28498.

Gottesfeld JM, Neely L, Trauger JW, Baird EE, Dervan PB (1997) Regulation of gene expression by small molecules. *Nature* **387**: 202-205.

Green MR (2005) Eukaryotic transcription activation: right on target. *Mol Cell* **18**: 399-402.

Gross CN, Irrinki A, Feder JN, Enns CA (1998) Co-trafficking of HFE, a nonclassical major histocompatibility complex class I protein, with the transferrin receptor implies a role in intracellular iron regulation. *J Biol Chem* **273**: 22068-22074.

Gunshin H, Starr CN, DiRenzon C, Fleming MD, Jin J, *et al.* (2005) Cybrd1 (duodenal cytochrome b) is not necessary for dietary iron absorption in mice. *Blood* **106**: 2879-2883.

Gunshin H, Mackenzie B, Berger UV, Gunshin Y, Romero MF, Boron WF, Nussberger S, Gollan JL, Hediger MA (1997) Cloning and characterisation of a mammalian proton-coupled metal-ion transporter. *Nature* **388**: 482-488.

Günther V, Lindert U, Schaffner W (2012) The taste of heavy metals: Gene regulation by MTF-1. *Biochim Biophys Acta* **1823**: 1416-1425.

Guo B, Brown FM, Phillips JD, Yu Y, Leibold EA (1995) Characterization and Expression of Iron Regulatory Protein 2 (IRP2). *J Biol Chem* **270**: 16529-16535.

Guo Y and Jamison DC (2005) The distribution of SNPs in human regulatory regions. *BMC Genomics* **6**: 140.

Gupta M and Liu JS (2005) De novo *cis*-regulatory module elicitation for eukaryotic genomes. *Proc Natl Acad Sci USA* **102**: 7079-7084.

available at http://meme.sdsc.edu/meme/cgi-bin/tomtom.cgi

Gurevich I, Zhang C, Aneskievich BJ (2010) Scanning for transcription factor binding by a variant EMSA. *Methods Mol Biol* **585**: 147-58.

Haber F and Weiss J (1934) The catalytic decomposition of hydrogen peroxide by iron salts. *Proc Soc A* **147**: 332-351.

Haile DJ (2000) Assignment of *Slc11a3* to mouse chromosome 1 band 1B and SLC11A3 to human chromosome 2q32 by *in situ* hybridization. *Cytogenet Cell Genet* **88**: 328-329.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95-98

Han O, Failla ML, Hill AD, Morris ER, Smith JC Jr (1995) Reduction of Fe(III) is required for uptake of nonheme iron by Caco-2 cells. *J Nutr* **125**: 1291-1299.

Harris ZL, Durley AP, Man TK, Gitlin JD (1999) Targeted gene disruption reveals an essential role for ceruloplasmin in cellular iron efflux. *Proc Natl Acad Sci USA* **96**: 10812–10817.

Häsler J and Strub K (2006) *Alu* elements as regulators of gene expression. *Nucl Acids Res* **34**: 5491-5497.

He L, Vasiliou K, Nebert DW (2009) Analysis and update of the human solute carrier (*SLC*) gene superfamily *Human Genomics* **3**: 195-206.

Heckman KL and Pease LR (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols* **2**: 924-932.

Hellman NE and Gitlin JD (2002) Ceruloplasmin metabolism and function. *Ann Rev Nutr* **22**: 430 -458.

Hentze MW and Kühn LC (1996) Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide and oxidative stress. *Proc Natl Acad Sci USA* **93**: 8174-8182.

Hentze MW, Muckenthaler MU, Andrews NC (2004) Balancing acts: Molecular control of mammalian iron metabolism. *Cell* **117**:285-297.

Hentze MW, Muckenthaler MU, Galy B, Camaschella C (2010) Two to Tango: Regulation of Mammalian Iron Metabolism. *Cell* **142**: 24-38.

Herndon JM (2005) Scientific basis of knowledge on Earth's composition. *Curr Sci* **88**: 1034-1037.

Hicken BL, Tucker DC, Barton JC (2003) Patient compliance with phlebotomy therapy for iron overload associated with hemochromatosis. *Am J Gastroenterol* **98:** 2072-2077.

Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SM, Snyder M (2002) GATA-1 binding sites mapped in the β-globin locus by using mammalian ChIP-chip analysis. *Pro. Natl Acad Sci USA* **99**: 2924-29.

Hu M, Yu J, Taylor JMG, Chinnaiyan AM, Qin ZS (2010) On the detection and refinement of transcription factor binding sites using ChIP-Seq data. *Nucl Acids Res* **38**: 2154-2167.

Hunter HN, Fulton DB, Ganz T, Vogel HJ (2002) The Solution Structure of Human Hepcidin, a Peptide Hormone with Antimicrobial Activity That Is Involved in Iron Uptake and Hereditary Hemochromatosis. *J Biol Chem* **277**: 37597-37603.

Hutchinson C, Geissler CA, Powell JJ, Bomford A (2007) Proton pump inhibitors suppress absorption of dietary non-haem iron in hereditary haemochromatosis. *Gut* **56**: 1291-1295.

Ip E, Farrell G, Hall P, Robertson G, Leclercq I (2004) Administration of the potent PPARalpha agonist, Wy-14,643, reverses nutritional fibrosis and steatohepatitis in mice. *Hepatology 39*: 1286-1296.

Jaenisch R and Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33**: 245-254.

Jeney V, Balla J, Yachie I, Varga Z, Vercellotti GM, Eaton JW, Balla G (2002) Pro-oxidant and cytotoxic effects of circulating heme. *Blood* **100**: 879-887.

Johansson HE and Theil EC (2002) Iron-response element structure and combinatorial RNA regulation. In: *Molecular and Cellular Iron Transport*. New York/Basel: Dekker. pp 237-253.

Jones PA and Takai D (2001) The Role of DNA Methylation in Mammalian Epigenetics. *Science* **293**: 1068-1070.

Jurka J and Smith T (1988) A fundamental division in the *Alu* family of repeated sequences. *Proc Natl Acad Sci USA* **85**: 4775-4778.

Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, *et al.* (2010) Variation in Transcription Factor Binding Among Humans. *Science* **328**: 232-235.

Katoh M and Katoh M (2004) Human FOX gene family (Review). *Int J Oncol* **25**: 1495-1500.

Kawabata H, Yang R, Hirama T, Vuong PT, Kawano S, Gombart AF, Koeffler HP (1999) Molecular cloning of transferrin receptor 2. *J Biol Chem* **274**: 20826-20832.

Kawabata H, Tong X, Kawanami T, *et al.* (2000) Transferrin receptor 2-alpha supports cell growth both in iron-chelated cultured cells and in vivo. *J Biol Chem* **275**: 16618-16625.

Kawakami H and Lönnerdal B (1991) Isolation and function of a receptor for human lactoferrin in human fetal intestinal brush-border membranes. *Am J Physiol* **261**: G841-846.

Kel A, Kel-Margoulis O, Borlak J, Tchekmenev D, Wingender E (2005) Databases and tools for *in silico* analysis of regulation of gene expression. *Handbook of Toxicogenomics: Strategies and Applications.* Ed J Borlak, WILEY-VCH p253-290.

Kell DB (2009) Iron behaving badly: inappropriate iron chelation as a major contributor to the aetiology of vascular and other progressive inflammatory and degenerative diseases. *BMC Med Genomics* **2**: 2.

Kono S and Miyajima H (2006) Molecular and pathological basis of aceruloplasminemia. *Biol Res* **39**: 15-23.

Koppenol WH (2001) The Haber-Weiss cycle – 70 years later. *Redox Rep* **6**: 229-234.

Kovács KA, Steinmann M, Magistretti PJ, Halfon O, Cardinaux JR (2003) CCAAT/enhancer-binding protein family members recruit the coactivator CREB-binding protein and trigger its phosphorylation. *J Biol Chem* **278**: 36959-65.

Kim TH, Barrera LO, Qu C, Van Calcar S, Trinklein ND, *et al.* (2005) Direct isolation and identification of promoters in the human genome. *Genome Res* **15**: 830-39.

Knudsen S (1999) Promoter 2.0: for the recognition of PolII promoter sequences. *Bioinformatics* **15**: 356-361. available at: http://www.cbs.dtu.dk/services/promoter/

Knutson M and Wessling-Resnick M (2003) Iron Metabolism in the Reticuloendothelial System. *Crit Rev Biochem Mol* **38**: 61-88.

Kono S (2012) Aceruloplasminemia. *Curr Drug Targets* **13**: 1190-1199.

Korkmaz KS, Elbi C, Korkmaz CG, Lodi M, Hager GL, Saatcioglu F (2002) Molecular cloning and characterization of *STAMP1*, a highly prostate-specific six transmembrane protein that is overexpressed in prostate cancer. *J Biol Chem* **277**: 36689-36696.

Krause A, Neitz S, Magert HJ, *et al.* (2000) LEAP-1, a novel highly disulfide-bonded human peptide, exhibits antimicrobial activity. *FEBS Lett* **480**: 147-150.

Krentz AJ and Friedmann PS (2006) Type 2 diabetes, psoriasis and thiazolidinediones. *Int. J Clin Pract* **60**: 362-363.

Krishnamurthy P, Ross DD, Nakanishi T, *et al.* (2004) The stem cell marker Bcrp/ABCG2 enhances hypoxic cell survival through interactions with heme. *J Biol Chem* **279**: 24218-24225.

Kühn LC (1998) Iron and gene expression: Molecular mechanisms regulating cellular iron homeostasis. *Nut Rev* **56**: S11-19.

Kushner JP, Porter JP, Olivieri NF (2001) Secondary iron overload. *Hematology* **1**: 47-61.

Kutty RK, Kutty G, Rodriguez IR, Chader GJ, Wiggert B (1994) Chromosomal Localization of the Human Heme Oxygenase Genes: Heme Oxygenase-1 (HMOX1) Maps to Chromosome 22q12 and Heme Oxygenase-2 (HMOX2) Maps to Chromosome 16p13.3. *Genomics* **20**: 513-516.

Lambert LA, Perri H, Meehan TJ (2005) Evolution of duplications in the transferrin family of proteins. *Comp Biochem Physiol B* **140**:11-25.

Lang D, Powell SK, Plummer RS, Young KP, Ruggeri BA (2007) PAX genes: roles in development, pathophysiology, and cancer. *Biochem Pharmacol* **73**: 1-14.

Lareau LF, Green RE, Bhatnager RS, Brenner SE (2004) The evolving roles of alternative splicing. *Curr Opin Struct Biol* **14**: 273-282.

Larocque R, Casapia M, Gotuzzo E, Gyorkos TW (2005) Relationship between intensity of soil-transmitted helminth infections and anaemia during pregnancy. *Am J Trop Med Hyg* **73**: 7783-789.

Laufberger V (1937) Sur la cristallisation de la ferritine. *Bull Soc Chim Biol* **19**: 1575-1582.

Lebrón JA and Bjorkman PJ (1999) The transferrin receptor binding site on HFE, the class I MHC-related protein mutated in hereditary hemochromatosis. *J Mol Biol* **289**: 1109-1118.

Leibold EA and Munro HN (1988) Cytoplasmic protein binds in vitro to a highly conserved sequence in the 5'untranslated region of ferritin heavy- and light-subunit mRNAs. *Proc Natl Acad Sci USA* **85**: 2171-2175.

Levay PF and Viljoen M (1995) Lactoferrin: A general review. *Haematologica* **80**: 252-267.

Lewis CE, McCarthy SP, Lorenzen J, McGee JO (1990) Differential effects of LPS, IFN-gamma and TNF alpha on the secretion of lysozyme by individual human mononuclear phagocytes: relationship to cell maturity. *Immunology* **69**: 402-408.

Li J, Wang L, Wang L, Li F (2012) Structure and transmembrane topology of slc11a1 TMD1–5 in lipid membranes *Peptide Sci* **98**: 224-233.

Li J, Ning G, Duncan SA (2000) Mammalian hepatocyte differentiation requires the transcription factor HNF-4α. *Genes Dev* **14**: 464-474.

Li Q and Johnston SA (2001) Is all DNA binding and transcription regulation by an activator physiologically relevant? *Mol Cell Biol* **21**: 2467-2474.

Lieu P, Heiskala M, Peterson PA, Yany Y (2001) The roles of iron in health and disease. *Mol Asp Med* **22**: 1-87.

Lin L, Goldberg YP, Ganz T (2005) Competitive regulation of hepcidin mRNA by soluble and cell-associated hemojuvelin. *Blood* **106**: 2884-2889.

Lin L, Nemeth E, Goodnough JB, *et al.* (2008) Soluble hemojuvelin is released by proprotein convertase-mediated cleavage at a conserved polybasic RNRR site. *Blood Cells Mol Dis* **40**: 122-131.

Lindley PF, Card G, Zaitseva I, *et al.* (1997) An X-ray structural study of human Ceruloplasmin in relation to ferroxidase activity. *J Biol Inorg Chem* **2**: 454-463.

Lombard M, Chua E, O'Toole P (1997) Regulation of intestinal non-haem iron absorption. *Gut* **40**: 435-439.

Looker AC, Dallman PR, Carroll MD, *et al.* (1997) Prevalence of iron deficiency in the United States. *JAMA* **277**: 973-976.

Loots G and Ovcharenko I (2004) rVista 2.0: evolutionary analysis of transcription factor binding sites. *Nucl Acids Res* **32**:W217-W221.

available at: http://genome.lbl.gov/vista/index.shtml

Lopez V, Suzuki YA, Lönnerdal B (2006) Ontogenic changes in lactoferrin receptor and DMT1 in mouse small intestine: implications for iron absorption during early life. *Biochem Cell Biol* **84**: 337-344.

Maines MD (2000) The heme oxygenase system and its functions in the brain. *Cell Mol Biol* **46**: 573-585.

Majuri R and Grasbeck R (1987) A rosette receptor assay with haem-microbeads. Demonstration of a haem receptor on K562 cells. *Scand J Haemotol* **38**: 21-25.

Mason B and Moore CB (1982) *Principles of Geochemistry*. 4th edn, New York: John Wiley & Sons, Inc.

Maston GA, Evans SK, Green MR (2006) Transcriptional Regulatory Elements in the Human Genome. *Annu Rev Genomics Hum Genet* **7**: 29-59.

Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF (2009) RNA regulation of epigenetic processes. *BioEssays* **31**: 51-59.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, *et al.* (2012) Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**: 1190-1195.

McCay CM (1973) *Notes on the History of Nutrition Research*. Bern: Huber.

McKie AT, Barrow D, Latunde-Dada GO, Rolfs A, Sager G, *et al.* (2001) An Iron-Regulated Ferric Reductase Associated with the Absorption of Dietary Iron. *Science* **291**: 1755-1759.

McKie AT, Marciani P, Rolfs A, Brennan K, Wehr K, Barrow D, Miret S, Bomford A, Peters TJ, Farzaneh F, Hediger MA, Hentze MW, Simpson RJ (2000) A novel duodenal iron-regulated transporter, *IREG1*, implicated in the basolateral transfer of iron to the circulation. *Mol Cell* **5**: 299-309.

Merika M and Orkin SH (1993) DNA-binding specificity of GATA family transcription factors. *Mol Cell Biol* **13**: 3999-4010.

Merryweather-Clarke AT, Cadet E, Bomford A, Capron D, Viprakasit V, *et al.* (2003) Digenic inheritance of mutations in *HAMP* and *HFE* results in different types of haemochromatosis. *Hum Mol Genet* **12**: 2241-2247.

Merryweather-Clarke AT, Pointon JJ, Shearman JD, Robson KJH (1997) Global prevalence of putative haemochromatosis mutations. *J Med Genet* **34**: 275-278.

Metz-Boutigue M-H, Jollès J, Mazurier J, Schoentgen F, Legrand D, Spik G, Montreuil J, Jollès P (1984) Human lactotransferrin: amino acid sequence and structural comparisons with other transferrins. *Eur J Biochem* **145**: 659-676.

Meyron-Holtz EG, Ghosh MC, Iwai K, LaVaute T, Brazzolotto X, *et al.* (2004) Genetic ablations of iron regulatory proteins 1 and 2 reveal why iron regulatory protein 2 dominates iron homeostasis. *EMBO J* **23**: 386-395.

Michalik L, Auwerx J, Berger JP, Chatterjee VK, Glass CK, *et al.* (2006) Peroxisome proliferator-activated receptors. *Pharmacol Rev* **58**: 726-41.

Miret S, Simpson RJ, McKie AT (2003) Physiology and Molecular Biology of Dietary Iron Absorption. *Annu Rev Nutr* **23**: 283-301.

Monsalve M, Wu Z, Adelmant G, Puigserver P, Fan M, Spiegelman BM (2000) Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol Cell* **6**: 307-316.

Montreuil J, Spik G, Mazurier J (1997) Transferrin superfamily: An outstanding model for studying biochemical evolution. *Glycoproteins* **29**: 203-242.

Nardone J, Lee DU, Mark Ansel K, Rao A (2004) Bioinformatics for the 'bench biologist': how to find regulatory regions in genomic DNA. *Nature Immunol* **5**: 768-774.

Naylor LH (1999) Reporter gene technology: the future looks bright. *Biochem Pharmacol* **58**: 749-757.

Naylor SL, Yang F, Cutshaw S, Barnett DR, Bowman BH (1985) Mapping ceruloplasmin cDNA to human chromosome 3. *Cytogenet Cell Genet* **40**: 711.

Nemeth E and Ganz T (2006) Hepcidin and iron-loading anaemias. *Haematologica* **91**: 727-732.

Nerlov C (2007) The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. *Trends Cell Biol* **17**: 318-324.

Nicolas G, Bennoun M, Devaux I, *et al.* (2001) Lack of hepcidin gene expression and severe tissue iron overload in upstream stimulatory factor 2 (USF2) knockout mice. *Proc Natl Acad Sci USA* **98**: 8780-8785.

Nicolas G, *et al.* (2002) The gene encoding the iron regulatory peptide hepcidin is regulated by anaemia, hypoxia, and inflammation. *J Clin Invest* **110**: 1037-1044.

Nicolas G, Viatte L, Lou DQ, *et al.* (2003) Constitutive hepcidin expression prevents iron overload in a mouse model of hemochromatosis. *Nat Genet* **34**: 97-101.

Niederkofler V, Salie R, Arber S (2005) Hemojuvelin is essential for dietary iron sensing, and its mutation leads to severe iron overload. *J Clin Invest* **115**: 2180-2186.

Nilsson R, Schultz IJ, Pierce EL, Soltis KA, Naranuntarat A, *et al.* (2009) Discovery of Genes Essential for Heme Biosynthesis through Large-Scale Gene Expression Analysis. *Cell Metab* **10**: 119-130.

Noe L, Kucherov G (2005) YASS: enhancing the sensitivity of DNA similarity search. *Nucl Acids Res* **33(2)**:W540-W543.

available at: http://bioinfo.lifl.fr/yass/index.php

Oakhill JS, Marritt SJ, Gareta EG, Cammack R, McKie AT (2008) Functional characterization of human duodenal cytochrome *b* (Cybrd1): Redox properties in relation to iron and ascorbate metabolism. *BBA – Bioenergetics* **1777**: 260-268.

Ohgami RS, Campagna DR, Antiochos B, Wood EB, Sharp JJ, Barker JE, Fleming MD (2005a) nm1054: A spontaneous, recessive, hypochromic, microcytic anaemia mutation in the mouse. *Blood* **106**:3625-3631.

Ohgami RS, Campagna DR, Greer EL, Antiochos B, McDonald A, Chen J, Sharp JJ, Fujiwara Y, Barker JE, Fleming MD (2005b) Identification of a ferrireductase required for efficient transferrin-dependent iron uptake in erythroid cells. *Nat Genet* **37**:1264-1269.

Okazaki Y, *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563-573.

Osborne TB and Campbell GF (1900) The protein constituents of egg white. *J Am Chem Soc* **22**: 422-426.

Pabo CO and Sauer RT (1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61**: 1053-95.

Pak M, Lopez MA, Gabayan V, Ganz T, Rivera S (2006) Suppression of hepcidin during anaemia requires erythropoietic activity. *Blood* **108**: 3730-3735.

Pakdaman R, Petitjean M, El Hage Chahine J-M (1998) Transferrins: A mechanism for iron uptake by lactoferrin. *Eur J Biochem* **254**:144-153.

Pantopoulos K (2004) Iron metabolism and the IRE/IRP regulatory system: an update. *Ann NY Acad Sci* **1012**: 1-13.

Papanikolaou G, Samuels ME, Ludwig EH, *et al.* (2004) Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. *Nat Genet* **36**: 77-82.

Papanikolaou G and Pantopoulos K (2005) Iron metabolism and toxicity. *Toxicol Appl Pharmacol* **202**: 199-211.

Parajes S, González-Quintela A, Campos J, Quinteiro C, Domínguez F, Loidi L (2010) Genetic study of the hepcidin gene (HAMP) promoter and functional analysis of the c.-582A > G variant. *BMC Genet* **10**:110.

Park I, Schaeffer E, Sidoli A, Baralle FE, Cohen GN, Zakin MM (1985) Organization of the human transferrin gene: Direct evidence that it originated by gene duplication (gene organization/intragenic recombination). *Proc Natl Acad Sci USA* **82**: 3149-3153.

Park CH, Valore EV, Waring AJ, Ganz T (2001) Hepcidin, a urinary antimicrobial peptide synthesized in the liver. *J Biol Chem* **276**: 7806-7810.

Parkkila S, Waheed A, Britton RS, *et al.* (1997) Association of the transferrin receptor in human placenta with HFE, the protein defective in hereditary haemochromatosis. *Proc Natl Acad Sci USA* **94**: 13198-13202.

Pastinen T and Hudson TJ (2004) Cis-acting regulatory variation in the human genome. *Science* **305**: 647-650.

Patel BN, David S (1997) A novel glycosylphosphatidylinositol-anchored form of ceruloplasmin is expressed by mammalian astrocytes. *J Biol Chem* **272**: 20185-20190.

Patterson AJ, Brown WJ, Roberts DCK (2001) Dietary and Supplement Treatment of Iron Deficiency Results in Improvements in General Health and Fatigue in Australian Women of Childbearing Age. *J Am Coll Nutr* **20**: 337-342.

Pedersen AG, Baldi P, Chauvin Y, Brunak S (1999) The biology of eukaryotic promoter prediction - a review. *Comput Chem* **23**: 191-207.

Pedersen TA, Bereshchenko O, Garcia-Silva S, Ermakova O, Kurz E, Mandrup S, Porse BT, Nerlov C (2007) Distinct C/EBPα motifs regulate lipogenic and gluconeogenic gene expression *in vivo*. *EMBO J* **26**: 1081-1093.

Pedro M (2002) Emerging bioinformatics for the metabolome. *Brief Bioinform* **3**: 134-145.

Pietrangelo A (2006) Hereditary hemochromatosis. *Ann Rev Nutr* **26**: 251-270.

Pietrangelo A (2010) Hepcidin in human iron disorders: Therapeutic implications. *J Hepatol* **54**: 173-181.

Pilpel Y, Sudarsanam P, Church GM (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* **29**: 153-159.

Polak P and Domany E (2006) *Alu* elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* **7**: 133.

Poleev A, Wendler F, Fickenscher H, *et al.* (1995) Distinct functional properties of three human paired-box-protein, PAX8, isoforms generated by alternative splicing in thyroid, kidney and Wilms' tumors. *Eur J Biochem* **228**: 899-911.

Ponka P, Beaumont C, Richardson DR (1998) Function and regulation of transferrin and ferritin. *Semin Hematol* **35**: 35-54.

Poskitt EME (2003) Historical Review. *Br J Haematol* **122**: 554–562.

Qiu A, Jansen M, Sakaris A, *et al.* (2006) Identification of an intestinal folate transporter and the molecular basis for hereditary folate malabsorption. *Cell* **127**: 917-928.

Quigley JG, Yang Z, Worthington MT, *et al.* (2004) Identification of a human heme exporter that is essential for erythropoiesis. *Cell* **118**: 757-766.

Rakhshandehroo M, Sanderson LM, Matilainen M, *et al.* (2007) Comprehensive analysis of PPARalpha-dependent regulation of hepatic lipid metabolism by expression profiling. *PPAR Res 2007* 26839.

Recalcati S, Minotti G, Cairo G (2010) Iron regulatory proteins: from molecular mechanisms to drug development. *Antioxid Redox Signal* **13**: 1593-1616.

Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, *et al.* (2002) Position effect protection and enhancer blocking by the chicken β-globin insulator are separable activities. *Proc Natl Acad Sci USA* **99**: 6883-88.

Reddy MB, Hurrell RF, Cook JD (2000) Estimation of nonheme-iron bioavailability from meal composition. *Am J Clin Nutr* **71**: 937-943.

Reese MG (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput Chem* **26(1)**: 51-56.
available at: http://www.fruitfly.org/seq_tools/promoter.html

Roetto A, Papanikolaou G, Politou M, *et al.* (2003) Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. *Nat Genet* **33**: 21-22.

Roetto A and Camaschella C (2005) New insights into iron homeostasis through the study of non-HFE hereditary hemochromatosis. *Best Pract Res Clin Haematol* **18**: 235-250.

Rouault TA, Hentze MW, Dancis A, *et al.* (1987) Influence of altered transcription on the translational control of human ferritin expression. *Proc Natl Acad Sci USA* **84**: 6335-6339.

Rouault TA (2006) The role of iron regulatory proteins in mammalian iron homeostasis and disease. *Nat Chem Biol* **2**: 406-414.

Rowold DJ and Herrera RJ (2000) *Alu* elements and the human genome. *Genetica* **108**: 57–72.

Rozen S and Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds). *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386

Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucl Acids Res* **32**: 91-94.
available at: http://jaspar.genereg.net/

Scarpulla RC (2002) Nuclear activators and coactivators in mammalian mitochondrial biogenesis. *Biochim Biophys Acta* ***1576***: 1-14.

Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M (2012) Linking disease associations with regulatory information in the human genome. *Genome Res* **22**: 1748-1759.

Schmid CD, Perier R, Praz V, Bucher P (2005) EPD in its twentieth year: towards complete promoter coverage of selected model organisms. *Nucl Acids Res* **34**: D82-D85.
available at: http://epd.vital-it.ch/

Schmucker S, Martelli A, Colin F, Page A, Wattenhofer-Donzé M, *et al.* (2011) Mammalian Frataxin: An Essential Function for Cellular Viability through an Interaction with a Preformed ISCU/ NFS1/ISD11 Iron-Sulfur Assembly Complex. *PLOS One* **6**: e16199- e16199.

Scholl TO (2005) Iron status during pregnancy: setting the stage for mother and infant. *Am J Clin Nutr* **81**: 1218S-1222S.

Selva DM, Hogeveen KN, Innis SM, Hammond GL (2007) Monosaccharide-induced lipogenesis regulates the human hepatic hormone-binding globulin gene. *J Clin Invest* **117**: 3979-3987.

Sendamarai AK, Ohgami RS, Fleming MD, Martin Lawrence C (2008) Structure of the membrane proximal oxidoreductase domain of human Steap3, the dominant ferrireductase of the erythroid transferrin cycle. *Proc Natl Acad Sci USA* **105**: 7410-7415.

Severyn CJ and Rotwein P (2010) Conserved proximal promoter elements control repulsive guidance molecule c/hemojuvelin (Hfe2) gene transcription in skeletal muscle. *Genomics* **96**: 342-351.

Seyfert HM, Tuckoricz A, Interthal H, Koczan D, Hobom G (1994). Structure of the bovine lactoferrin-encoding gene and its promoter. *Gene* **143** (2): 265-269.

Shamm RL, Phatak PD, West C, *et al.* (2005) Autosomal dominant hereditary hemochromatosis associated with a novel ferroportin mutation and unusual clinical features. *Blood Cells Mol Dis* **34**: 157-161.

Shaulian E and Karin M (2002) AP-1 as a regulator of cell life and death. *Nat Cell Biol* **4**: E131-E136.

Shayeghi M, Latunde-Dada GO, Oakhill JS, *et al.* (2005) Identification of an intestinal heme transporter. *Cell* **122**: 789-801.

Sheftel AD, Mason AB, Ponka P (2012) The long history of iron in the Universe and in health and disease. *Biochim Biophys Acta* **1820**: 161-187.

Silva M, de Freitas Bonomo L, de Paula Oliveira R, de Lima WG, Eustáquio Silva M, Pedrosa ML (2010) Effects of the interaction of diabetes and iron supplementation on hepatic and pancreatic tissues, oxidative stress markers, and liver peroxisome proliferator-activated receptor-α expression. *J Clin Biochem Nutr* **49**: 102-108.

Silva M, Silva ME, de Paula H, Carneiro CM, Pedrosa ML (2008) Iron overload alters glucose homeostasis, causes liver steatosis, and increases serum triacylglycerols in rats. *Nutr Res* **28**: 391-398.

Simó R, Barbosa-Desongles A, Sáez-Lopez C, Lecube A, Hernandez C, Selva DM (2012) Molecular mechanism of TNFα-induced down-regulation of SHBG expression. *Mol Endocrinol* **26**: 438-446.

Smale ST and Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449-79.

Smith SR, Ghosh MC, Ollivierre-Wilson H, Hang Tong W, Rouault TA (2006) Complete loss of iron regulatory proteins 1 and 2 prevents viability of murine zygotes beyond the blastocyst stage of embryonic development. *Blood Cells Mol Dis* **36**: 283-287.

Sobczak K and Krzyzosiak WJ (2002) Structural determinants of BRCA1 translational regulation. *J Biol Chem* **277**: 17349-17358.

Spiegelman BM and Heinrich R (2004) Biological control through regulated transcriptional coactivators. *Cell* **119**: 157-67.

Steger DJ, Haswell ES, Miller AL, Wente SR, O'Shea EK (2003) Regulation of Chromatin Remodeling by Inositol Polyphosphates. *Science* **299**: 114-116.

Stein J, Hartmann F, Dignass AU (2010) Diagnosis and management of iron deficiency anemia in patients with IBD. *Nat Rev Gastroenterol Hepatol* **7**: 599.

Stocker R (1990) Induction of Haem Oxygenase as a Defence Against Oxidative Stress. *Free Radical Res* **9**: 101-112.

Stoltzfus RJ (2003) Iron deficiency: Global prevalence and consequences. *Food Nutr Bull* **24**: 99-103.

Strandberg Pedersen N and Morling N (2009) Iron Stores in Blood Donors Evaluated by Serum Ferritin. *Eur J Haematol* **20**: 70-76.

Stuart JJ, Egry LA, Wong GH, Kaspar RL (2000) The 3'-UTR of human MnSOD mRNA hybridizes to a small cytoplasmic RNA and inhibits gene expression. *Biochem Biophys Res Commun* **274**: 641-648.

Szutorisz H, Dillon N, Tora L (2005) The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem Sci* **30**: 593-99.

Takahashi N, Ortel TL, Putman FW (1984) Single-chain structure of human ceruloplasmin: The complete amino acid sequence of the whole molecule. *Proc Natl Acad Sci USA* **81**: 390-394.

Tallkvist J, Bowlus CL, Lönnerdal B (2000) Functional and molecular responses of human intestinal Caco-2 cells to iron treatment. *Am J Clin Nutr* **72**: 770-775.

Tandy S, Williams M, Leggett A, Lopez-Jimenez, Dedes M, *et al.* (2000) Nramp2 Expression Is Associated with pH-dependent Iron Uptake across the Apical Membrane of Human Intestinal Caco-2 Cells. *J Biol Chem* **275**: 1023-1029.

The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature 489*: 57-74.

The ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **22**: 636-640.

Theil E. (1987) Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annu Rev Biochem* **56**: 289-315.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl Acids Res* **22**: 4673-4680.

available at: http://www.clustal.org/clustal2/

Thomson AM, Rogers JT, Leedman PJ (1999) Iron-regulatory proteins, iron-responsive elements and ferritin mRNA translation. *Int J Biochem Cell Biol* **31**: 1139-1152.

Tomilin NV (1999) Control of genes by mammalian retroposons. *Int Rev Cytol* **186**: 1-48.

Torti M and Torti SV (2002) Regulation of ferritin genes and protein. *Blood* **99**: 3505-3516.

Trinder D, Olynyk JK, Sly WS, Morgan EH (2002) Iron uptake from plasma transferrin by the duodenum is impaired in the Hfe knockout mouse. *Proc Natl Acad Sci USA* **99**: 5623-5626.

Troadec M-B, Ward DM, Lo E, Kaplan J, De Domenico I (2010) Induction of FPN1 transcription by MTF-1 reveals a role for ferroportin in transition metal efflux. *Blood* **116**: 4657-4664.

Trujillo MA, Letovsky J, Maguire HF, Lopez-Cabrera M, Siddioui A (1991) Functional analysis of a liver-specific enhancer of the hepatitis B virus. *Proc Natl Acad Sci USA* **88**: 3797-3801.

Turlin B and Deugnier Y (2002) Iron overload disorders. *Clin Liver Dis* 6: 481-496.

Valenti L, Fracanzani AL, Dongiovanni P, Bugianesi E, Marchesini G, *et al.* (2007) Iron Depletion by Phlebotomy Improves Insulin Resistance in Patients With Nonalcoholic Fatty Liver Disease and Hyperferritinemia: Evidence from a Case-Control Study. *Am J Gastroenterol* **102**: 1-8.

VanGuilder HD, Vrana KE, Freeman WM (2008) Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **44**: 619-626.

van Helden J (2003) Regulatory sequence analysis tools. *Nucl Acids Res* **31**: 3593-3596.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, *et al.* (2001) The sequence of the human genome. *Science* **291**: 1304-51.

Venter M and Warnich L (2009) *In silico* promoters: modelling of *cis*-regulatory context facilitates target prediction. *J Cell Mol Med* **13**: 270-278.

Vilar JM and Saiz L (2005) DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev* **15**: 136-44.

von Recklinghausen FD (1889) Über Hämochromatose. *Taggeblatt Versamml Deutscher Naturforscher und Aertze* **62**: 324-325.

Waheed A, Parkkila S, Zhou XY, *et al.* (1999) Hereditary haemochromatosis: effects of C282Y and H63D mutations on association with beta2-microglobulin, intracellular processing, and cell surface expression of the HFE protein in COS-7 cells. *Proc Natl Acad Sci USA* **94**: 12384-12389.

Wallander ML, Leibold EA, Eisenstein RS (2006) Molecular control of vertebrate iron homeostasis by iron regulatory proteins. *Biochim Biophys Acta* **1763**: 668-689.

Wang RH, Li C, Xu X, *et al.* (2005) A role of SMAD4 in iron metabolism through the positive regulation of hepcidin expression. *Cell Metab* **2**: 399-409.

Wang J, Chen G, Lee J, Pantopoulos K (2008) Iron-dependent degradation of IRP2 requires its C-terminal region and IRP structural integrity. *BMC Mol Biol* **9**: 1-11.

Wang J and Pantopoulos K (2011) Regulation of cellular iron metabolism. *Biochem J* **434**: 365-381.

Wang W, Knovich MA, Coffman LG, Torti FM, Torti SV (2010) Serum ferritin: Past, present and future. *Biochim Biophys Acta* **1800**: 760-769.

Wang W, Zhong J, Wang Y-Q (2010) Comparative genomic analysis reveals the evolutionary conservation of *Pax* gene family. *Genes Genet Syst* **85**: 193-206.

Ward PP, Mendoza-Meneses M, Cunningham GA, Conneely OM (2003) Iron status in mice carrying a targeted disruption of lactoferrin. *Mol Cell Biol* **23**: 178-185.

Warren AJ (2002) Eukaryotic transcription factors. *Curr Opin Struct Biol* **12**:107-114.

Wasserman WW and Sandelin A (2004) Applied bioinformatics for identification of regulatory elements. *Nature Rev Genet* **5**: 276-287.

Wertime T and Muhly JD, eds. (1980) *The Coming of the Age of Iron*. New Haven.

West AR and Oates PS (2008) Mechanisms of heme iron absorption: Current questions and controversies. *World J Gastroenterol* **14**: 4101-4110.

West Ap, Bennett MJ, Sellers VM, *et al.* (2000) Comparison of the interactions of transferrin receptor and transferrin receptor 2 with transferrin and the hereditary hemochromatosis protein HFE. *J Biol Chem* **275**: 38135-38138.

WHO (World Health Organization) (2004) Assessing the iron status of populations. Report of a joint World Health Organization/Centres for Disease Control and Prevention technical consultation.

Wilson NK, Foster SD, Wang X, Knezevic K, Schütte, *et al.* (2010) Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell* **7**: 532-544.

Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnauser R, Pruss M, Schacherer F, Thiele S, Urbach S (2001) The TRANSFAC system on gene expression regulation. *Nucl Acids Res* **29**: 281-283.

available at: http://www.gene-regulation.com/pub/databases.html/

Worwood M, Summers M, Miller F, Jacobs A, Whittaker JA (2008) Ferritin in Blood Cells from Normal Subjects and Patients with Leukaemia. *Br J Haematol* **28**: 27-35.

Yamashiro DJ, Tycko B, Fluss SR, Maxfield FR (1984) Segregation of transferrin to a mildly acidic (pH 6.5) para-Golgi compartment in the recycling pathway. *Cell* **37**: 789-800.

Yang F, Naylor SL, Lum JB, Cutshaw S, McCombs JL, Naberhaus KH, McGill JR, Adrian GS, Barnett DR, Bowman BH (1986) Characterization, mapping, and expression of the human ceruloplasmin gene. *Proc Natl Acad Sci USA* **83**: 3257-3261.

Yoshida T, Biro P, Cohen T, Müller RM, Shibahara S (1988) Human heme oxygenase cDNA and induction of its mRNA by hemin. *Eur J Biochem* **171**: 457-461.

Yoshida K, Furihata K, Takeda S, *et al.* (1995) A mutation in the ceruloplasmin gene is associated with systemic hemosiderosis in humans. *Nat Genet* **9**: 267-272.

Zimmermann MB and Hurrell RF (2007) Nutritional iron deficiency. *Lancet* **370**: 511-520.

6.2 WEBSITE REFERENCES

BLAST: Basic Local Alignment Search Tool
available at: http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome

CLC Sequence Viewer 6.7.1
available at: http://www.clcbio.com/

ClustalW
available at: http://www.clustal.org/clustal2/

ENCODE Project
available at: http://www.encodeproject.org/ENCODE/pubs.html

Ensembl
available at: www.ensembl.org

Eukaryotic Promoter Database
available at: http://epd.vital-it.ch/

JASPAR CORE
available at: http://jaspar.genereg.net/

MEME
available at: http://meme.sdsc.edu/meme/intro.html

Neural Network Promoter Prediction
available at: http://www.fruitfly.org/seq_tools/promoter.html

Promoter 2.0 Prediction Server

available at: http://www.cbs.dtu.dk/services/promoter/

RestrictionMapper
available at: http://www.restrictionmapper.org/

TOMTOM Motif Comparison Tool

available at http://meme.sdsc.edu/meme/cgi-bin/tomtom.cgi

TRANSFAC®7

available at: http://www.gene-regulation.com/pub/databases.html/

VISTA Tools for Comparative Genomics

available at: http://genome.lbl.gov/vista/index.shtml

YASS: Genomic Similarity Search Tool

available at: http://bioinfo.lifl.fr/yass/index.php

APPENDICES

Appendix 1: Conference Presentations and Publications

Selected Conference Presentations

* NJ Strickland, MG Zaahl (2012) *In silico* analyses of promoter regulatory targets in the iron metabolism pathway. (Oral). Joint Congress of the Southern African Genetics Society South African Society for Bioinformatics and Computational Biology, Stellenbosch, 10-12 September.

* NJ Strickland, MG Zaahl (2011) *In silico* analyses of promoter regulatory targets in the iron metabolism pathway. (Poster). Joint Congress of the Southern African Society of Human Genetics and Africa Society for Human Genetics, Cape Town, 6-9 March.

NJ Strickland, T Matsha, MG Zaahl (2009) Molecular Genetic Analysis of Ceruloplasmin in Oesophageal Cancer. (Poster). International Bioiron Society Biannual Meeting (IBIS), Porto, Portugal, 7-11 June.

SH Krikler, CA Whittington, M-A Hallendorff, NJ Strickland, VR Human, NW McGregor, JA Vervalle, MG Zaahl (2009) Mutation analysis of seven genes involved in iron metabolism in a Mennonite population. (Poster). International Bioiron Society Biannual Meeting (IBIS), Porto, Portugal, 7-11 June.

NJ Strickland, T Matsha, MG Zaahl (2009) Molecular Genetic Analysis of Ceruloplasmin in Oesophageal Cancer. (Poster). South African Society of Human Genetics (SASHG) Annual Meeting, Spier Estate, Stellenbosch, 5-8 April.

*** Presentation directly related to the current study.**

Previous Publications Related to Work in the Iron Metabolism Field

Strickland NJ, Matsha T, Erasmus RT, Zaahl MG (2011) Molecular analysis of *Ceruloplasmin* in a South African cohort presenting with oesophageal cancer. *Int J Cancer* **131**: 623-632.

Panton NA, Strickland NJ, Hift RJ, Warnich LW, Zaahl MG (2012) Iron homeostasis in porphyria cutanea tarda: mutation analysis of promoter regions of *CP*, *CYBRD1*, *HAMP* and *SLC40A1*. *J Clin Pathol*; doi: 10.1136/jclinpath-2012-200988.

Appendix 2: Gene Promoter Sequences

The promoter (5' to 3') region of each of the genes investigated in this study was acquired from Ensembl (reference numbers indicated). The primer binding sites for primers designed for the amplification of the 2 kb promoter region are shaded in grey. Primer binding sites for the primers designed for the amplification of the CR element are underlined. The translation initiation site (ATG) is indicated in bold red text. The 5'UTR (black text), the CR element (green text) and the introns (blue text) are indicated in lowercase text. The first exon is indicated in uppercase text; untranscribed regions in purple and transcribed regions in black text.

***Aconitase* (*ACO1*)**; ENSG00000122729.

ttgtaagcttcctgaagcctccccagcatgcagaactgtgagccaattaaacctccttccctatgaattacccagtctc
agggttctttacagcagcgtgagaacggtctaatacagacacctttggaatatccaggtggggttgccacatatggctg
caaatcaagagtgtgaaggtcagaaagttaactaagaccaatgacagagacatagcaggcttcagctagaagtggcagg
cagtctgcatactgagaaaatagatgaagaaaaactcagcaaatggagaaacatataaagcaagaggagagagtgtagg
gaaagaccttaaaaaagagggacaaaagagggaagaggtgactcctaagagcggagggctgaaaagccaaaaggaggtt
aacatttctaggaaaacgaatcgccagtgtcaaataccacagagtagtgagtaaggacaggaacagaggacaggccctg
ggattcacgatctgcacgtcactggtgacaacgtccaagagcaacagactgggggatgtggaactatttccttcacgg
ttttgaccagcgtgcttactaataaatctcaattcccttcatagctctgatttgattgttttctgcgaaggattctcag
ttacacggatgttatcattgaacattagatactttgctttccgaatggctgcctcattaagaatcgcaatttagctgga
tgtgtcctcatcggacttccccagccttctactcactctcaccttgcccctactctccctacctcccaaggggactgga
gcatgatttttcagggattgtggtgcacctctgccatatggtggacattcaaagaaaatttaattactaatgtgaatgg
aaaggatttaggaagtgggtgggaaaacagtctcctctcatattggaaatacagttacttccccttcaaaataaatat
aagtatttatatttcaaagttaaatataactagacgatagttgtaccctatagtgcagtatgacagggaaggtgtaggc
tttggtgtcaaaccaacctaaggactgagtctcagctgttgcgcattgcacacctgcaaagacgggtaatattctttac
ttgtaaggtggttggaagagctgaactgaataacgtggaaagagtcgatcacagggctttgtgcagaataggagggcag
tataaagtagcaggggggctattagtactggctattgtttactcatcctataagctgcttattcagaaaactacagtaa
ctttccaaataatttgggcaggaaaacctcaccacctttcatctagaaagatctaatcacttctttgaacgccattcac
ccaagagggcacaactcctaccaacatgaaaagcagatcttgtgaaagcccacttgctgccatctcccagccatttctc
cgacttcagaccggtgaacatcgtgcattagaaagtgttgttacatgtccaaggaaatctacatatagaacttttgcca
cggaaatgctgaagtcctattctgacagcgtgactccgatgtccacttggcaggctttgcgcaaagaattaaggcgagg
gtggagagggtgggtagggtggggacggacaaagtgagaagggaggcgggccaagtgaccacacgtacatgaagcctct
ccagttatggccaaatcatttttaaaaataattccatttcttccccacccttggtcacttcctaacgccggctgggg
agtcacatgctaagctccaaagtgagcaacagacaccctcgtcccaggtcccatgctggcagaagcggacccgggccga
gggatcagccctcccacttccagcggaggagcccttcgcaaccctagcctccagcccggggcgaagtcctccaggcccg
cgggggcgcccgcaggctccgccccttcttcggctagctgcccgtccgtcccgcccctgctccgcccttcccgcccctg
ctccgcccatcgcgcggcggccccgGGGGCGAGAGGGGGCGGAGCGTGGAGGCGGCAGCTGGAACCGCGCAGCGCACGG
GAACGCGTCCCGCTGCTTGGGTCAGGTTCGCCGGTCGCGGGAGCCCCGCCGTGCAGTCGGAG

***Ceruloplasmin* (*CP*)**; ENSG00000047457.

acagtgttactgtttaaccatttcatttcaccaccaatatgaggggtgtgctatgcatacattctggaaccctcagcaa
ctatttgaacaggtttgtttgttggatggttggttggtacaatttgctgaaacaagtgactctgggcatggttacggtg
ctctgatcaacatctttccatgcacaagattggggctgagtttgagcactgagtgaaaaatgaaaagcaaatggctgag
acagttgctggttcacatcgtggcttcatgccaattagaaaaagaactgtccccttctccatgccctggggaggataa
acggctgaaaaatgccccccgtctgtgtggaccattaggggaaaagcagcttcttgcaaggcacagggtatggccggtg
gcatctgagcaagattttagtccttcgtcagtcctctcagtggacatgacggcacaattgttcaaggcagtcctagaag
gaggccttggggcatcctgggcaacatgattcagttttctgtaccaagccatttagttaagattattaccagctggtga

gagaagctcctccactgtctgccttaacataggtcatttcaattaacctgagattttcagatgaggaaactgaagctgg
gctgaagtatcttacttgaggtcatttggtagcatcagatttttccctagcactccttattctccttctctgctttatt
tttctccatagcattcaccactaatattgtgtatttcatttttcatttttatattttcagtttccccaagtagaatgta
agttgcatgatggtaggtgttttgttctctgcggtattcccagtgagtagaacagaacttggcacagactgctgtattc
ccagtgagtagaacagaacttggctcattcaataatatttcctaaatgaaggaattgttcaaatatcgcatgctttctt
ctataccaatctatttcaacaagtggccttcagcaaatacacaacgttttgaggtatttgagagaaaaaaaattatact
aacacaaagtgtcaactgttatcaggattttcttcagctaaaaaaaaaagccatttcatacttttagacatatacttcc
tgactaaaaacagggttttaagaaaaagtcttccaccttaaaattcatatccacttcaaagagaaaaatttcagtctaa
tgataaactggggtactttctatttactaagactaatctttgcctctgaagacaggtttagttatggtttgtttatttg
atgattaaaacaaagttattagcccctgttaggctctgctagttatttgcatatagcaatacagaatgatctggtaaag
tttttatgcagatttggtccaagctcataaaacagataaattcttccaaattaaattcttatttttgtttacttaaatt
tggcatgctagatatcatgtttcatatcataaagctgcaattaatcatttgttagacttttagaaatagtcatgcacca
cataataatgtttgagtaaatgacaaactgcatatgtgatggtggtcccataagattataataacatatttttactgta
cattttctatgtttgcatatgtatagatacagaaatatcattgtgttacaattgtctatagcattcagcatagttatag
actgtacaggtttgtagcctaggagcaataggtataccatatagcctacctagatgtatagtaggctatatcatataag
tctgtgtaagtacactctgtgatagtcacacaatgacaaaatagcctaacaatgcatttctcagaacgtatccctgtca
ttaagaatgactgtagtaaaaagctacatgttttgattcatttacataattcttatcctgatttactaaaagtgcact
gtttgcaggacactgtgctgtgacatattgtcagaggaacaaagggagcatttagacacgttctctgccctcctggaat
ttacacaatgaatggaaaaaggcacAGAGTTATGCACACCCTAATGCCTCCAACAATAACTGTTGACTTTTTATTTTCA
GTCAGAGAAGCCTGGCAACCAAGAACTGTTTTTTTGGTGGTTTACGAGAACTTAACTGAATTGGAAAATATTTGCTTTA
ATGAAACAATTTACTCTTGTGCAACACTAAATTGTGTCAATCAAGCAAATAAGGAAGAAAGTCTTATTTATAAAATTGC
CTGCTCCTGATTTTACTTCATTTCTTCTCAGGCTCCAAGAAGGGGAAAAAA**ATG**

***Cytochrome b reductase 1* (*CYBRD1*);** ENSG00000071967.

actctgagcccttaatcctgccctactaccatggcttgccactcctatttccttagcctgctttgaaaatcatccccgt
atacttcctgcacattctaaggtctttcaagacggcaacaagacaatatagcaaagtagtccttggagatgagctacag
aagagggctagaagtgaacgttcccaaatgcattctggaagcccctggcagtcagaaccctttttcttataatagcatt
agataatgctgacattgtttttgtccagtggagaagggtgaggataatgttggaaagagggcagggagcaagcttcact
tcatcaaggaggggctttggttcaaattggtttgggtgggcgggcactttgaaaatccatttgtcacactaaacggcaa
gtccaggtccaggaggttcctgtcttccctctctcaagagcaaactgcaagtagtttcattgcaggataaggccaagcc
caccgctgcccatgctgtttttgttttgtttccagacagagtctagctgtcaccaggctgcagtgcagtggtgcgatca
tacctctcctgagctcaagtgattctcttgcctccacctcccaaagccctgggattacagctgtgagacacctcatggg
gacccggtttactgggttttttttgttttttgttttttgtttcacagagtctcactttgtcacccaagctg
aagtgcagtggtgcaacctcagctcactgcagcctcgaccttctgggctcaagtgatcctccttcctcagtcccccaag
tagctggggctacaggtgcatgcatttgtattttcagtagagacagggtttcaccttgttgcccaggctgttctcaaac
tcctggactcaagtgatctgcccgccttagcctcccaaagtgctaagattacaggtgtgaggcactgcgcctggccttt
actgttaacttaaaacaaaaaattataaatttgaaaaaagaggagactttatttcttataaaaggttatagcctgcaag
gaggccattccataggctgataaacatagcctctggcctaagaccagagacaggcacttggaaggcagaggggttgggg
taggagctttatgctgaacagtttggccaaacatacatacgtaacaggttacaggaggagctatgaatattaatggagg
tggtccttacacatgcatattgaacaaacatgcatgtaacatgttctctttggggtggagacttaacatttaattgtat
tacttcaaatacatttaaatgtattacttcaaacctacacttcaaaaggtcttttcaggacgtgaatgcatacaagtgc
acaatccctgtacactggccagaaccggtccatggtcggtcttcttatcatgaaaaagttcctgaaatcagcccagtga
aagctgtagttctggctggtgcaacaggggttcagctggtcagcatctgtgaactgattaagttgtaattgttttaata
ttgcttatctcaagccagtgcttgtttagcctctagaggaaaagaaaaccctttgtggcagttagaccatagtttatttc
ttaagtgtaggagtgtgtgacttaggtccttttttataatttgatgtcttattgctacaaagagtctgttctgtccgcct
tatgatttctattttaaacattaatgctagtcagctgttgagtctaaattccaaaatggaggggggtagacttcccttcc
ggctgtagctagaaactcagctttaaggtttttctggggtctgcttggccaaggaggtccattcagtcagtgaggggct
taggattttatttttagtttacaccaccttacaggttcctccttaacacccttttgatgtcagagacagccaattctcca
gcactgtggctaaggaggcttcgttagtagttaaaagcacacacataaccctacccaacggatccctctcttctcttaa
ccaagagggcgcaccactgccctactcaacctccccacaaataaggacgagacacgggaagtgttaaaagcaagtggtg
agtataaaatcctcggaattttctctttgagcaactaactaggtctgttactgaagccctctgcgagcttggatgctgg
acgagggagagggtgaggccaccaggcaatgagcgccctcggcggccgcgtgatcccgggggtggggcCCATTTCTGAG
TTGGGGCCAGCTCCCCACCCCCAAGAGGCCCCACATTCCGGGCCAGCAGCCCAGAAAGTCCCTCCCCGCAGGCGGAGAC
AGCCCCAAGAAGTCGACGCCCCGGTCCCGCCGCCCGGCCACTACCCAGAGGGCTGCCGCCGCCTCTCCAAGTTCTTGTG
GCCCCCGCGGTGCGGAGTATGGGGCGCTG**ATG**

***Ferritin Heavy Polypeptide 1* (*FTH1*)**; ENSG00000167996.

tatttcttgttaactttattccaaggtactttaagcttttgttgtcaatgtaagtgaaaaacatttttttttccattct
ttttttccccatattttctaactggtggttactgttataaaggaaagctttggatcctatatatggtgactttgctatac
tcttatcctaagagattcccagtgtaaggtttctaaaaacaaagctaatgaggtgtctctatgctgtctctaggatcaa
gtttaacctctttttatggaacacaaccctgcccgatcttgcccctattctccaccctaaccagcatccccccaacataa
actgtactttccaggccaacattctttctttctttctttttttttttttttgagacggagtttgttttgtcttcc
aggctggagtgcaaaggcgcgatctcggctcactgcaacctccgcctcccaggttcaagcgattctcctgcctcagcct
ccagaatagctaggattacaggcgcatgccaccacgcccggctaatttttgtattttcagtagagaagggtttagccat
gttagttagccaggctgatctccaactccgacctcaagtgatccgcccgcctcggcctcccaaaatgctgggattacag
gcatgagccaccgcgcccagccccaggcaacatattttcttaaggcagctttaacaggccatgcatttccacatttcca
cacctttgcatatgctgttcactctgctccaaattcccttccctgtttctctcccgcacttcaccttcctttaacttt
ttcttcaaactcagctcacgtgtcttgcttagggagggcttccccgacctcctcctatcccagactgggtaaggtactc
ctggtctagcagcatccttctgtttaccttattcttgtagctctagcaccactaaaacgtttttcctgtaaataacata
ggtgcatttttaaatgaaaccgtgaacaagaaatcaccctaattccgtcggcaatttattttagtatttcagtagagtt
gctttttttaaactaaggtttggaagtctagcttccagctgcaggtttgtgagcatcctgaaattcagttcttccaccga
tgcgtcaaaagacgcgacagaacagtggacctcgaagccaggacctaggtaagggcagggcctagctctagctctgcca
cccataaattcacatccaccctgtgcctgggactcagcttagccacacacaccggccccacccctccggggtctctttc
acctgtcacccgccattaaccccgtcagcttctctccaatgacaccccagggttgtcctacagggccaccgaagcctga
agctcctgtgagtggggccgataagaggccattttgaaaagtcctcagcatgggcttcgcctttagatccacgtcgttt
ctcctttctaacccgagatgtgtcccttgggtgcgaagcgacaccagcgattctctcggaggccccggtggctcacccc
cacgactccccggggaggcggtccccatccctagggaaacccggcccgcccagcctttgttgcagaaaccgcacacgga
gccatatttgctttcagccaaaggcagacgtggtcggaggagtcggaagggcaggggcggacccgaaacccgccgtgcg
ccccacttctccatcgcgctccttcctggactctgaggaagagaaaatgcgctcgaccacctaaggaccgcgtctccac
cttccaccttcccaatcatctgcagcttcccaatctgcagcaaggtgtcacaagggcttctccgcagctatgatttcta
acaccatgccccgggccaggagcctcagggtgcccgttcgggaaaatgggagccgaatcaggatcaccccatgcgcccc
cgcacccttcccccgccggttccaacgcccgggcgcccctagaccgggggggggttctgaggtggacttcctgcgcct
cctcgcagcttccctccaactccactaaacgggcacagagacgccaccgctgtcccagaggcagtcggctaccggtccc
cgctcccgagctccgccagagcgcgcgagggcctccagcggccgccctccccacagcaggggcggggtcccgcgccc
accggaaggagcgggctcggggcgggcggcgctgattggccggggcgggcctgacgccgacgcggctATAAGAGACCAC
AAGCGACCCGCAGGGCCAGACGTTCTTCGCCGAGAGTCGTCGGGGTTTCCTGCTTCAACAGTGCTTGGACGGAACCCGG
CGCTCGTTCCCCACCCCGGCCGGCCGCCCATAGCCAGCCCTCCGTCACCTCTTCACCGCACCCTCGGACTGCCCCAAGG
CCCCCGCCGCCGCTCCAGCGCCGCGCAGCCACCGCCGCCGCCGCCGCCTCTCCTTAGTCGCCGCCATGACGACCGCGTC
CACCTCGCAGGTGCGCCAGAACTACCACCAGGACTCAGAGGCCGCCATCAACCGCCAGATCAACCTGGAGCTCTACGCC
TCCTACGTTTACCTGTCCATG

***Ferritin Light Polypeptide* (*FTL*)**; ENSG00000087086.

cagggtctctgtacccctctgtctccgggtctgttttcccctctctctgggtctctgtccccctctctctcagggtctc
tgtctgagtctctttttctctgggcctctgtctccccacccacgttctctatggtcttcagatcctgtctctctcagc
ccctcctctccctccacctatgggccccaaggcccctctcggccgactaaaggtctctgggctgaggtatgacatttgg
taaatgagtaactcgggccccacagtcaacggggtgctttggacggctgcccctcctccatggactcttcccagagacc
ccaagtcaggggagagagaggggcagaggatcaatgagtgtacactgagccaccacccagagtgaggggacaatatggc
cacaggcaggatcccagaaaaaaaaaaattatgggaaccaaaaagacacggagacaaagcacacagccggcttccaaca
agggggaccttggggtaggggcccgatttgggtctgagccaggtctagttctgagggtcccaccaaaaaagtcgggg
agtgccaaggggatgaggtcagcgagtggccccacccccatccgtccccgcgggggccccttcccgcgctgagctcagc
atgactcagcagtcgctgcctgcaacaaggaccccagataccggcctggcggggcatctgcagatcccctcggtcaggg
gctgccccgcctgggctgttcggagcagctcctctgtagggcagggcctgacctcaccccacaccccagtggcttcagc
tccctgcccagcgcacgggctgcctccatcagggcacttaagtagagtcaaggtttgcgccgactccgtctcccagaaa
tcaacgtggaccaggcccacttaaaaaatatgcattagtcgaagacagacgtctcgtccataggtccccagactgtttg
tccacaacgcagggcttctctttgtgggcctgaaggagcacacgtcaggcaagacacgagggcaacttcagagacactg
ttcagggccacggagaattctcgaagtctgtgagttgtctccaaaacagcctccaggaaatgcaatcccagagacaccg
caaagctcaaaacagacttttcattcctccccatttccactcccagaaaaagcctgggcccctgaaagagctccaaggt
atccccaaagtcccagcccatctttggagggttccagagggctgtccagggttccagaagtcaccctacctcatacaat
gctcctcctgggcttcccgcactgggtgaaacccacggctccagaataactgccctttccaacttgcagagcacctcaa
gaacctaaagatttccccggggagggctccaaaatatagagagtttgcatcgtgaaccccccacccctgtaaatcccagg
actccagtccccgcaacactccatcttccagagatctccaggggtccagaaataacacctcacagccttccaagcttcc
caaggatccggagactccttatctcgcagcgcaaacctcaggcaactacaggtcctccctgcaccttgaggtgcccct
ccccccactcaaggttccagaaaaatccatgccctcggaaaccccagaatcttactttcccgaggcaccccaggacccc

agggtcatcagttcacagggctgttagtgctcccataaaactctagcctcccagagggaccccagcccccaccctcccg
cccacgaacccctgcatttccagaatcagccccagggccccaacccccccaagcccccatttcacaacacgctggcgct
acaggcgcgtgacttccccttgctttggggcgggggctgagactcctatgtgctccggattggtcaggcacggccttc
ggccccgcctcctgccaccgcagattggccgctagccctccccgagcgccctgcctccgagggccggcgcaccataaaa
gaagccgccctagccacgtcccctcGCAGTTCGGCGGTCCCGCGGGTCTGTCTCTTGCTTCAACAGTGTTTGGACGGAA
CAGATCCGGGGACTCTCTTCCAGCCTCCGACCGCCCTCCGATTTCCTCTCCGCTTGCAACCTCCGGGACCATCTTCTCG
GCCATCTCCTGCTTCTGGGACCTGCCAGCACCGTTTTTGTGGTTAGCTCCTTCTTGCCAACCAACC**ATG**

***Hepcidin antimicrobial peptide* (*HAMP*)**; ENSG00000105697.

gatgcagcgatgcctggtagggctgggggctgctcctgtgtctccccaggtgagcacacccctattcactgggccctgc
ttcagcctgcagcaccttcaactcccaggagctgggcttgccactctgctcaccttgtggagctccatctgcctttcc
tccccaattcccccactccctgcactcgtctcttcccacaagagccctgtctccttttcctagctattcccatctgagg
ccatctttattcatttagtttttagagacagggtttcactctcacccaggctggggtgcagtggcacacaatcacggct
cactgcagccttgaccaactacaggtgcgtagcaccacagccaagtttttgtatagatggggtctcgctttgttaccca
ggctgtgacaagaggagcctcccacgtggtgtggatgaggaggcagatggcagggcctgtgcatttctgtgcttgagtg
ggccttgaaagtggttcagcaaccaggaagaagtgttcattcctcgacaacaacatccccgggctctggtgacttggct
gacactggatggccctggaatgaaaaaggcaaagaggcaaaatgtgcaagggcccatctggaaccaaggtttgttgatc
ccctgggccgtgtgcaccctgagctgggcctggtagtggaaaggaatgaaggcactgcagtcaggcagcctgggttcat
cccccagctagtggtgtcctaaggaaccggctccccaaaaacatccctggcttgtagtgcttgccaatttctgggtgtc
aagactcccactgctgctgatttcaggataccagcatgatgccactgaatgcagagtttcgagatgtgcatggtctgct
atgttgagccaggtctagcataccgctgtgccctgctgtgttttaggggagatggggaaacctggtgggtaagagcaaa
agccctggagtcaggctgtccaggctagaatctcagctctgcctctggctgagcaagcttgggccatgccctgatctct
gccttcagtgccttttctgtaaagtgaaggaaatgagtgtccgacggggaggaggttcctaaaagggagcagggtctgg
ggagcccaggcctctggggttgggtgactgagaaggcagcccctgaatacagagcagagctgaaggtggggcagtaagt
gctgctgggagaacaggcagcacaggctgagttggtgcagaagtgagtcaacatatgtgccatcgtataaaatgtactc
atcggactgtagatgtt**agctattactattactgcta**ttttatgttttatagacagggtctcactctgtcacccaggct
ggagtgcagtcacacaatcatagctcactgcaacctcagcctcctgggcttaagcgatctgcctcagcctcccaagtag
ctgggactacagatgtgtgc**caccacgcctggctaaatttg**tttaaaatttttttgtagagatggggtctccctatgt
tgcccaggctagtcttgaacttctgggctcaagcgaccctcctgccttggcctcccaaattgctgggattacaggcata
agccactgtgctgggccatattactgctgtcatttatggccaaaagtttgctcaaacattttccagttaccagagccac
atctcaagggtctgacactgggaaaacaccacgtgcggatcgggcacacgctgatgcttgccctgctcagggctatcta
gtgttccctgccagaacctatgcacgtgtggtgagagcttaaagcaatggatgcttccccaacatgccagacactcct
gaggagcctggcggctgctggccatgccccgtgtgcatgtaggcgatggggaagtgagtggaggagagcggaaccttga
ttctgctcatcaaactgcttaaccgctgaagcaaaaggggaacttttttcccgatcagcagaatgacatcgtgatggg
gaaagggctccccagatggctggtgagcagtgtgtgtctgtgaccccgtctgccccacccccctgaacacacctctgccg
gctgagggtgacacaaccctgttccctgtcgctctgttcccgcttatctctcccgccttttcggcgccaccaccttctt
ggaaatgagacagagcaaaggggaggggctcagaccaccgcctcccctggcaggccccataaaagcGACTGTCACTCG
GTCCCAGACACCAGAGCAAGCTCAAGACCCAGCAGTGGGACAGCCAGACAGACGGCACG**ATG**GCACTGAGCTCCCAGAT
CTGGGCCGCTTGCCTCCTGCTCCTCCTCCTCCTCGCCAGCCTGACCAGTGGCTCTGTTTTCCCACAACAG

***Hephaestin* (*HEPH*)**; ENSG00000089472.

tacccagagcaagcctagggctgtcacacatcacagcaatgcagagcaacctttccaaagcataacagagtctgctggt
cagaagtggtggtggggagtggcttcagagatgtggtttctgggggacaaccccctcttccctgcacccctccaccttt
ccccagaggtcttctttttaagaaactgctaagggcaagctcatcagaggatcctgaggtaaaggggttagactgtctg
tgccaaggcactgcccatggcttttaatttctcctggtacctgatactgtcacaaaatttaactgtggaaaatatagg
aaggaaaacaacatttactgggggtctgaacatgctgagcactgtgctaagggttcacacaaatgccagctgatttaat
cctcacaatatttcagtcatggaagtgttattattcctgttctgcagatgagggaccagaggctcgaaaaagttaagtg
atttgcccgaggtcacatagctcagaaaaaggcagccaagatacaatccttagtctttctgactccaaatgccatgctt
ctcaccctagaccatgttgactctaatgtgaaggtccctggggtcctctcctctttgtagccaccccagggccatgtgt
tattgatggcccatagatgagtaacctggaaatattggttgaatgaataaacgaaaagatgttctagtgacagagttta
acacatgaaggacccttaaggatattctggtcttacggttttccaatctgtttttagtgtcctagacatttcttactgg
tgctccagacacaatcacattgtctaggaaaccagtcatacccagtggacaggttcctctgtctcttcacccttgccca
tttctatctccttccctgccttacctagaacaaatcactttcatcttccataagtagtgggtgtttactaagattctgt
ttggggaaagggttctgctgccaaaattaattttgagaaacccaaatctagaccacctttttcatgttataggtgaaga

aactgagtgatagaaaggcaaaaagacttgtccagggtcactgtgtgctctttagtagcataactagaatgtgaagcca
ggattcttgaaacattcttctttttgacacaagttaatacagtttgtgtcctaacgaggaaataattatttagatttc
aacaattatttagaattgaaaaattaaaaattataactctaatttttgcctttactatgttccagttattgtgctaaat
agttaatacgtatagccttatttaagttcttataacagtcttataaaataggtaatattattatctccatttgaaaact
gaggtttaggggattaaattactttaatcatagaatcaggaaatggcagagttgggatttcagttcagaatatctgat
ttcagttttcatgctcttaacagtgatgccatattgcaaatgtctatattgaaaaaaaagtctatttagcctgctgctt
gttttgtaaattaatttttattggcacaatgatataccttgagagctacaaagacaaatttgtgtagttatggcaaaca
gaaactgtatgctctggaaataaatgatttagtagctagccctttatagaaaaaatctgccaacctctggtctgtagac
accaactttaaaatgtttttcaagtgacctccaggagaaggaaccagaaatgtaccccggatttgcagaaaaggtgggt
gctgtttgtactctcagtgaaattggtgaatgaataatacaaacctggaaacctggcctacctggcctcctgcagtctt
cactgtgttgattcaagtctattatgacctgttcttgttgtctttcatgagtcagagcagaggccttatcccttcccct
ccctgtcataaatgccttagagcttggacttggatctcagcttaatcagttttgtctccacttcccctttcttgcttca
aggttcacatgtgtgcgaggcagatAATTCTAGGGGCTGCCATGGCTTGGCTAAGCTTCCCTTGACCACTGAGCATTTC
TAAGGGAGTTGAGGCTGGTGGCTCCTCCTTCCTTCCTACTGGTGCTTCCACCTGCCTTGGTCTGAGTTGCAGTCCATGG
GGCAGCGCCTAAGTGTCTGAGCACACTTAAGAATCTCTAGTGGTTTATGACCCAGACTTTGCCCTACCACCTCAGTCTT
CTGAATGTTCTCTTCCCTGGACCCTGCTCCAGACACTTTAAATTCAGAAGAGGAAAATGTGCCCAGCCTGCCTGGAGAA
AAGTGTCTGCTCCTAGCCAAGATCTCCTCATCACAAA

***Hereditary Haemochromatosis* (*HFE*)**; ENSG00000010704.

tactaaatgttccaaatagaaagaaagtagattaaagattgtctagggctgggaggggaggaggaggagaatggagga
tctgggaggcgatgaattaagggtactgaattaggggaacatggtttctttttggggagatagaaatgtaaaatttgg
taatggttgcacatctctgtaaatatactaaaatccattaatttttgcatttttatttttatttttatttttgagacg
gagtttcgctcttgttgcccaggctggagtgcaatggcatgatctcggctcaccacaacctccgcctcccaggttcaag
cgattctcctgtctcagcttcccgagtagctggaattacaggtgcatgccaccacacccagctaattttttttttttga
gatgaagtcttgctcttgtctcccatgctggagtgtgatggcacgatcttagctcactgcaacctacgcctcccgggtt
caagcgattctcctgcctcagcctctaatgtagctgggattacagttgcctgctatcacgcccggctaatttttgtatt
tttagtaaagacggggtttcaccatgttggccaggctggtctcgaacccctgatctcaggtgatctgccttcctcggcc
gcccaaagtgctgggattacaagcgtgagccactgagccgggccaatttttgcattttacatagatgaattatatggta
tgctaattatatcttaccaaaaattgaaaaaaaggaatagtactatcagccccaatgtgcccatcatcaatattctaca
tgtttccaatgttattgtatctgtttgcctacagtcgaggccctgatatcctgtttgattttcttgaattgccaaaatt
tgcatacatgcttacaaaaataatgcctgttgaatttgctagatatgtaaaggtttggagcaaatcaggtgtattaaat
ttattaatattgtttgaaatgtctaaggcaataattcccaaacttcgttgagggagaaggaaagctttaaaatcccat
tgcccaggtggcatcccatactgttactgggaattatgcattgggatggatcctttaaccgaggagattattatagccg
gagctctgaaccagcaatctcagttcttgtgatagtgagcaaagaactacaaactaacaccaaaatgcaagcttaaagc
aaagtttattgaagcacaataatacactctgagggacagcgggcttatttctgcgaagtgaactcagcacttctttaca
gagctcaaggtgcttttatggggtttgtggggaggagttgaggtttgggctgtatctgagtgacaggatgatgttattt
gattgaagtgtatagctatacaatctaaaattaaactgtgcatggtcttacctataatttgttaagaaaagcctcccag
ggatgggggggcaaaactgtatgtaaattctattataatgatggcatgatgaacttggggtgaacttgaagacaggctt
ttgtgttgttgggcatgtgccaccttagggaatttccacctgtaccctccttttctctttctccaggatattttggccac
agactttatcataaactccatcccttagggtggcattagggtagtcttgggcctgaatttaggtgggccagtggctgtc
ttagtgacagcctttccgctctcttctgtcatcccctcccaactgctaatgtctaactacctaacaattacccattaaa
tcagtgtgtctggggttaggagcaggcctcaatatgtttaatcattctccagataatcccaatactgtaaagtttgtga
aacacttgtcagataattcaattatgaaggctgtggaaggtgtttcagtaggatctaattggttaatgttatgacttaa
ttaatttgaatcaaaaaacaaaatgaaaaagctttatatttctaagtcaaataagacataagttggtctaaggttgaga
taaaatttttaaatgtatgattgaattttgaaaatcataaatatttaaatatctaaagttcagatcagaacattgcgaa
gctactttccccaatcaacaacacccccttcaggatttaaaaaccaaggggacactggatcacctagtgtttcacaagc
aggtaccttctgctgtaggagagagagaactaaagttctgaaagacctgttgcttttcaccaggaagTTTTACTGGGCA
TCTCCTGAGCCTAGGCAATAGCTGTAGGGTGACTTCTGGAGCCATCCCCGTTTCCCCGCCCCCCAAAAGAAGCGGAGAT
TTAACGGGGACGTGCGGCCAGAGCTGGGGAA**ATG**

***Hemojuvelin* (*HFE2*)**; ENSG00000168509.

ttttctggcattatagtatatcccacttaggctgtcatagtcaaattattatgttttaaagtcacttgaatagtttggt
tagaagcattttacatttctataatgaaactgcttgtgatatggcctctaatggttgagaaatatttgtcatatatata
tacctgagaagagtgctatgagggcctctagactctgtattaaaatagagccaactggtaaagatggcttagtgattgt
gttggttattactgagtgtcaatttgattggattgaaggatacaaagtattgatcctgggtgtgtctgtgagggtgttg
ccaaaagaaattaacatttgagtcagtgggctgggaaaggcagatccacccttaatctgggtgagcacaatctaattca
ctgccagcacagctagaataaaaagcaggcagaaaaatatgaaaggagagactggcctagcctcccagcctacatattt

ctcccatgctggatgcttcctgcccttgaacatcagactccaagttcttcaattttgagactgagactggctctccttg
cccctcaagcttgcagacagcctactgtgggaccctgtgatcgtgtaagttaatacttaataaattcccctttatttat
atatctacctatatagatatccatatctatatagatattaataaatctagagagacagaaagcagactggtgatggcca
gtctagatggctagatagatagacatggatatagatatagatctctatatagatagaggtagatacagatatagatata
tgccctattagttctgttcctctagagaaccctaatacagtgaccgtatttggaatcggtccttctgttaatttcactt
ggcaagtactaaaagatgatgatctcagatatacctatggctgcaaaaacatgacatggctaaatcccttggttgcagt
atctctttcttttttaaggggggtgggggggcgggtctcactgttgcccaggctggagtgcaatggcgttatcatagc
tcactgcagcctcaaactcctgcgctcaagtgaccctcctgcctcagctcccaaagtgctgagattttgcaatatttat
ggtcacaagattatgttattccataaaagtatctttctgaggctaggcatgttggttcacacttgtaatcccagcactc
tgagaggctgagatggaaggattcattgaggcaaggagttcaagaccagcctggtcaacatagtgagacctcatctcgg
aaggaaggaaggaaggagggagggaggaagggagggagtgaaggaaggaaggaaggaaggaaggaaggaaggaaggaag
gaaggaaggaaggaaaagtatatttttgaatctttttctatttctccaactctttctttagaagaattctatttccatt
ctttcttcacctctttgcctttgttagccttctctccaagcaaatcgggagcctttatttttgtgtattcatgaggga
gaggaagatgaattgctgtacaaactaaagtaatgaaaatggagtaggtaggaggatagacagctgcaaggatctgagc
tggatagactgaacaaaccctcatcctaagcaactcacagctcagatttcttctctggacagctggctttttcgtcct
tctgaaatactctgcaaagataggagaggggctatgaactacctctgctatggatcttattcaaagtcagctacctcct
agatactatctgtagaacctaaatgtaatattcagcatagcagggatgaacatggtaaatgaaaggtatccaattgccc
actgtaatttttaaaggccaggagctcaacattattgaaaatgctggagggctgcctggagtaggcagtgaccacagag
tcacacaagctggaattggatatccaacttgtctgtcatatttctctcctccctccctgacttggcactcaatactcca
tattctttctaatcctctaaccctccccactcccccaactcccacaccctaccccccaccaacgttcctggaattttgga
cttagctatttttaaaaccgtcaactcagtagccacctccctccctgctcagctgtccagtactctggccagccatata
ctccccctttcccccccataccaaacCTTCTCTGGTTCCCTGACCTCAGTGAGACAGCAGCCGGCCTGGGGACCTGGGGGA
GACACGGAGGACCCCCTGGCTGGAGCTGACCCACAGAGTAGGGAATCATGGCTGGAGAATTGGATAGCAGAGTAATGTT
TGACCTCTGGAAACAgtaagtcaaaatgaaattgcaattcctttaataagcttttatattgaagttagacttttataaa
attacaaacacctacttggatgtctctcgtccaaatgctgggatctctccctaccaaggtgccccaatctccatttctc
tttctgtcttatttctttctggcctctggcctctagcttttgaagtttaattctctgtctctcctctggcagtcttag
ccctctctttaccttattacctcaagactcctgatgaagttttagaaggagttccctacgtcctctattctgtagtttt
cttaccaaggccaaatatgacctcagatgatgagtcactgataccttctatcctgcccccacttagcaatgcccttca
cattgagattccaagcatgggggctgctccctgtaaatgatttctccccacaactctagtccctccattctattctccc
tcttgcaggactcttcccccaatcatatccttacccataagataggggagttaggcaggagggatttagccccctctcca
actcctgtcatcataaaagactgagaacttcagaatttgaaaagaagagattaatggaaggagtgatatttgggaaaat
acaagaactgttgacttagaaaaaacaaatattgatttgcatgtttggtttgcatcccattattccatgagagagggag
attaaaattgcagctctctagagctgatgaaaagagattggtttccttttcatttgaatactgatattctagacgggat
gggtatgccacccttaatccttcttgtgttctgacacaaaggaggaaaagaaatgtatgactcctagagggcatctcct
cctaatggagagggacaaataagaagtatgtttctgaaatattttcaggtcctaattttactagggtacccactaggat
tactggtatctgatctagccccatgattcctccatctttgacatacctgctgtttggtagctcagaatggagcaataca
gtggactctgccccctgagttcactcaaccttccctccacccccacctaggggtattgcacaagggccctaaaagtgg
ccacaggaacagggcaaagaggcttaactgccacacttatagtttgaggaactccaatctccccaaattccagtctgtt
catccttttcttgatctccccagattcactccacattatccttaccaatcttcaattcttctctctctccatgtccagc
caaatttctttttttcagTCACTTACAGGGCTTCCGGTCAAAATTCACTAGGTAGGAGGGTCATCAGCTGGGAAGAACCG
GCGCCTGGGAAACCTGGCTGGATAGGTATG

***Heme Oxygenase 1* (*HMOX1*)**; ENSG00000100292.

cccagaaagagaccttgaagcatctattcattcaggcatgcattcagtattaactgtgacctaccaggcactattccag
gaactgggaatttacaaagcaaagtagtaggcaaggtctccacctttgtgctttacatgccactgttggggctcagaaa
gtgataccaggacaactggtgctttgacatgatgagaggccttaggagctgcctcagagtcaaggtcctgtaagcttg
tcttgttcccccgacaagcacagggagagacgctctctggaatttccttatctgaccaagaacacttctttacaaaaga
aatgcaattgtgttaagtctttttttcctttgagacagagtctcgctcttgtcacccaggctggagtgcagtggcacta
tcttggctcactgcaacctcctcctcctcccgggttcaagcgattctcctgcctcagcctcccaaagtgctgggattac
aggtgtgagccaccgcgcccggccagtgttaagccctttttctagtaatctcatcaaatatccaggaaagatcaaccac
tggagagagaaagagactgggagtcatcaccagacccagacagatttacctgttcttctgaggacagtgccaagagatt
acctgggggactttatctgcctaggacaacctttgtccctgtgcggctccacctccaccttcccttaaagtcggccttt
cacctccagggccattccttcttgctaatgatttactgtctttcaaaagaattgtctgcattccctatctcccttctcc
cctataaaaaaggctgggtagcctctgtactccactgggttacagggtcaccattcttccgtgattaccccattacaat
acatttttgtatgtcttttctcctcttaacctgacttttgtcggttggttttcggggaaccttcagaggaagaagggga

attttttcaaggccccttagctcacccttgggaggatgatcctttccaatacaatctcagaagtgcctctgggctgta
gtggccctaggctgaattccaggaagttttttttttttttttgttttttttttttgagggacagcgtcttgttc
tgttgcccaggttagaatacagtagcgtggtcacagctcactccagcctctacatcccaggctcaagtgaacctccagc
ctcagcctcccaagtagctgggaccacaggcatgtgccaccatgcccagctaatttatttttatattttgtagagacagg
gtctccctatgttgcccaggccagtctcgaactcaaagcaatcttcccacctcgactgggctcaaagcgctcttcccac
ctcaacctcccaaagtactgggactacaggtgtgagctaccatgccaggcctgaaagccatcttaaaaaaaaaatctta
gaatgagaatcacagtattgggaaaggactgtatgaatcatctggtccattcgttttgtcctctgggttcacccagtga
ccctatttcccccgagttctaaggagtccacctcatgcagaattgattcaataggcgatcagcaagggccagctctgct
ctgggccctgagcaggcactgagtataagtcagacctgaatgtgcctggaagagtgtcccacgcattccagcagggaag
cagtttgtatgacaggtgtcccagtccaggcggataccaggtgctgccagagtgtggaggaggcaggcggggacttagt
ctcctccctgggtttggacactggcatcctgctttatgtgtgacaccactgcacccctctgagcctcggtttccccatc
tgtaaaatagaagcgatctaccctcacaggtcagttgtagggatgaaccatgaaaatactagagtctctgtttttgac
aggaactcaaaaaacagatcctaaatgtacatttaaagagggtgtgaggaggcaagcagtcagcagaggattccagcag
gtgacattttagggagctggagacagcagagcctggggttgctaagttcctgatgttgcccaccaggctattgctctga
gcagcgctgcctcccagctttctggaaccttctgggacgcctggggtgcatcaagtcccaaggggacagggagcagaag
gggggctctggaaggagcaaaatcacacccagagcctgcagcttctcagatttccttaaaggtttGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTTTTCTCTAAAAGTCCTATGGCCAGACTTTGT
TTCCCAAGGGTCATATGACTGCTCCTCTCCACCCCACACTGGCCCGGGGCGGGCTGGGCGCGGGCCCCTGCGGGTGTTG
CAACGCCCGGCCAGAAAGTGGGCATCAGCTGTTCCGCCTGGCCCACGTGACCCGCCGAGCATAAATGTGACCGGCCGCG
GCTCCGGCAGTCAACGCCTGCCTCCTCTCGAGCGTCCTCAGCGCAGCCGCCGCCCGCGGAGCCAGCACGAACGAGCCCA
GCACCGGCCGGATG

***Iron-responsive Element-binding Protein 2* (*IREB2*)**; ENSG00000136381.

aagtgctgggattacaggcgtgagccaccgccccggccttatctttttttgagacagggtttcattctgtctcccaggt
tggagtgcagtggcacaatcatagctcactgtaatcttgaatttctgggctcaagtgatcctcctgcctctgtctcacg
gtagctggtactacaggcatgcaccactatgtctggctaactctttaagtttttgtagagacagggtctcaccatgttg
cccaggctagtctcaaactcctagcatcaagcgatcctgcagctttggcttcccaaagtgctgggattacaagcttgag
ctaccaactgtgcctggcccagtcgaatattatatgtcctgtaaaaccaaactcattcgttataaatagatacaagcat
ctgactactcaattttgtcatctaatatagacatgcctgtgcagctacgtattataatacatattattttctaataaat
tactttcctatttttctcctttatattacagttaggtcattaagtagattatttatcgtgtatgtagattatattacct
atgaatttcatgctgggattgtaaaggagatgttacaaaatatagcaaggtgaataaaggaaatatattggagtttggg
ttgacagggcttggatcaatgattaaagggagtggatgaaggagtcaggtataatgtgttgatttctggcttgggcaac
caggtggatggtggcgccattcgctaatacaggcaatacaaaaggaatagtaggtttaaggagggaaatgacgagttca
gttttggagatttttagtttgaggtgccttgggatgtggaggcatcccatgggtactaggatatacacctccgggttta
ggagataggctcaggtcacaggcagactgggaagcatcaacatattaactaataactgaagtcaaaggaagtgtgagat
gggccagggaaaatggacagagggaggaggtacaaggccagagccctgggacacatttaatgtccaatatatttttttt
tgagacagggtcttgctgtgttgctcaggctggagtgcagtggtgcaatcatggctcaccacagcctcaatctcctggg
cccaagtgatcctcccacctcagcctcctgagtagctgagactacaggtgcatgccaccacactcagctaatttttaaa
ttttttttgtagagatagggtctcactacgttgctgtgtttggtcttgaacttttcagctcaagccatcctcccacttt
ggcctcctaaagtgctggaattacaggcatgagctactgtgcccatcaagtgcccagtacatttaaaagctgggcggag
gaagaggtctgcagaggagcctgagatcaagcacccttaggagaagaaacaaaagtagagtctgaagtcacagaagctg
atggaagagtatttttcaacaatgggagagaaccagtcaagattgagatatgtgagattagaaagtggagttatagctg
gccttggctaaaggcatttctgggagtggaagccagactggaagaggataagaagtgggtgggggagagccgagtatga
gcgactctttctggtaagttgctgcgacaggagataatgctgggtaggtggagaacacagggaccggaggagctgaggc
ggaaggatggagaggcttaagcatgtgttaactgccactgggaagaagtcggcaatcacggaagctaccggagggtaag
aaaaaattcattcattcaacagttgagcgcctcccacgcgcccgatcagtatggccgcccccacttgaaaacacgcgtg
tgggccgcccacgtctgacaagttaatgcaaggctttatagttaggggaaagcggttctccttgagctctttctcctag
cagttccgattccgaccctgactccaacagacaccttgcgggaacgcaaacaccgctcgaattcatgaccccaatagaa
aacttaaggccgcaactaagtaacggatcgctgcgaaggccaaactagccacgccaacgcccccactggagcctcccca
gcgctccgcccccgctcgcgagaacagcggcgacggcgcgagaaatcgctttctggttagctccgcccccttccctttct
ttgttttcctgtccgacgatctcgcgggagttaggcgacaaatcccgcgagcgcagacccggggctgGCTCTGCTGCTC
TCGCGATATTTGCGCGAGCCTGCTTCCTTCTTTCCTCCCTTGCCAGTCCGCCTGTCTTCCTCCCCGTCTTCCCTGCCCG
GCCTCCCCCTTCTTCCCCCGCTGGCCCCCTCCCCGGAGGGATAATATGGTCTCCGGCGATG

***Lactotransferrin* (*LTF*)**; ENSG00000012223.

agggagcctctgtgatagaagtgcattcctgtgactgtccctcccccatagtgctgtgagctcctgagggcaaggacag
acccctcttctctgtctctcatggcatctaccagttctggcatgaatgaatgaatgaatgaatgaatgaatgctgttaa
tagatcagtgaaacttcattgttttttggggcagactcatcccctacatcagcctcctctgcaaatggccttgaggggc
tgctctgcctgtgtccagatgctcacatccctgccctggggctgggctgttccaacaggcacagcaggaaaacagtctg
ccctgttgcctacccatgctgccttcagggcactccttaagctgagcctcttggtagtggccccaggttctctgtgttc
ttgccaacactgtaacatacttaagagggcccagggccctctgacctcccagtcatccttttttaaaatttggactcag
aaaaagagacacgggttatgatgtttcttaattcttttataatgatgaaaaggcaaagtcttgttgccaatttaggtac
aaagatgcttcagcactcctgggaattagtcaattttgtatttcttcagtatttttgaaagaacttattgcaattattg
atgatggcaactttaaatggtgcaatatcatgtttccaaacaatgagagaccttggatctgtcaccccaaaacccagct
ggtgattctagcaccaaatctcagaccccagtctcatggcaggcaagcatgatccctgattagcctccaccctgtgcct
tggcaggatcatgcccagaaatggaggggctccccagcctccttaatggcctctccaccttgggctgagcctgcttctc
catggctaggtgcccactgcatgctcactcttggcagagctggctccctggggcattgcctatctgctccaggaaatgc
tttttagtaccaagtagtctaagcagagtcaagaccagcttttcagaataagcagatttcagagtcacatattggtcag
actcacctgccaagaaaggtgttcaggtgacctattatttccctcactgttagtacctttttttcttcacttagttaccc
ttttttttttttggtgggggttacagagtcttactctattgcccaggctgcagtgcagtgacatgatcatggctcactgc
caccttcatctcccaggctcaaatggtcctcccactttagcctcccaagtagctgggaccataggcatacaccaccatg
ctgggctaatttttgtatttttgtagagatgggggtttccctatgttgcccaggctagtcttgaactcctgggctcaa
gcgatcctcccatcttggcctcccaaagtgctgggattacaggcatgagccactgtgccctgcctagttactcttgggc
taagttcacatccatacacacaggatattctttctgaggcccccaatgtgtcccacaggcaccatgctgtatgtgacac
tcccctagagatggatgtttagtttgcttccaactgattaatggcatgcagtggtgcctggaaacatttgtacctgggg
tgctgtgtgtcatgggaatgtatttacgagatgtattcttagaagcagtattctagcttttgaattttaaaatctgaca
tttatggcgattgttaaaatgaggttaccatttcctactgaatactatcaacaccaaaaaagaagaaggaggagatgga
gaaaaaaaagacaaaaaaaaaaaaaaagtggtagggcatcttagccatagggcatctttctcattggcaaataagaacat
ggaaccagccttgggtggtggccattcccctctgaggtccctgtctgttttctgggagctgtattgtgggtctcagcag
ggcagggagataccccatgggcagcttgcctgagactctgggcagcctctcttttctctgtcagctgtccctaggctgc
tgctgggggtggtcgggtcatcttttcaactctcagctcactgctgagccaaggtgaaagcaaatccacctgccctaac
tggctcctaggcaccttcaaggtcatctgctgaagaagatagcagtctcacaggtcaaggcgatcttcaagtaaagacc
ctctgctctgtgtcctgccctctagaaggcactgagaccagagctgggacagggctcaggggggctgcGACTCCTAGGGG
CTTGCAGACTAGTGGGAGAGAAAGAACATCGCAGCAGCCAGGCAGAACCAGGACAGGTGAGGTGCAGGCTGGCTTTCCT
CTCGCAGCGCGGTGTGGAGTCCTGTCCTGCCTCAGGGCTTTTCGGAGCCTGGATCCTCAAGGAACAAGTAGACCTGGCC
GCGGGGAGTGGGGAGGGAAGGGGTGTCTATTGGGCAACAGGGCGGGGCAAAGCCCTGAATAAAGGGGCGCAGGGCAGGC
GCAAGTGGCAGAGCCTTCGTTTGCCAAGTCGCCTCCAGACCGCAGACATGAAACTTGTCTTCCTCGTCCTGCTGTTCCT
CGGGGCCCTCG

***Solute Carrier Family 11* (*proton-coupled divalent metal ion transporters*) (*SLC11A2*)**; ENSG00000110911.

ggagttttctgttcttctgtgactaccctatgttattcttgtctcaatttttaatggaagggaattgtttaaggtgta
caagatggaggtcccaatatggtgtggtggttgagtcttggctcgtgggtctaacagatttgtgttcaaagcctggctt
tgccactaactagctgtgtccttaggcaaattatactccaagtctcagattcctaatttgcaaaatgagaataataata
aataatacttcctgccccattagtttattatgaggattgttttttggtagagtcatggtctcgctatgttgcccaggc
tgttttcaaactcctggcctcaagtgatctgcctgcctcggcctcccaaagtgctgggattgcaggcgtgaatttttgt
gccagcctactacaaggattaatgcttgtgaagtgttaaatatagcgtttggtacacaataaaaattcaattaaggttg
tcagttactcttaagtataaattggggaccaagaaggcaaaaatggagtgtgtttttgcagatggacggaggggagcc
aggacacagctcagcttgctcatgctcagagagttaagctgctgaccctgaaggcaggggcacacaactgtatgtggg
agccagactaagcagtcgagacagggcagacagtgtgagaaagctgttgatgagagctgctgctgaataaaatcatctt
tcacctgtctacagctccccaagtgttctttctgctcatccacccactacctccggacctcaacatgacctttggcgta
gtctaaacctgacactcataacagttacagatttctttgtcccttctacactcccacttttttgagacagggtcttgct
ctgtcgcagaggctggagtgctctggtgtgatcacgacgcactgcagccttgacctcctagcctcaaatgatcctccca
cctcagcctcccaagaagctgggatgttccacaggcacgttccactacactcaggtaattttaaaatttttgtagaga
gggggtctccttatgttgcccagggtaagaattcctaaactcaaataggctttttcctgccaagactttagttcgtggg
tgattttacaggtgcttctattttcccttcaagtggaacaaattggcccccagtccaatctgcagagacagctttgcct
tggggaagttacctgatgtagaacaaggacaggagacaggtcaagttgtagttatttgactttcgtgattaagagagcc
ctgaagcaaagagaccttttcatctggatccctgggcaggagatccataattctttctcctgactcaatggttgttctag
gtaggaaccagaaccacttgctggaagcctgcattgacttatttgaatcgattactgcttttctttcttttccattcta
agctattatgtttttatggtactaagcagatttcttgaaaacttctagaaaactgcctgctacctgcaagaaactgtt
ttcttctctcctgtattttacgccaccccaatgcttctcctatgcagctttgtcgtctaagttatttccctgaggcaaa
atttccccattctttcagctcccctcctcccagattgagaaattctaggcttgccctggacttctgtcttgtagtctct
gggtttttttccttttctttttctttttccatcaggcaccaagtccggagtttgcgtggctctgatgtctgctcgtggatc
actgattgaccaggctgaagccagagggtcctttagatcccgagggataatagagcaggcagcatttccacccacgttc
tgagactcctcataattgcctcagtgctgcacgtacctatatggaaaatacaagtgcactgtcaggcttggggttaatg

gtttcacctcacgtctcagaggagcagagtcctagtgcacctgtggagcctgaaccagctgcactccaagaggagcaga
gaagggaagcaatttcttactccacCCCTTTCAACCACTGGCACTCACAGGCTTGCCACCTACTCTTGCTCCTGCCAAA
TGgtacagagatgaaatcttagaataaaatttttgaactcaataaattgttaattaatatctaggagtggcagatatat
aatgcttctcttttcacgacagGTCAAAGATGAAATCTGATTAGGCTTTGGAGGGCGTGGTTTGCCTCTTGAAATTCCT
TGCTCGTCTTTAGAAAGCTGTGTTGCTTAGAGCACCCAAAAGTGCCCTTTATCTAACTGAGGGAGGCGCACTTGTCTTC
ATATATAGAGGCAGGAGCTGGCATTGGGAAAGTCAAACTAGTTCTGCACC**ATG**AGGAAGAAGCAGCTGAAGACGGAGGC
AGCTCCACACTGTGAACTAA

***Solute Carrier Family 40 A Member 1* (*SLC40A1*)**; ENSG00000138449.

ttattctattttcattttcaatttctttccatttcataagccccaactaacatatctcaactattctcattaattaggat
ttaagattccccttccacagctcactcactggttttaatgtcatgaaccttgcttgttttctttccaaagtatctttcct
acttgcttcactgcagtatcccatttgttctttataacattctctcggttttggcttctggggactccttggtgacaatg
gagaacagagactccatatcacattttgaggtatatgctgacttagccactctttccacaaatacttcttaaacagtgtc
tacgggaagtaatttttttttaaccaaccgaacttttaaaaagccacctggaagttggggacagcatgacacccccttct
tataatatgtaataattaatgtaaagcacttggaggtccttgttgctgagacacactctgacaaatgtgcaaactatcac
tatcctttgtgccttgttgaggacctgtcaaaaaattcatctctaataagggatcgccataaagtggaggggtttattta
ccagttcactccttcactttagttcactcatctgagtatactttggtttattctcaaggatgcatgtataaatgacccta
ttctccacggttagcaaaccaaaacctctctaaattgggagaaagaacttgaatgtcgtgtgaaagaaaggaaaaaagag
ggcttgaaaaggcaaaaagcagtgaaaagcagtagttaaactgtaacccctatctctactcaccagaccacatcccaacc
gaatccagctgcacccaccccgtagacctttggggctcctgattgagagtgcagatacagggcacatactcatgcgtggc
tcccttcactactgggtgtcagtctggcctgtgctcaaggtgtggcatctggttggagtttcaatatgtaggatccacta
ccagggttttcgtgagattaagaagggtaaggtaacctactggcaaaaggggctggtatgccccaggggttggtttggca
cagcaggattaaaacgaagtcaaccaaggctagagtctggtgttctttagtcattcacctcaccctccaggaggccaccg
aatggctttatctggacagggacagatccagggacacaactgggataaccggtattcctggtaagcctgtcactggggca
tttggaggttaggtggggaagggacgcgcgcgtgggggcggggtgagagggtacagagggagaaggaatgatggtgaagg
gtttgctggggctgcagcatcctcattctgtctccaggacggatttggaggccccagtttggggatacgggtaggtctgt
aactcgctgcgggacttcacctttgcaagcctccgtttgctcctctcaagagggatggactctgatctttgcgccccctt
cctgccctttgattctggttctttgagggaagccctgctatgcagtccggggaaggaaaggcattctctgctgcaggcgg
gccggaatgggacggccaggaaagcggcctctgtggcatgaattatatttatttagatacctgtattaaaaattattttc
gttaaaaaaggaatccccacccaccaagctcgcgctggagctttgcactgcgaccgtccggcggccgccctttccctgaa
ctgcggggtagccggtgcgcagcgctttcttccagcacctgacgcttagtttcgcgcagaatctccctacgccgccccgc
cggctccacgcgccttcctccttttcccagccccacggccgccccccgaggttgccctcgcggcttcccggagagcagga
aaacccggggagtggaacgcgtcgaggcgaaggtccccgcaagccgcgcagggtgtctgcggccggttggacgcttgcgc
ccggggtgggcgactcctccgggcaagggcgcggggacggcccggcgcgcaaggttgacgggagctcgtctcgcgccgcg
GGGACGCCCGGGCGGCCCTGAAGGGGACGGGGCGGCCCCAGTCGGAGGTCGCAGGGAGCTCCGCCCCCGACTCGGTATAA
GAGCTGGGCCCGGCCCACGGCGGCGGCGGCGGCGGCGGAGAGAGCTGGCTCAGGGCGTCCGCTAGGCTCGGACGACCTGC
TGAGCCTCCCAAACCGCTTCCATAAGGCTTTGCCTTTCCAACTTCAGCTACAGTGTTAGCTAAGTTTGGAAAGAAGGAAA
AAAGAAAATCCCTGGGCCCCTTTTCTTTTGTTCTTTGCCAAAGTCGTCGTTGTAGTCTTTTTGCCCAAGGCTGTTGTGTT
TTTAGAGGTGCTATCTCCAGTTCCTTGCACTCCTGTTAACAAGCACCTCAGCGAGAGCAGCAGCAGCGATAGCAGCCGCA
GAAGAGCCAGCGGGGTCGCCTAGTGTC**ATG**

***STEAP Family Member 3* (*STEAP3*)**; ENSG00000076351.

gaacaagcaccaaaggtcaacggcgtgggcggtggcctctctgccaaagtgtgtgggcccaagtagggtattcacctcc
ctccctgtctgtgagaaatttcagacccctctacttttccagcaggagttcatgttccttttaagataatcctctata
gacctctggcagcccagccttcaaaatgtctccccctgtctctcttactcatggagagccccgcttccccgcctcaaaa
aaaagtcttcctctctcttactcatgcatggagagccccacaaaaagacactggttcccagggattcttccaaagatgc
tggaagacctctcagtgtttgcatttcccaagctcggtttccttgtgctgtccttgggcattaaaggaaggctattcag
ccatccccatcagcaggagcaggcatcaaaagaacagctactgcaggggagcatattaggggcaccagaggggctag
gaactgggtcttgtcttctcaaagtttacaacccactgcaggggccactcaccagccctggcaccctaagggcaggaac
ctccattgggcacgtgtgtctccagcactgggagcactgtctggcacctggggcactgtgagtgttgaatggataaagg
acctgtaagaattgccataagaggtgttgccctggggcaaggtgggacaggaaagttcttgcaagagaagagaggctct
ggccttgggccctggctgagcagagagctgggaggtgggtattcgggtaggaggaaagctgtggaggggaaagcaaac
tgcctgcgggggtgggggggttctcggcgggagttgggggtggagggaaaggcgctttacagcagcactccctccacag
ggagaccagagctctgcctcttcctccccaccgtgccctttggggtccagttgtggggccttccttgaggcccagctcc
tgcctctggcctccggaggcttatctccccaacagaggctggctccccagaaaacaggattcttagacttcctaaaaat
aacctgaagctccttatcagaggctaatagtgtatgtctcccaggaatactctagaaagagtaccaggggtcaggaacc

cccactcagcctccacgaatctgagcggaaaagcccgggtctcctcacctgtaaagaaaataactgcagccggcaaatg
tgaccacaggtgagaaggggcttcggaccagggcgcacacctctagagaggtgcccacagacagcctccccttacttac
gccctgcttcgctgcagtgctgtgaagccatcggcattggaagtgcggagggcagcgggcagtgccccaggggaaggg
gggacgggtagggtttcagagcccagaaggcctagacagggaggcagctccgggagaaaggtacccgccatcccaccgg
tcttggaacctgaggactgctgggaagggtcaaaagtctgccctaagggtggacaagggacagagattgggggggacc
acaaaggcctgggcgcaaccttgttgagagggaagggatccggacccaatctggagccctggtctcttcccacctgcaa
ccgtggctgccacttggcctcggttataaaatggacaaggattcgaggtgctcctcaggttaagagaggggggcctaga
ggtcaccctggcctcggtgccctccgtggccgggttaagaggaggtcccggagttctgctcacttcagccgtgtgccgg
gcactgcaaatcaggaagtgttggcgccggctggcgacctcccgcctggggccaggggaggagggtggttggacgctgc
caccgctgccggggctgtgcagggctgggcggggagcgaggacccggcggctcctgattgcggccccgggggaggtggc
cgagccggataagctgcggcgggctggagggcggccacctcccctgcaggtccggccctcccgggcgggtggggcgcgg
gggaggaggagcctcgggccgagccACCGCCTTCGCCGCGGACCTTCAGCTGCCGCGGTCGCTCCGAGCGGCGGGCCGC
AGAGgtgagtgtaccctcccccggtctccgctcacgtgtgtgtgcgcgcgttagATGACATTTATTCATTTTATGCATC
CTGGGTTCTACTGGTCGTCCCACCTCAGTTCCTGTAGCAAAGAGACTTGAGTCTGAGCCACTAATTATCACCCGTGAGG
TTTCCTCCCCGAGCAGGAAGCAGCAGGCCAGAGCTGCGCTCTCTCAGTGCACTCTCCAACCAAGCATCAGTCACCACTC
CCGGTCCAGCCCCTGTGGCCAAGAGCTGGCGTGCAGGCTGCGGGAGGCAGCTGGCTGTGCAAGACCCTGGCAGGGCCCT
CGCCTCCTGAGAAACCGAGAGTCAGAACCAAAGCCAGGCTGTCCTGGTTGGAGACTGAGCCAGAAAGGGTGGCTCACCT
CACGGTGAGGCTGTCGAGTGACCTGAGAGCCTCAGACCCTCACGTCAGCCGG**ATG**TCGCACCAGCCTGCTGTTG

***Transferrin* (*TF*)**; ENSG00000091513.

tgtcctcttgagcaaaagccaaggccctggcctctccatctaggtgtgtgtagatgacgtgggagcaccagttctctat
gtcttgtctgagcttccttcatatcccatgaggctgtatatttccttcccagcacctgtcttgggcttggactcaaagg
acccttctgaggacagggcatagagctggctgctctggagaggactcagcctggccagtaggtagttgctccctgctga
ggggctctggctaggctatgggcctgtggctggaggactggcctcagcggggcttagtctacgaatgtagatgggctcc
agaaaaggagaatcactgaccatcttctttccatgccctgcgttgagcttccctgagtttgtggcatgggctcaacctt
tcgtaactgcacttcctgctctctggcagaccctgcaagcttactacagggcctgtccatgcatcagctgtatgtgtgc
atgctgctctgctgggtgccttccagttagccttgcccctgacttttcctgctcatctcacatccttacctcaaggtct
ccaaatgaatccagaccactgcctagagcctgctgcaggggcagtaggaatatggccctcatgctaacagtcttctttc
cctatggctcaagcacaaggggcccgtggacgggcaaggggggagtaagacagggcagcagtggtcactctccctcgct
ggtgaggctgatgtgttgggaaagccagacttgacttctgtcctcactgtaaacagaaagcttctgtagtggactctaa
agtgggaggtatgtaacccctttgcatatgagataatcttttggagtgtgggagaacataatctaacttccagttatat
ttattgttattgcaaaagattaacaaattatcattcattcatattgaatgtgtgaattggccctatctcccacctgagt
ctgcagtcatgacagatggagcctttcaggtgtcctaagggaagatggggttcacctcacaatgctaaaggtgaaaagg
ggagtgccgaccctgccctttgttttggcatttttatttccatactttgtggtttgtatgtgcccagttaaacggatgt
acagattagataatctggctttaactccaaccttcacgagatgcacacatgccttaaaaagattcttacaaagaaacta
gaattaggagacaattctgacaacacaaaaacaagtgaacaaaaatcaaatgactgacctgacgtgtccactaggcttg
atatgagcttggtcaactaaaacatgagacagaatgaatgaccatctggtcaggttggataagaagtaacttttggcaa
aatctctgaagtctggatttacacttcgtggaaggaatcttgacctttccccaagtatattatgccttgagataataat
cctgtataaggccgcataagcaaaacgacatttgaaaatgagtttcatctttagctcatccttttacaaatttctattc
cttgtgcatgattgaagtgtacacatttttttagttaaaatataagcatatagaaggcacaagttcaatatttttcatga
taaggatgtgtaattaaactggttaggagactcatcgactagaggggacatggtggccccaggctgtaagaacaggcc
acaccgtccactgggccgcttgctttgtgctaagatgacactttgttctgagcctcacagtgtcttgaccatgttcctg
gaaccttcttgttggagggagttcatcttcccctatgactctgtccctagtctaaggtgtcccacaggaagcttgaggg
cgggaagttttccagcccaggagcctgagctcagcggggcaggaagagggagcagctcctccgtgggggacctttgaga
gcccaggagcaggatttcgagggacacctggtggggagcaaaaggtgctgagtctgtctttgaccttgagcccagcttg
tttctcctgcatcctcccccaaaagGGGCTTTGCCTGTCATTCTGCAGTTCTAGTGTGGGGTCTGGGCGCAGTTCTTTT
CCCTCTCCAGCCTCGGAGTCTTCCTCTGTGGACTGCGCAGATAGGACTGGTGGCACGGACCAGCTCTGCAGCCCTGGAG
TCAGGAGCAGAGCCCCCCGGCTCCCAGCCCGCCGTAGCCGCTCCTGGCACCGAGCGAGCCGCGATGACAATGGCTGCAT
TGTGCTTCATGTCCCTTCCCATCAACATTTCTGTGCTGGACTCCTTCCACTCGCGGGTCGTCTCCAGAGCTCAGAAAAT
GAGGTGATCAGTGGGACGAGTAAGGAAGGGGGGTTGGGAGAGGGGCGATTGGGCAACCCGGCTGCACAAACACGGGAGG
TCAAAGATTGCGCCCAGCCCGCCCAGGCCGGGAATGGAATAAAGGGACGCGGGGCGCCGGAGGCTGCACAGAAGCGAGT
CCGACTGTGCTCGCTGCTCAGCGCCGCACCCGGAAG**ATG**

***Transferrin Receptor 1* (*TFRC*)**; ENSG00000072274.

gacagagtcttgctctgctgcccaggctgaagggcagtggtgtgatctcagctcactgcagccacccccctcccgggtt
caaacgattctcctgcctcagcctctggagtagctgggactacaggtgtgcgccaccacgccagactaaattttgtatt
tttagttgtgatggggttttatcatgctggccaggctggtcttgaactcctgacctcagaagatccacctgcctcagtc
tcccaaagtgttgggattacagggattagccatcgcatctggccacctaaacctatgtcctccatgaggctataagatt
ttcacctgcaaataccagcattgtttgtttctttttaaaagatacttcagtcaaggaaaggcagaaagccaggcatgtt
agctcatgcctataatcccagcactttgggagaccggaggcaggaggattgtttgagcccaggagttcaagactagtct

gggctacatagcaaaccctgcgtcaaattaaaaataaaaaataaataggctgggcatggtggctcatgcctgttatccc
aacattttaggaggccaaggtgggcagatcacctgaggtcaggaattcgagacaagcctggccaacatgataaaatccc
gtcactcctcaaaatacaaaaattagcctggcatggtggcacacacctgtagtcccagctacttgggaggctgaggcag
gagaatcgcttgaatccggaggcagaggttgcagtgagcccgagatcatgccactgtactccagccagggtgacagagc
aagacactgtcttaaaagtaaaaataataaaaataaataaaattaataattaaagaaataaagaaggaacaccacaggg
gagcaggcctggtgtctgaatgttaaggtaggccctctgtggatgtgcatccttaacctttcctgattagatcattaa
aggtgcagcaatgctcctgctcatatgacttttttttttttgagatggaatcttgctctgtcgcccaggttggagtg
caatggcgcgatcttggctcactgcaacctccgcctcttgggttcaagcgattcttctgcctcagcctcctgagtagct
gggattacaggcgcgtgccactatgcctggctaatttttttttgaaatttttttagtagagatggggtttcaccatattg
gtcaggctggtctcgaactccagacctcgtgatccgctcacctcggcctcccaaagtgttaggattgcaggcgtgagcc
accgcgcctggcaactcatatgacttttaaacacttctctttggggagagagattttagagaatgggagtaaagagaaa
ggcggtgagaatcccaagtactttctctatcttagagcagaagaaataaagggaagagtaaaagcccaaggcccagaaa
cggatttgcttcttaaagggctgttctgtgtccctacgggcctgaaggtcagtttatgtgcaatgctttttaaggtatg
ggttgctttcttgcattgtagctactacctccctggcctgctggttactgaccagagggagggagagggtgtgattgtt
tggaaagagtttaaacagtgacaagaattgaattttatttatttgttattcattgcacaatccaataacagcgatggca
atgaaaacgaaagttaatctacaatcctatcaccctattaagtcacttgtttctattttccctttattccctcatgca
ttaaacattacgcaaagcactccgctagtcttggttaaagtgcagcttgtatacattctgaataaaaatttgcattttg
cttttctttcctttttgtatttctgaggctgcaaaatacatcttcacaagagtggctttaatagttacctaacattcc
atcctgcacgtacatttcaaggcactgccgcagtgcaatatccaacatttaagcgtcaacgccaacagattggacctag
cactgcagggactggaggttggtcaaggtcaggcggaggctgccaggctaccagggtggaggaaggcgccgaggcagag
gccagtgcgcccatcgcgcggctcctcggggcacctgctgccttggcgccttttcccttggccttcgcctcgcccgcag
cgccctccgcatagggccccgcccgctgcgcgcgcatccccgccccccgggcgatctgtcagagcacCTCGCGAGCGTA
CGTGCCTCAGGAAGTGACGCACAGCCCCCCTGGGGGCCGGGGGCGGGGCCAGGCTATAAACCGCCGGTTAGGGGCCGCC
ATCCCCTCAGAGCGTCGGGATATCGGGTGGCGGCTCGGGACGGAGGACGCGCTAGTgtgagtgcgggcttctagaacta
caccgaccctcgtgtcctcccttcatcctgcggggctggctggagcggccgctccggtgctgtccagcagccataggga
gccgcacggggagcgggaaagcggtcgcggccccaggcggggcggccgggatggagcggggccgcgagcctgtggggaa
ggggctgtggcggcgcctcgagcggctgcaggtacacggggtcggcggctgtgcgcagaggcgtccctgcgcctctcgt
cccttcgcctctcgtcccttcccttctctctgccttcttgcccgcctcctcggtcacagagcgacgaatgacgagacca
ggtgtacccccactgtcgctctcagccccggggacttcgggtcctcgcccttgaaggccgcaggccctagtgcgccggc
tccgggctgcgggtccgggagcgcgggcgcagccaaggtgcagctgcgcggcgtgcggcgccgggggaacacgtggctg
ctccaggaagtcgccccagggaacggctgggattcgtgggtgaccttgggctcctaaaccttcggttccccgggctccg
ggcgggccccgttctcactgcggaaagggacaaagctgtccccgattcggttactagcgtgttacaggcattaattaga
tacgggtttctgtaaacttacccttcggctctggtacaacctgatcttgccatgctctgcctgttcgccttgtgtgtgg
attctgttttctcaaggcagaatccgactgaacgccgcaggtgttcacattaagaatattgctagcagttctcactaca
agacgggagccataggagtgacttatttgtgaaagatacttggagattagcggtgtggtcagagggctgtgctggattg
atgggcgccggagaggccgattgtgtcacattctgctggacagttcttttaaggttgggagggtgggtaagaaaataca
ttctgattcggctcttttcggataacgctttccctctgcttactgcttgtagagacctcacttacccgggggactggat
tctgcagctttttccatttcttccccgttgataagaggattaaagtaggaaattgtatttggctaccccatgcttata
tgctaagcttactgtgaaaattaaggttaggctgtctagagtacttgtggaccctgctggttcccccgcccccggcc
ctgaattttgggttacaatatgcttctcaagaaaactcagagttaagataatttttgtcatcgattttgcaaggttata
tatagcatagcatccttcaatttcctttggatgttcttatcaagtaatggtggctgtaaacctagcctgtgttgaagat
tgttagatagacagcagtttggctaactccgccctaaggcgaggcagttacttgcctctgataatgtatttaaatgtca
tggagcaagattcccagctaaacctgaatcgatcacaatgctagttaaaaatgaccctcggctgggcgcattggctccc
gcctgtaatcccagcactttgggaggccaaggcaggcggatcacctaaggtcaggagttcgagtccagtctgactaaca
tggtgaaatcctgtctctactaaaaacacaaaaattagccgggcatggtggctcacgtctgtaattccagcactttggg
aggccgaggcaggtggatcacttgtggtgaggagttcgagactagcctggccaacatggtgaaaccccgtctctacgaa
aaatacaaaaattagctgggcatggtggtgcacacctgtaatcccagccactcaggaggctgaggctcgagagtcgctt
taacctgggaggcagaggttgtagtgagctgagatcgtgctactgcactccagcctgcgcgacagcgagatgccatctc
aaaaacaacaacaaaaaaactccctactaaaccagaagagtgatggggacgggaaaagattcctggtccaatgttttca
ttatatttttcatatcatttggaatctcatgcatcaggcatgccccagtactgttaaagacaatatttttactcattat
caggtagttcaataccagttattacaggatagggaagtcagtcagagaaggcttcatagagatgaagccttgagctgag
cctggaagaaatgaggaggacctcaaggaacagagaagaatgttggtgggtaaaaacaggggctggattctgtctgatt
ttggagaaaatagagggtagtgaagttgttaggcagtgaatgtgttccatttcatggttcaaaacgtggggctacgttt
tgcctacctgagcttcattattaatgtgagaaattgaattgttgttttcagtcaccattgagaacaccaaagataacac
aagtagcttagattattatttatttatttatttttgacagagtttcacgcttgttgcccaggctgtagtgcaatggtgt
gatctcggctccctgcaacctccacctcccccaggttcaagtgattctcctgcctcagcctccccagtagctgggatta

caggcatgcgccaccatgcccggctaattttgtatttttttggtagagatgggatttctccatgtttggtcaggctagt
ctcaaactctcaacctcaggtgatctgctcgcctcggcctcccaaagtgttgctgggattacaggcgtgagccaccatg
cccggccatcgctttgtgttcttaaaattaatttaaaacaagaaaacttgaggaatgatggttcagatgagtgttaaaa
cttgcaagatgtttttcctagaaaaggatgattaaaattggttcagggtggggccttccagttctggctctaatgatt
gggtccttactgttctggtggagtggtagtgataagctttttgtaacagaaggggcaaaagattgtgcttctggctggg
cgcagtggctcacgccttcgatttccttgggatgttcttatcaagtaactaatcttagcactttgggaggccaggtggg
tggatcaccctaggtcagaagttccagaccagcctggccaacgtggtaaaatcccgtctctactaaaaatacaaaaatt
agctgggcatggtggtgggcacctgtaatcccagctactcggagtctgaggtgggaaaatggcttgaacccgggagggg
gaggttgcagtgagccgagatcctgccattgcactccagcctaggcaacaagaccgaaactctgtctcaagaaaaaaaa
aaagattatgcttctaagaatatggtctattgcatatggacactggtttttaaagcttgaatttaaaaagaatttttg
taataggttttagagctttttcttttttcttttttattatttttttgagatggagtttcactccgtcacccaggctggag
tgcagtggtgccatctcggctcactgcaacctccacctcccgggttcaagcgattctcctgcctcagcctccccagtag
ctgggattacaggtgcccaccaccacgcccatctaatttttgtatttttagtagagatggggtttcaccaccttggcca
ggctggtctcaaactcctgacctcgtgatccacctgcctcggcctcccaaagtgttgggattacaggcatgagccaccg
tgcctggccagaaagcttttttcttaatgccagttataataccctttgctttacaatttgcgtatacctcacagcttgct
ttgtgccgtgtgtgtttaatatgccctaagtgtactcgtatataggaaaagctttaaaatggtatccaataaatgtcca
ttttgcaaatctcaggataatgtcattttcaagtatctagatgccctcaaaccaaattaactggaaaattcagatttca
gttaaggcctttaacttaacttaggactttgattacagagatgtattttgctttgacaaaagttaaggttaatgtctta
aatttccagagatgataacacaatttctataacccagtaacccatttaacttcgcttagattatcaattttcagccttg
aacaggccacttagtttctctaataactggcctcttcatctgtaaaacagcgtaatctgtgtttcatgtttcttcttgt
gtcaaagatcgaatcacttgtaggaactattcattcagtcaggatcagtatttcttccccttcccattggagatggttt
gtggtatttgtagcataactggtcttgtctgccttagtcttaattttgtctttttgtatgtgtcactttctttttttt
tgagacagagttttgctcttgttttccaggctggagtgcagtggcgcgatcttggctcactgcaagctccacctcctgg
gttcacgccattctcctgcctcaggctcccgagtggctgggattacaggcgcccaccaccacgtccggctaatttttgt
atttttagtagagacagggtttctccatgttggccaggctggtcttgaactcctgacctcaggtgatccacccatctca
gcctcccaaagtgctaggattacaggtgtgagccaccccgcccagcagtatgtgtcacttttgtctactcagaacatag
tggttcgtacttataattgcagcactttgggaggccaaggtgggaggattgtttgaggccaggagttcaagaccagcct
gggcaacatagggtgaccttgtctctacaaaaagaaaaaaaaagaattatgtatataacaaacttaaaaaggatccca
cttaatttagttgtcatgtaaactatggttaaatacctttctagaaaagtgataatgtattttaaaaatttaatgtat
catttagcctggtaatagaaagctcactaatctgatacagtagtatctttctcaataattctctattctgatacctagG
TTCTTCTGTGTGGCAGTTCAGA**ATG**

***Transferrin Receptor 2* (*TFR2*)**; ENSG00000106327.

taatcccagcactttgagaggccaaggcgggaggatcatttgagcccaaaatttctagaccagcctgggcaacaaagcg
agacctcatctctacaaaaaatttaaaaagcaactggcaatgtggcgcatacctgtagtcccacctacttgggaggcca
aggttggaggatcactggaacctggaaggttgaggctgcagtgagctgtgccactgcattccaacctgggtgacagagc
aagatcctgtctcaaaaaaaacccaaaaaaattgggtgtgtattttccacttagagcacgtctcaatttggaccagggc
tggtgagctacaggaagctagtggccactgcattggacagcacatgtgtgtaggctgagaaagatcaaagaagcccagc
cgtggtagctcacgcctgtaatcccaacattatgggaggctgaggtggaacgattacttgagcccaggagtttgagacc
aggctgggcaacatagtgagacaccatctctacaaaaaataaaaattaggctgggcgcggtggctcacgcttgtaatcc
cagcactttgggaggccgaggcgggcggatcatctgaggtcaggagtttgagaccagcttggccaacatggcgaaaccc
tgtctctactaaaaatacaaaaattagccgggcgtgttggcacgcgcctgtaatcccagctacccagcaggctgaagca
ggagaatcgctggaacctgggaggcggaggctgcagtgagccaagatcgcaccactgcattccagcctgggcgacagag
caagactccatctcaaaaaaataaaataaaaataaaaataattaaaattaaaatttaaaaaaatagagagaagggaca
ctctgagatgctcattgcaggaaccctcaaggggcaagttgtaatgccgaaggctgcagatggggtgccacagtgttgg
aatggtgaggtggagggaatatcttccttgggaaagagggagactggcggtgttccaaagccctgatggagtgagaaga
ggacacctggtggggaaggactgtgtgtgtgtaggggtgttgagtgccaggtgcaggttgcctggcctggggctgtgc
gtgccatccccccagctcaggaacagggcctagctcagggcagagactgctgttcaataatacttgttgaatgaatgaa
ttattgattgaatgatatgaagaggatactctccctctgaatatgaaaggggacagaccgaggaggggaggatgaggtg
gggaggggcttcagagggatgtgaacccaaggggtggttggggcctgttggaggattcccaggggctgcacctgcctgt
ccacacctccctccaaggccaccctgaccctgatattacccccctccacccagggcactttccagcagatgtggatctcc
aagcaggaatatgaggagggcgggaagcagtgcgtggagcgaaagtgcccctgatggcactcctccccacacacctgct
cccaagctcagatggaagtcccttaaccccccatgccacattgccccctcctcctttccctcttgtcctcattaatggt
gatgtttctgggttgaaagaagtaaaaatgttttaagaaaaaaaagtttggcctccgtggggtggggagtgggcacagg
ggtctggagggaggtgaggggagggagtctcacttctacacccatgccgatcccacagtgggtgctgggacacagcccc

aggggccatggattcctcaggtttcctaggggttgtgacccagggagggccacatgacgaggaccgggggttgggggag
ttgctgggggctaacagttgtggtgggaggtgggaggggctgcagggtgcagtggaggggcacggcctggggtgagtgt
ggcaggaggtgggtggagcaggaggtgaagcctggatgggcggaacaggaggtggggcctggtgggcggggcctcagcg
taggccttaggtctcaggggagcagACGCCTGAGCACCCCAGGACCTGCGCTCAGgtgagatgcgctatgccttccttc
ctggggctcttggtggtgacagacatggggcggggactagggcgctggtgcatcccagcagctcggaacttctctggac
ctgagtaggagctgatgtttcccaggggaaacccgaaagggaaacttggtgggaccctgagccctcctcttcagagcta
cttgtcttggggccctcggaagggtttggggacctgcgccccctgcctgtttagcagtttatgcctgtgacatgtttct
ggggtgaggctaactggcctttcatcagcctcagaacccaaagaggtagaaacccatgtgatccaagtccttcggtgac
caaacgggccggtgcccagggaggaccatgacttgtccagggtcacccagctctgcccacaccttggcggggactctga
tcactcctccccactgacccacatatccctgaccccaggacccctatctgggagctctcgctggtctcatctgaccatg
tgctcccacctgtgcccctccttctccagcagccagcaggagtctgagaacgacctttcaaaagagacacacatgcaca
ccaccagaccagacccccaagtcacaccactgtggtgcctccagcttagaatgccactggggcaagtgtcaggacaggg
aggacagctgatatgtcacaggaccggggcactgggggactctgaaggccagatgaggaaggggcagatgtcttagggg
tgaatcttccacctgggcatggcggggcttaggaaggggtccccaggatccagggaggggtaaaagcaggagaggggag
ggggcacccatgtctctgccctcccccactatcggcccccttcccccacccctgatgggaactaggaggccaaagtccg
ccccaaggtcaaaaaattgattgttttgcagGGAGGAGGCAGGCGTGGGGCTGTGGAGAGATTGGCAGGGGAGAGCACA
GCCCCTTGTGCTCTGGCCTGGACCCGGGGGCCACGTCTGGAAGGCTGGACTGAGGCCAGGACTGTGCCCCCACCCTTGG
GGGTGGTGAGGAGCAGCCTTGGTTCAGGCTGCCTGCCAGGACTGATAAGGGGCCCTCCTAGGGCTCCCACAAACGGTTT
ATCGGTTTATCGCTGGGGGACAGCCTGCAGGCTTCAGGAGGGGACACAAGC**ATG**

Appendix 3: pGL4 Vector Maps

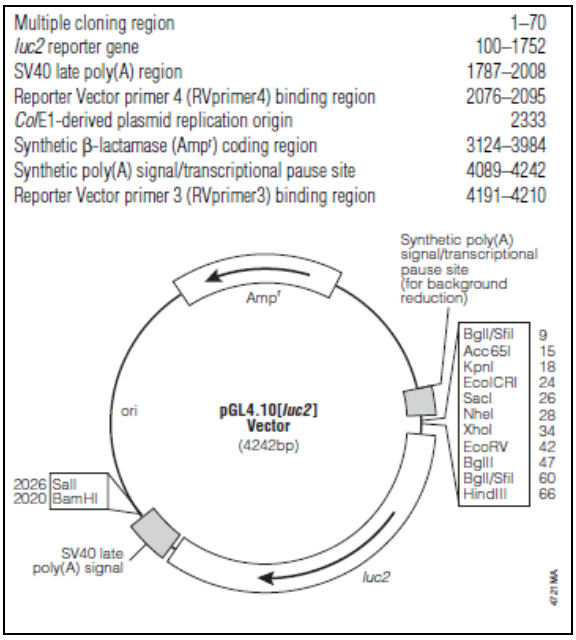


Figure A3.1 pGL4.10[*luc2*] vector map.

Source: http://www.promega.com/products/vectors

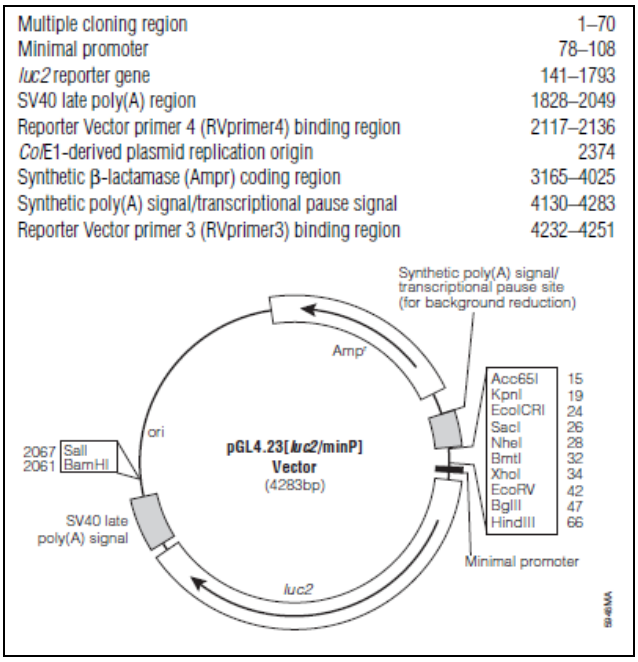


Figure A3.2 pGL4.23[*luc2*/minP] vector map.

Source: http://www.promega.com/products/vectors

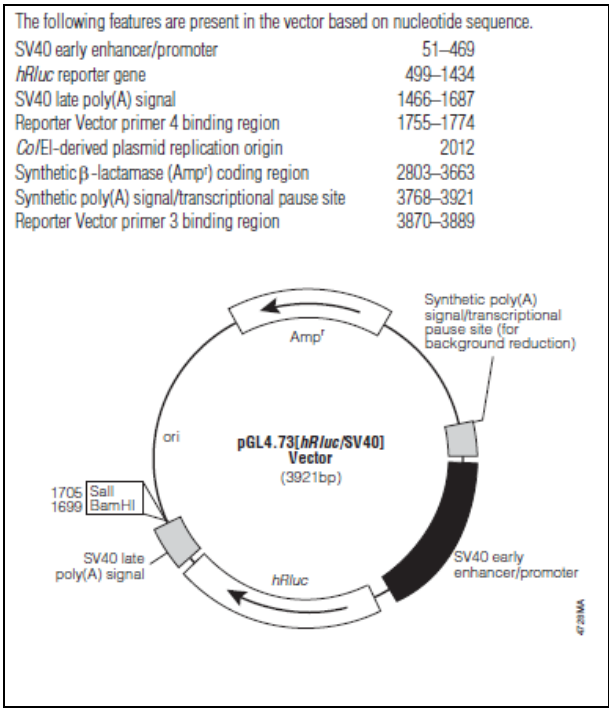


Figure A3.3 pGL4.73[*hRluc*/SV40] vector map.

Source: http://www.promega.com/products/vectors

Appendix 4: *In Silico* Promoter Analyses Supplementary Data.

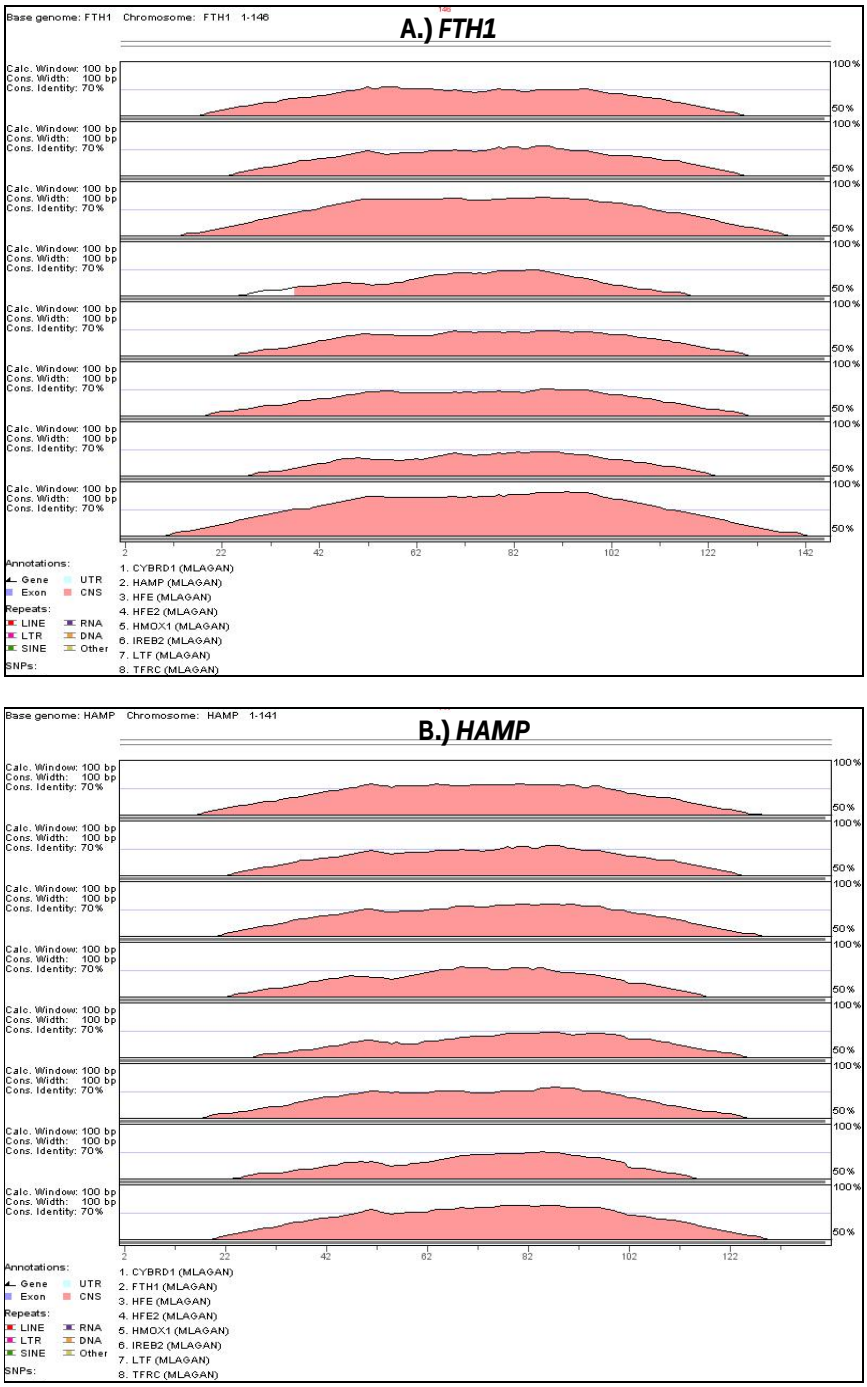


Figure S1 mVISTA sequence alignments of the CR in each of the nine iron metabolism genes.

Pink coloured areas indicate regions of sequence similarity. Each gene in turn was utilized as the base genome for visualization of the alignments: **A.**) *FTH1* and **B.**) *HAMP*.

Abbreviations: *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

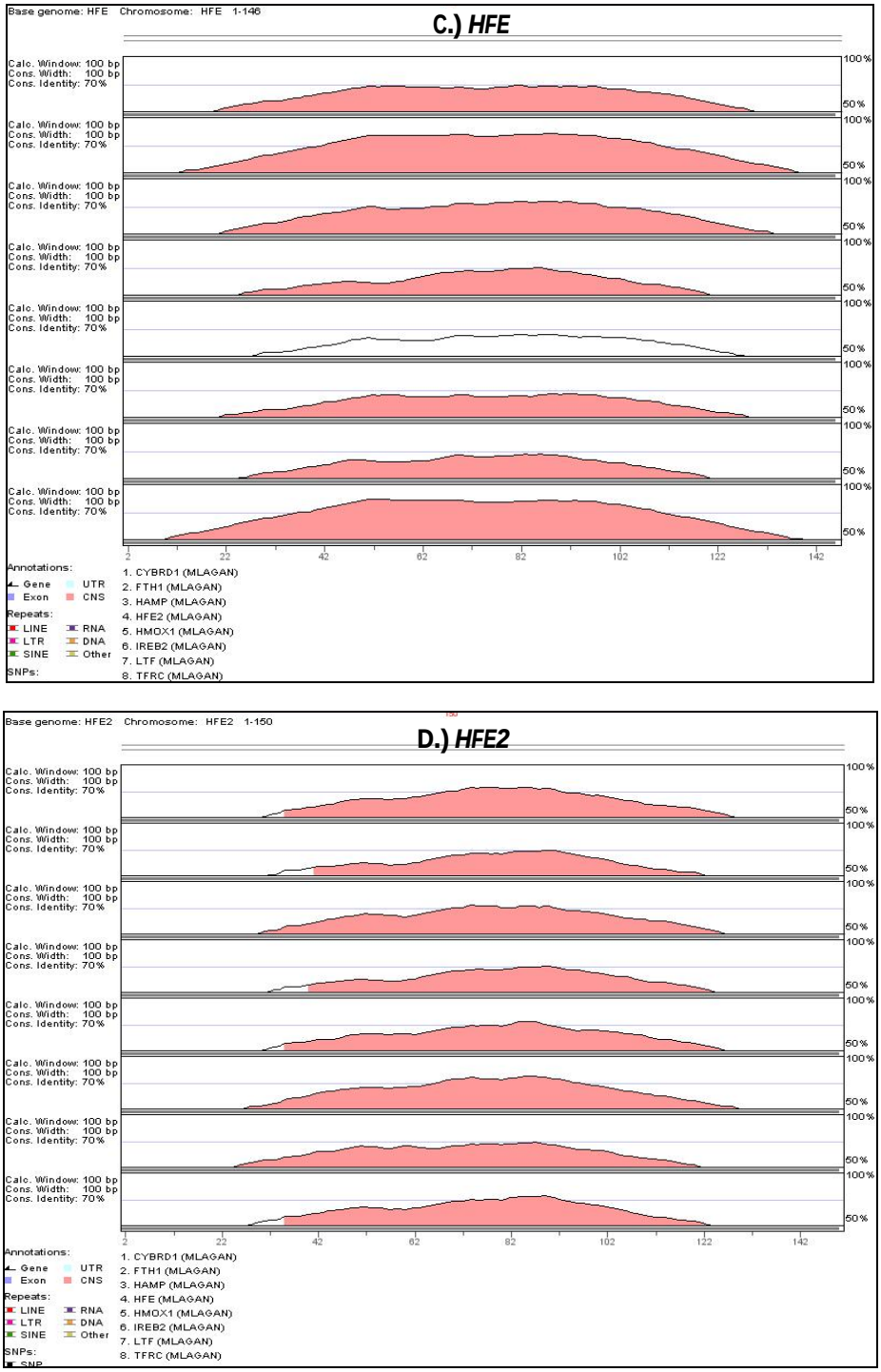


Figure S1 Cont. mVISTA sequence alignments of the CR in each of the nine iron metabolism genes.

Pink coloured areas indicate regions of sequence similarity. Each gene in turn was utilized as the base genome for visualization of the alignments: **C.**) *HFE* and **D.**) *HFE2.*

Abbreviations: *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

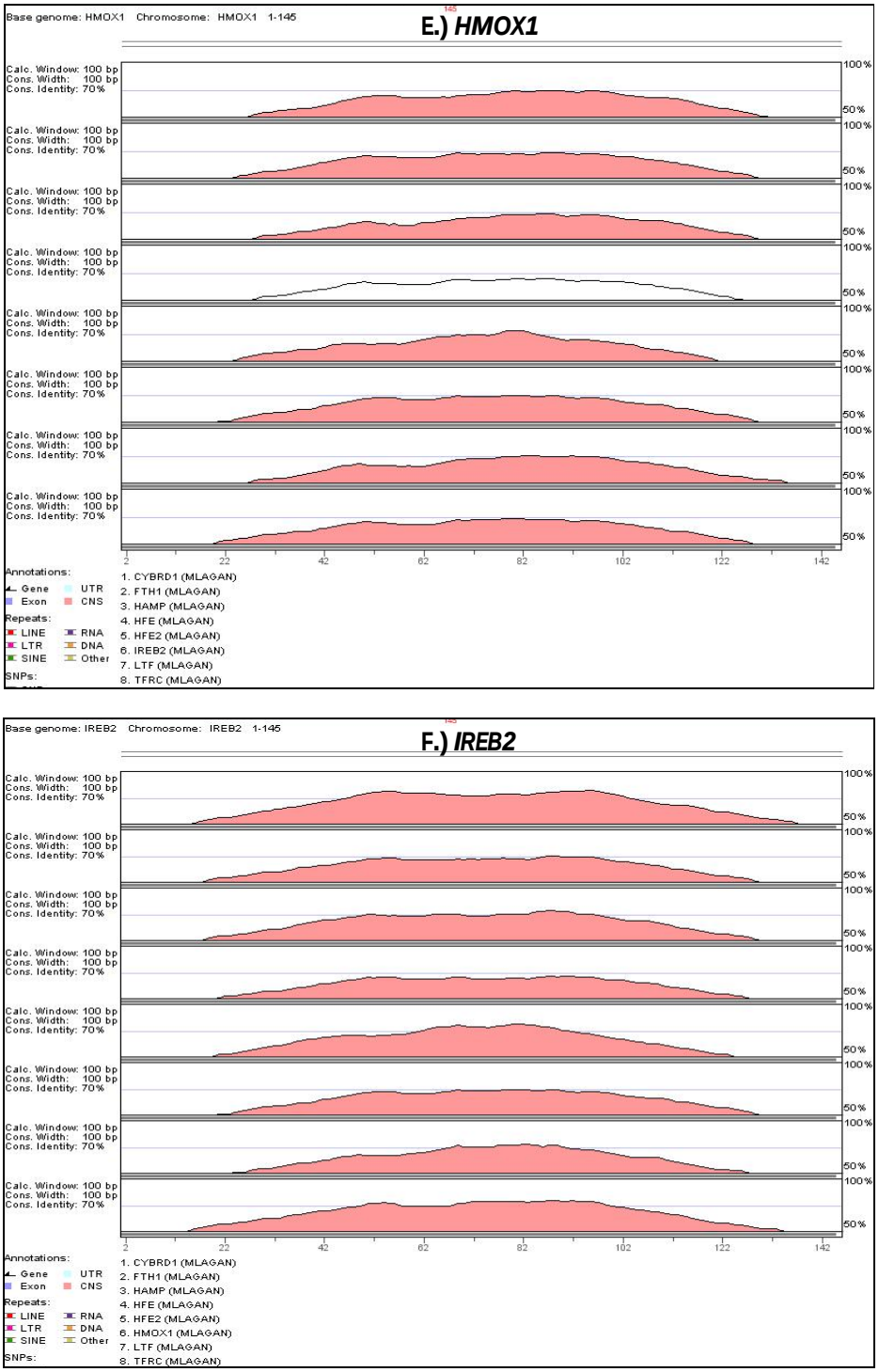


Figure S1 Cont. mVISTA sequence alignments of the CR in each of the nine iron metabolism genes.

Pink coloured areas indicate regions of sequence similarity. Each gene in turn was utilized as the base genome for visualization of the alignments: **E.**) *HMOX1* and **F.**) *IREB2*.

Abbreviations: *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

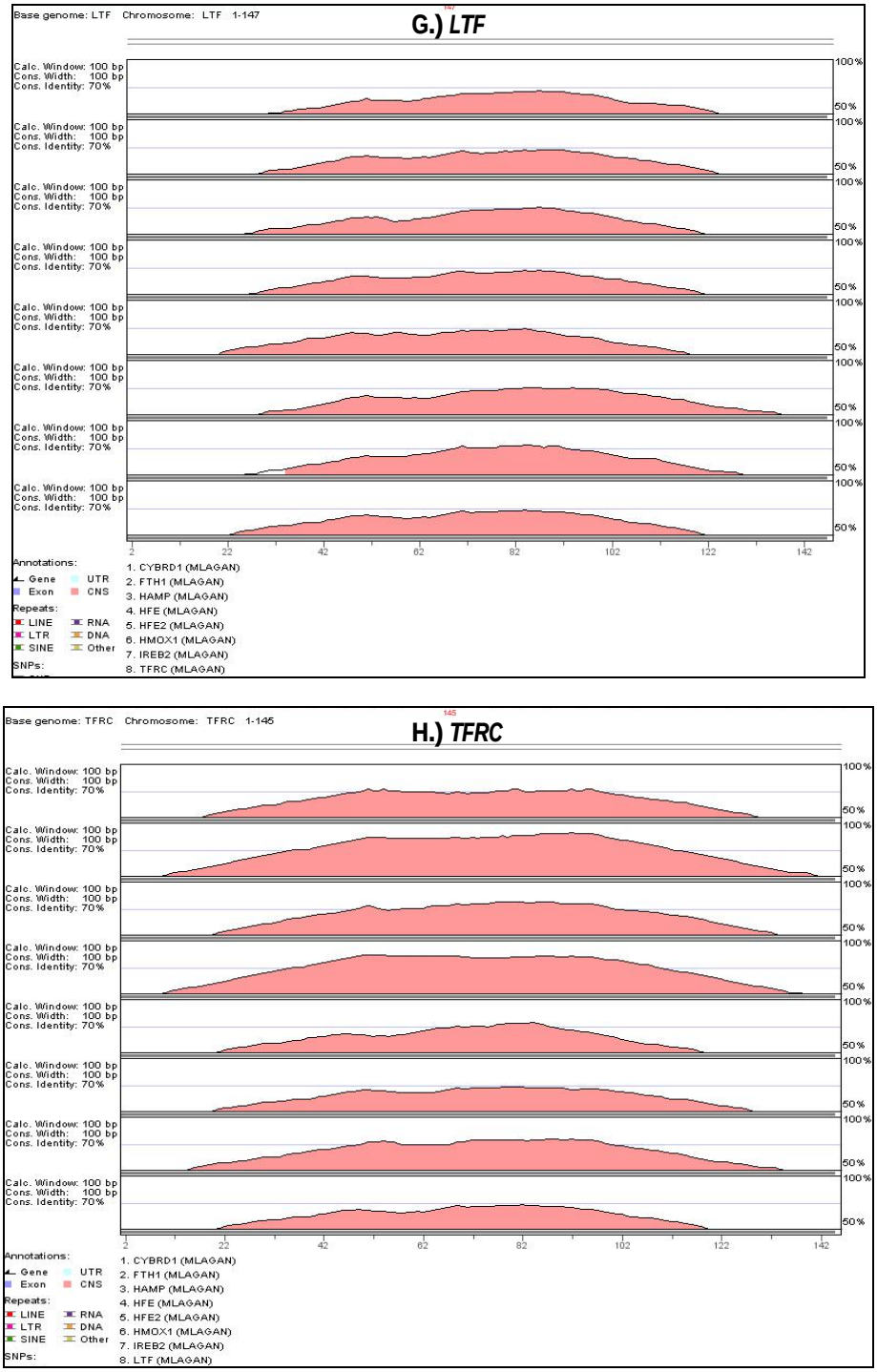


Figure S1 Cont. mVISTA sequence alignments of the CR in each of the nine iron metabolism genes.

Pink coloured areas indicate regions of sequence similarity. Each gene in turn was utilized as the base genome for visualization of the alignments: **G.**) *LTF* and **H.**) *TFRC*.

Abbreviations: *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor protein 1 gene.

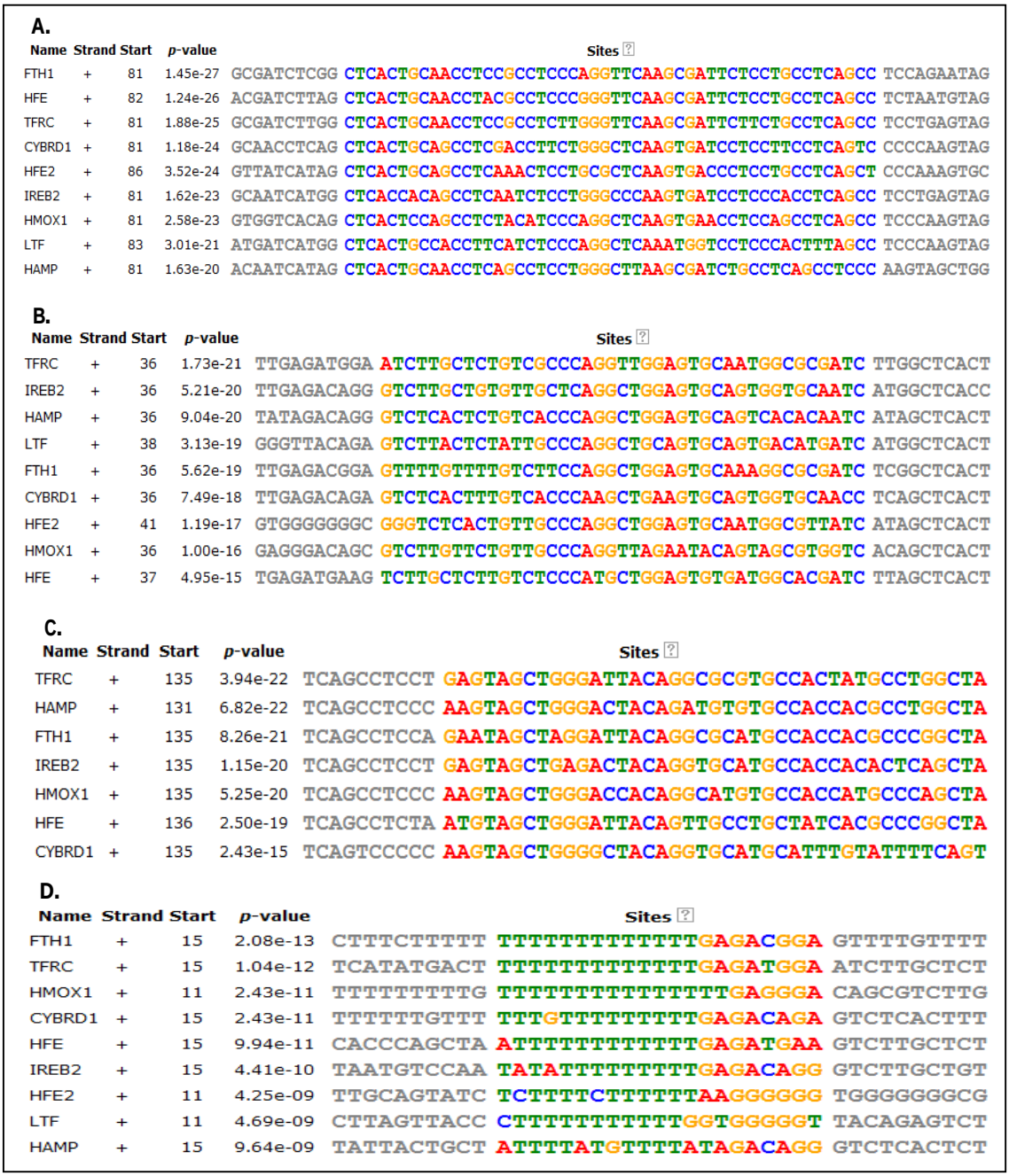


Figure S2 CRM discovery using MEME.

The sequences of the detected motifs in the CR of each gene are shown to demonstrate the level of nucleotide conservation at each site. **A**.) Motif 1; **B**.) Motif 2; **C**.) Motif 3; **D**.) Motif 4. Individual *p*-value scores for the four motifs are listed.

Abbreviations: CR, conserved region; CRM, *cis*-regulatory module; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; *LTF*, Lactotransferrin gene; *TFRC*, Transferrin receptor 1 gene.

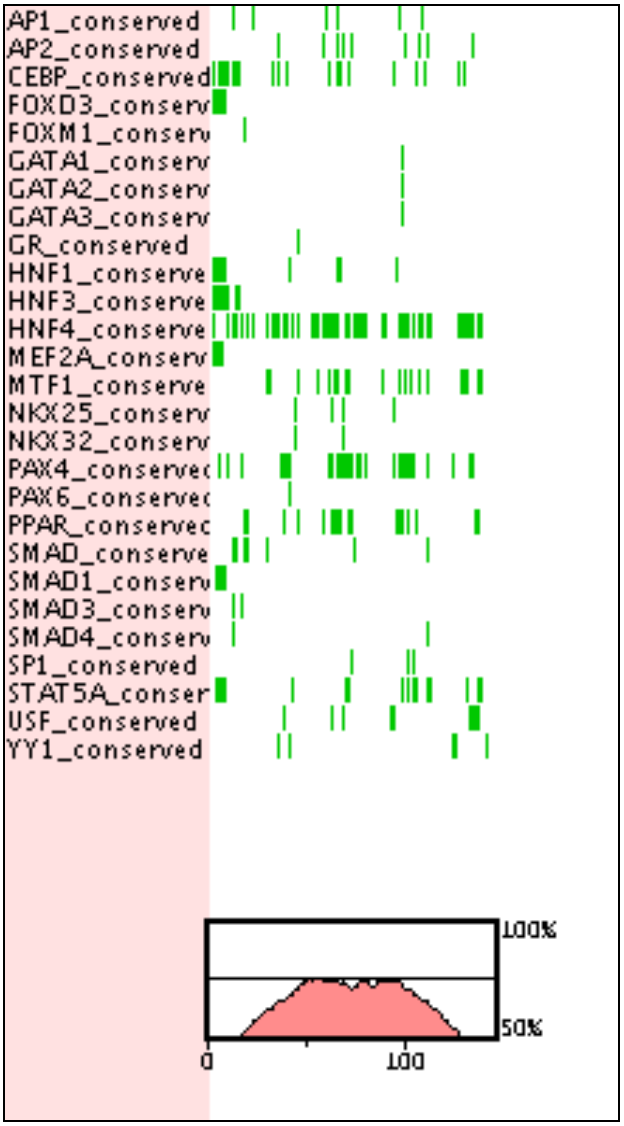


Figure S3 rVISTA TfBS prediction of the CR.

Green bars denote the alignment and conservation of the TfBSs sites indicated on the left of the figure. The pink shaded area at the bottom of the figure indicates the consensus CR sequence submitted to rVISTA. **Abbreviations:** CR, conserved region; CEBP, CCAAT enhancer binding protein; FOXD3, Forkhead box protein D3; FOXM1, Forkhead box protein M1; FOXO3a, Forkhead box O3a; FOXO4, Forkhead box O4; GATA-1, GATA-binding protein 1; GATA-2, GATA-binding protein 2; GATA-3, GATA-binding protein 3; HNF-1, Hepatocyte nuclear factor 1; HNF-3, Hepatocyte nuclear factor 3 HNF-4, Hepatocyte nuclear factor 4; HNF-4A, Hepatocyte nuclear factor 4A; MEF2A, Myocyte-specific enhancer factor 2A; NF-1, Nuclear transcription factor 1, NKX2-5, NK 2 homeobox 5; NKX3-1, NK 3 homeobox 1; PAX4, Paired box 4; PAX5, Paired box 5; PAX6, Paired box 6; PPAR, Peroxisome proliferator-activated receptor; RXR, Retinoid X receptor; SMAD-1, mothers against decapentaplegic homolog 1; SMAD-3, mothers against decapentaplegic homolog 3; SMAD-4, mothers against decapentaplegic homolog 4; SP1, Specificity protein 1; USF, Upstream transcription factor; YY1, Ying Yang 1.

Appendix 5: Functional Promoter Analyses Supplementary Data.

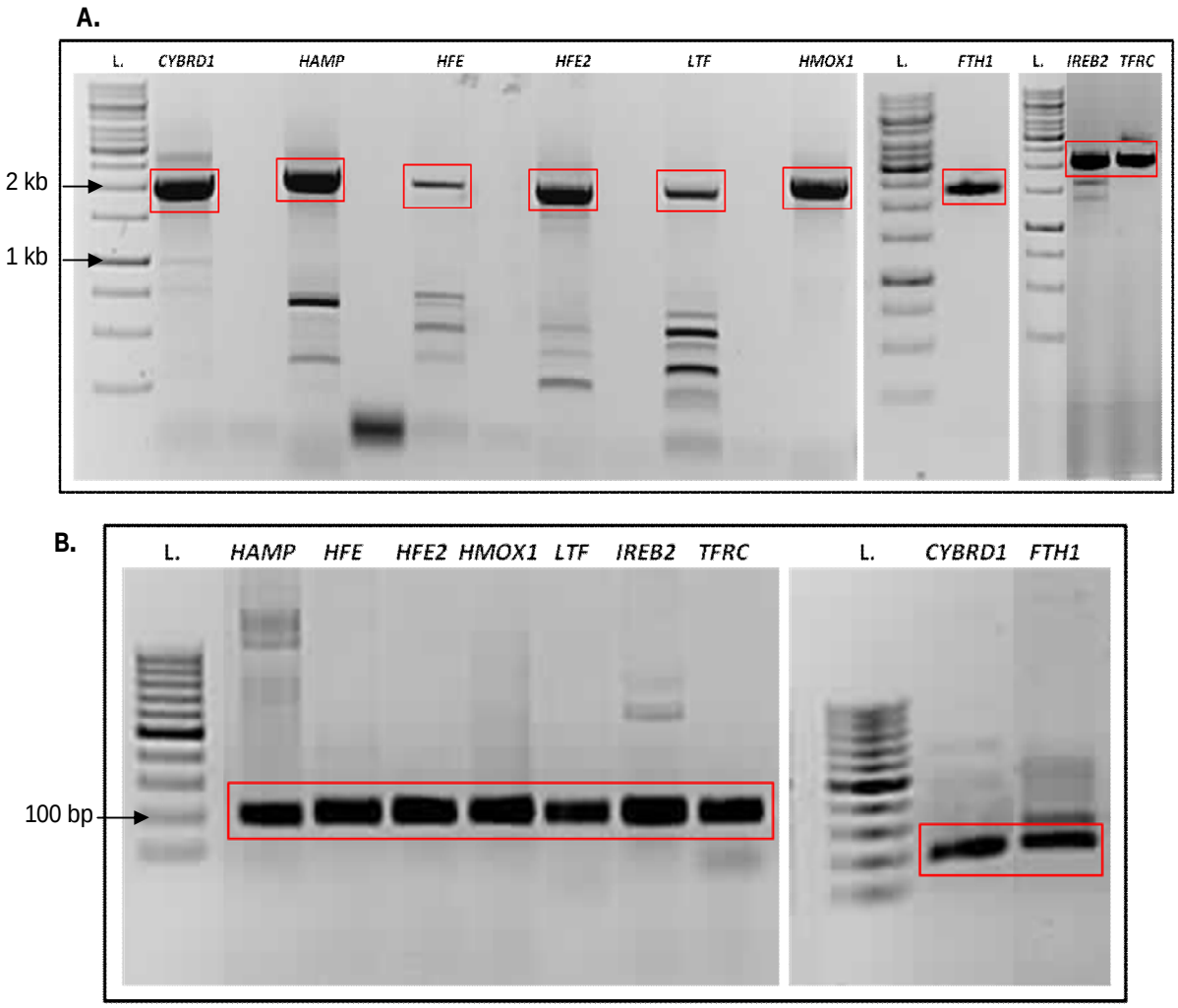


Figure S4 Agarose gel electrophoresis of PCR products.

PCR products highlighted in red were excised from the agarose gel and used in a Wizard® SV Gel and PCR Clean-up. **A.**) 2 kb fragment PCR amplification and **B.**) 140 bp CR element PCR amplification.

Abbreviations: bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; kb, kilobase pair; L., Ladder; *LTF*, Lactotransferrin gene; PCR, polymerase chain reaction; *TFRC*, Transferrin receptor protein 1 gene.

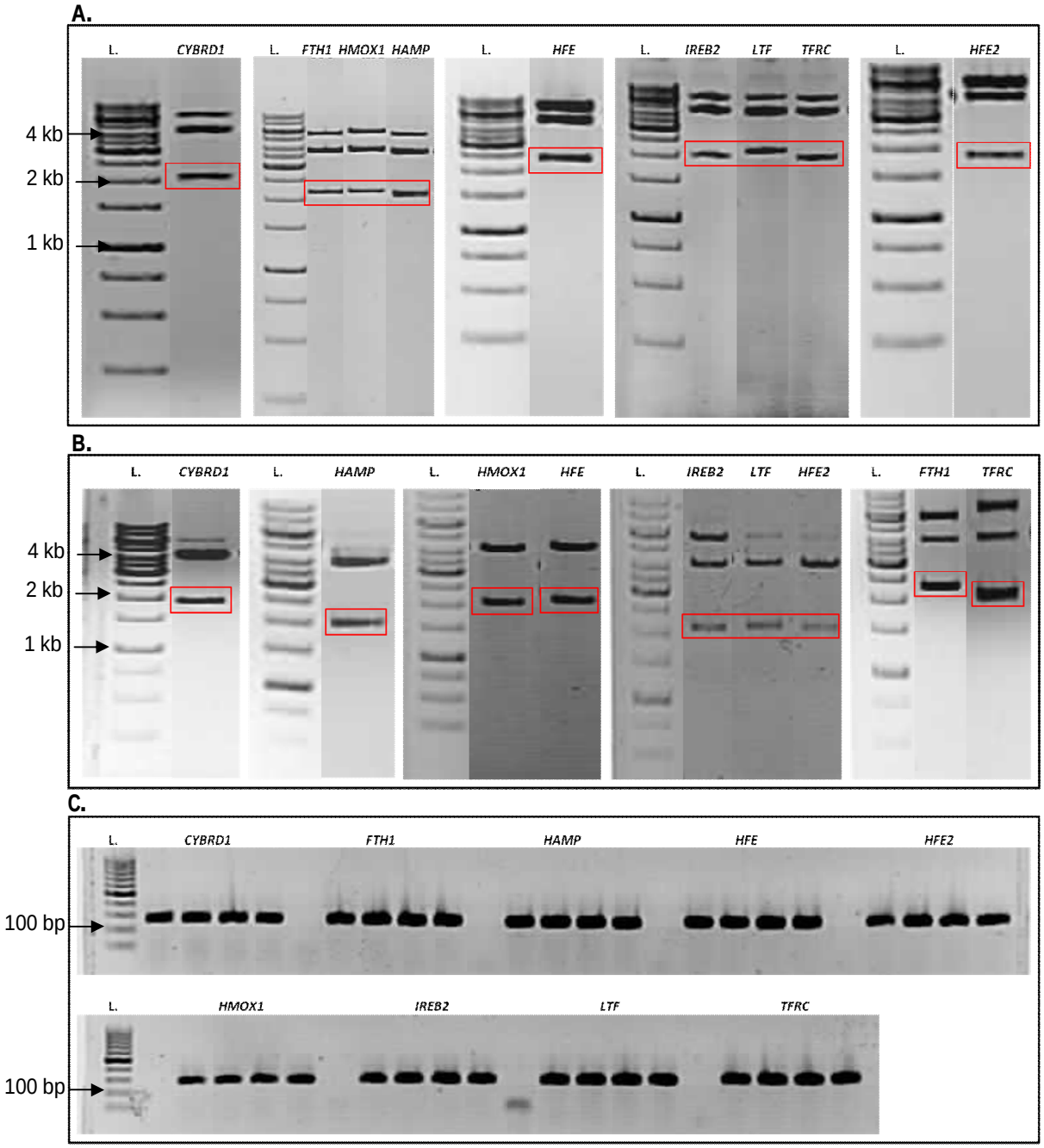


Figure S5 Clone selection.

Colonies were subjected to plasmid minipreps followed by double digestion of the extracted plasmids. Digestion products were electrophoresed on a 1% agarose gel and inserts of the correct size are highlighted in red. Larger bands above the insert of interest represent digested pGL4 plasmids (~4 kb) or undigested constructs (~6 kb). **A.**) 2 kb promoter constructs and **B.**) 1.86 kb CR-removed promoter constructs. **C.**) Colonies were subjected to colony PCR amplification to identify positive clones. For each gene, four separate colonies were chosen and were electrophoresed adjacent to each other on the 1% agarose gel.

Abbreviations: bp, base pair; *CYBRD1*, Cytochrome b reductase 1 gene; *FTH1*, Ferritin heavy polypeptide 1 gene; *HAMP*, Hepcidin antimicrobial peptide gene; *HFE*, Haemochromatosis gene, *HFE2*, Hemojuvelin gene; *HMOX1*, Haem oxygenase 1 gene; *IREB2*, Iron-responsive element-binding protein 2 gene; kb, kilobase pair; L., Ladder; *LTF*, Lactotransferrin gene; PCR, polymerase chain reaction; *TFRC*, Transferrin receptor protein 1 gene.

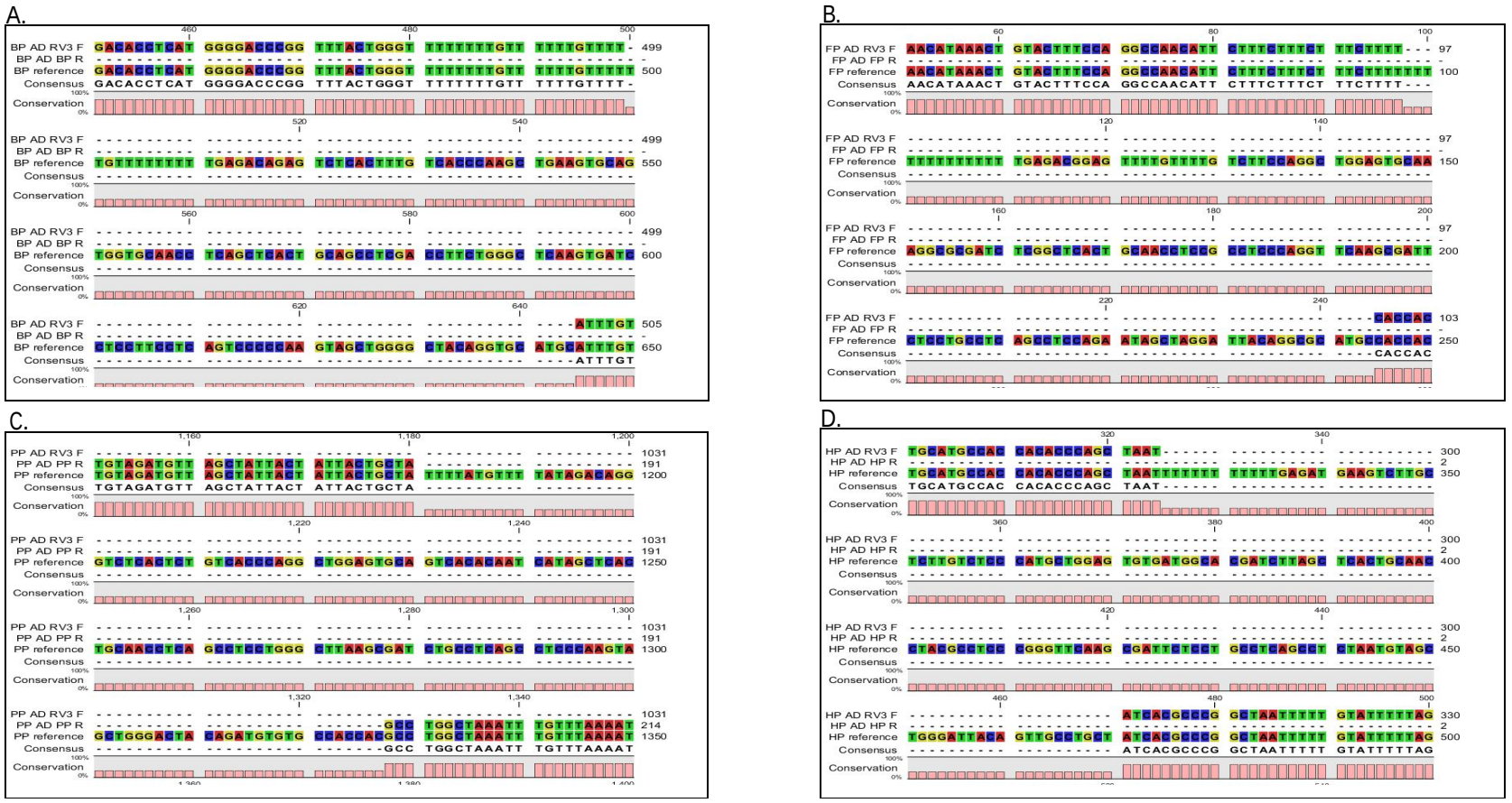
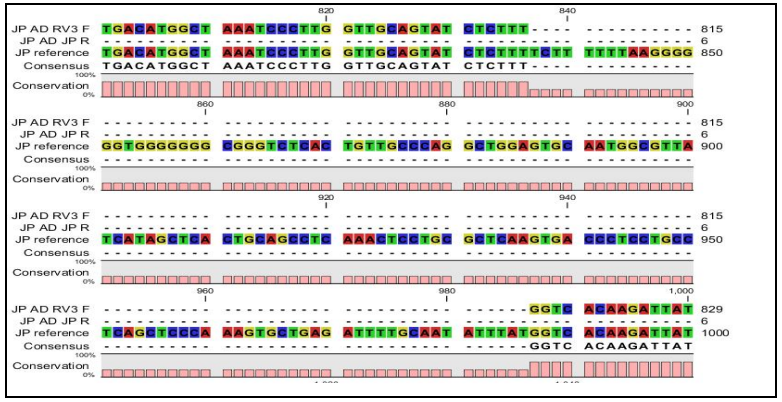


Figure S6 CLC Sequence Viewer 6.7.1 alignments of the 1.86 kb CR-removed promoter.

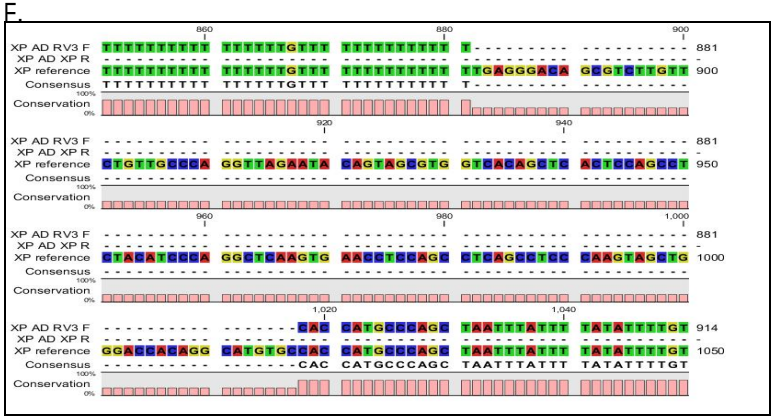
Sequencing results from the 1.86 kb CR-removed promoter constructs were aligned to the 2 kb reference sequence from each gene in order to confirm accurate deletion of the 140 bp CR element. **A.**) *CYBRD1*, **B.**) *FTH1*, **C.**) *HAMP* and **D.**) *HFE*.

Abbreviations: BP, Cytochrome b reductase 1 gene; F, Forward primer; FP, Ferritin heavy polypeptide 1 gene; HP, Haemochromatosis gene, PP, Hepcidin antimicrobial peptide gene; R, Reverse primer; RV3, Plasmid-specific forward primer.

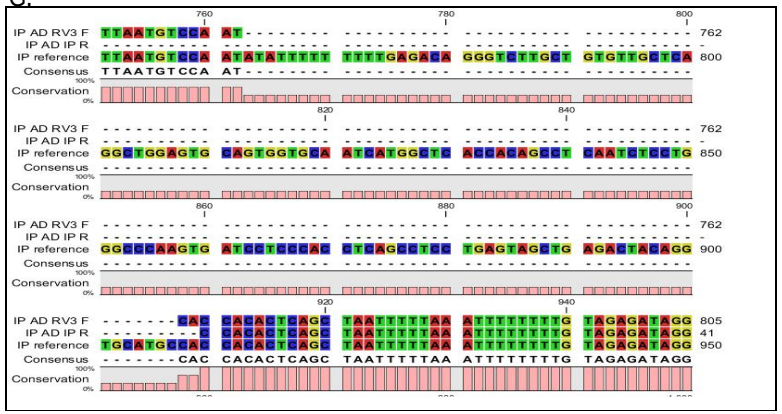
E.



F.



G.



H.

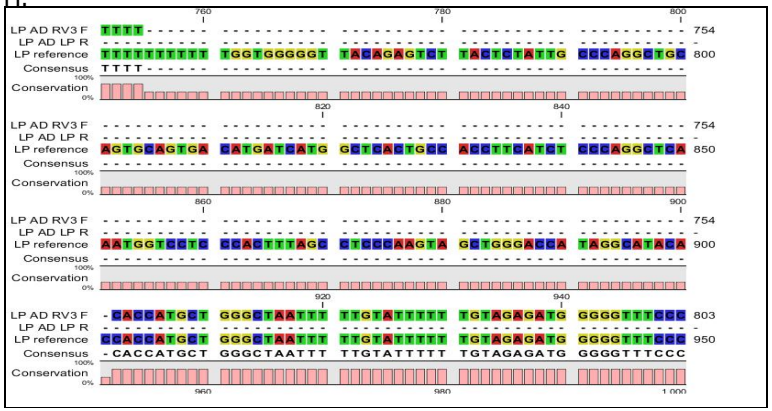


Figure S6 Cont. CLC Sequence Viewer 6.7.1 alignments of the 1.86 kb CR-removed promoter.

Sequencing results from the 1.86 kb CR-removed promoter constructs were aligned to the 2 kb reference sequence from each gene in order to confirm accurate deletion of the 140 bp CR element. **E.**) *HFE2*, **F.**) *HMOX1*, **G.**) *IREB2* and **H.**) *LTF*.

Abbreviations: F, Forward primer; IP, Iron-responsive element-binding protein 2 gene; JP, Hemojuvelin gene; LP, Lactotransferrin gene; R, Reverse primer; RV3, Plasmid-specific forward primer; XP, Haem oxygenase 1 gene.

I.

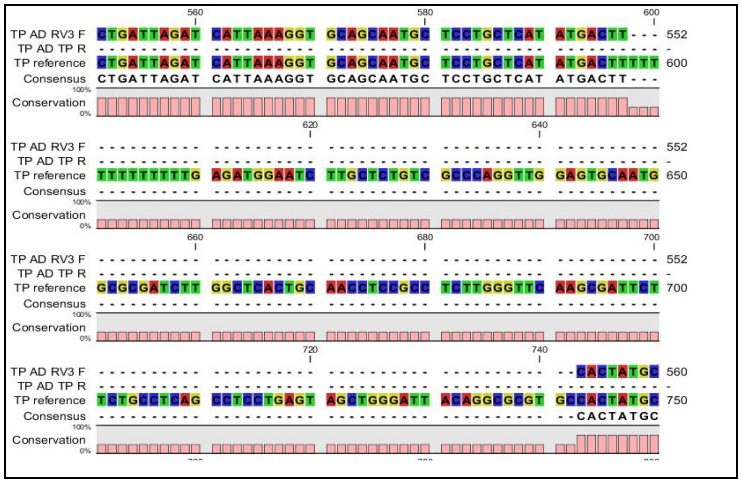


Figure S6 Cont. CLC Sequence Viewer 6.7.1 alignments of the 1.86 kb CR-removed promoter.

Sequencing results from the 1.86 kb CR-removed promoter constructs were aligned to the 2 kb reference sequence from each gene in order to confirm accurate deletion of the 140 bp CR element. **I.**) *TFRC*.

Abbreviations: F, Forward primer; R, Reverse primer; RV3, Plasmid-specific forward primer; TP, Transferrin receptor protein 1 gene.

Normalized Luciferase Data

Normalized values were calculated by dividing *firefly* luciferase values by the internal control *Renilla* luciferase vaules for each construct. Boxes shaded in grey indicate the construct and condition used as reference for calculating the fold change in expression levels. Values followed by an asterisk were excluded from further analysis as they did not lie within 20% of each other.

Table S1.1 Normalized luciferase data for *CYBRD1* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	10.18	1.18
	140 bp (dH_2O)	2	9.83	
	140 bp (dH_2O)	3	10.33	
	140 bp (FAC)	1	10.38	1.22
	140 bp (FAC)	2	10.72	
	140 bp (FAC)	3	10.46	
	pGL4.23 (dH_2O)	1	8.55	1.00
	pGL4.23 (dH_2O)	2	8.63	
	pGL4.23 (dH_2O)	3	1118.48 *	
	pGL4.23 (FAC)	1	7.48	0.95
	pGL4.23 (FAC)	2	8.63	
	pGL4.23 (FAC)	3	8.34	
2	140 bp (dH_2O)	1	2.75	1.07
	140 bp (dH_2O)	2	2.84	
	140 bp (dH_2O)	3	2.74	
	140 bp (FAC)	1	3.08	1.18
	140 bp (FAC)	2	2.85	
	140 bp (FAC)	3	3.24	
	pGL4.23 (dH_2O)	1	2.49	1.00
	pGL4.23 (dH_2O)	2	2.84	
	pGL4.23 (dH_2O)	3	2.48	
	pGL4.23 (FAC)	1	3.02	1.15
	pGL4.23 (FAC)	2	3.04	
	pGL4.23 (FAC)	3	2.91	
3	140 bp (dH_2O)	1	0.66	0.93
	140 bp (dH_2O)	2	0.73	
	140 bp (dH_2O)	3	0.82	
	140 bp (FAC)	1	0.78	1.11
	140 bp (FAC)	2	0.96	
	140 bp (FAC)	3	0.87	
	pGL4.23 (dH_2O)	1	0.74	1.00
	pGL4.23 (dH_2O)	2	0.83	
	pGL4.23 (dH_2O)	3	0.79	
	pGL4.23 (FAC)	1	0.93	1.17
	pGL4.23 (FAC)	2	0.44 *	
	pGL4.23 (FAC)	3	0.91	

Table S1.2 Normalized luciferase data for *CYBRD1* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	1.88 *	1.00
	2 kb (dH_2O)	2	2.43	
	2 kb (dH_2O)	3	3.05	
	2 kb (FAC)	1	3.39	1.27
	2 kb (FAC)	2	3.46	
	2 kb (FAC)	3	3.58	
	1.8 kb (dH_2O)	1	20.02	7.59
	1.8 kb (dH_2O)	2	20.27	
	1.8 kb (dH_2O)	3	22.08	
	1.8 kb (FAC)	1	15.71	5.59
	1.8 kb (FAC)	2	14.95	
	1.8 kb (FAC)	3	1.50 *	
	pGL4.10 (dH_2O)	1	0.01	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.01	0.00
	pGL4.10 (FAC)	2	0.01	
	pGL4.10 (FAC)	3	0.01	
2	2 kb (dH_2O)	1	1.87	1.00
	2 kb (dH_2O)	2	2.08	
	2 kb (dH_2O)	3	2.28	
	2 kb (FAC)	1	1.90 *	1.19
	2 kb (FAC)	2	2.46	
	2 kb (FAC)	3	2.49	
	1.8 kb (dH_2O)	1	29.81	13.05
	1.8 kb (dH_2O)	2	24.32	
	1.8 kb (dH_2O)	3	27.11	
	1.8 kb (FAC)	1	22.88	10.92
	1.8 kb (FAC)	2	21.73	
	1.8 kb (FAC)	3	23.42	
	pGL4.10 (dH_2O)	1	0.01 *	0.00
	pGL4.10 (dH_2O)	2	0.01	
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	0.00
	pGL4.10 (FAC)	2	0.01 *	
	pGL4.10 (FAC)	3	0.01	
3	2 kb (dH_2O)	1	2.86	1.00
	2 kb (dH_2O)	2	2.46	
	2 kb (dH_2O)	3	2.42	
	2 kb (FAC)	1	3.38 *	1.90
	2 kb (FAC)	2	4.86	
	2 kb (FAC)	3	4.92	
	1.8 kb (dH_2O)	1	38.07	15.06
	1.8 kb (dH_2O)	2	51.30 *	
	1.8 kb (dH_2O)	3	39.61	
	1.8 kb (FAC)	1	25.79	10.44
	1.8 kb (FAC)	2	19.23 *	
	1.8 kb (FAC)	3	28.05	
	pGL4.10 (dH_2O)	1	0.01 *	0.00
	pGL4.10 (dH_2O)	2	0.01	
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	0.00
	pGL4.10 (FAC)	2	0.01	
	pGL4.10 (FAC)	3	0.01	

Table S1.3 Normalized luciferase data for *CYBRD1* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	0.45	0.96
	140 bp (dH_2O)	2	0.43	
	140 bp (dH_2O)	3	0.46	
	140 bp (FAC)	1	0.58 *	0.90
	140 bp (FAC)	2	0.44	
	140 bp (FAC)	3	0.40	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.57	1.23
	pGL4.23 (FAC)	2	0.62	
	pGL4.23 (FAC)	3	0.52	
2	140 bp (dH_2O)	1	1.50 *	0.77
	140 bp (dH_2O)	2	0.55	
	140 bp (dH_2O)	3	0.58	
	140 bp (FAC)	1	0.59	0.76
	140 bp (FAC)	2	0.52	
	140 bp (FAC)	3	2.58 *	
	pGL4.23 (dH_2O)	1	0.71	1.00
	pGL4.23 (dH_2O)	2	0.76	
	pGL4.23 (dH_2O)	3	0.72	
	pGL4.23 (FAC)	1	0.72	0.94
	pGL4.23 (FAC)	2	0.67	
	pGL4.23 (FAC)	3	0.68	
3	140 bp (dH_2O)	1	0.33	0.83
	140 bp (dH_2O)	2	0.45 *	
	140 bp (dH_2O)	3	0.38	
	140 bp (FAC)	1	0.31	0.81
	140 bp (FAC)	2	0.37	
	140 bp (FAC)	3	0.36	
	pGL4.23 (dH_2O)	1	0.39	1.00
	pGL4.23 (dH_2O)	2	0.46	
	pGL4.23 (dH_2O)	3	0.43	
	pGL4.23 (FAC)	1	0.54	1.26
	pGL4.23 (FAC)	2	0.60	
	pGL4.23 (FAC)	3	0.47	

Table S1.4 Normalized luciferase data for *CYBRD1* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	0.75	1.00
	2 kb (dH_2O)	2	1.35 *	
	2 kb (dH_2O)	3	0.78	
	2 kb (FAC)	1	0.71	0.92
	2 kb (FAC)	2	0.69	
	2 kb (FAC)	3	1133.33 *	
	1.8 kb (dH_2O)	1	2.33	3.02
	1.8 kb (dH_2O)	2	2.31	
	1.8 kb (dH_2O)	3	2.29	
	1.8 kb (FAC)	1	2.14	2.97
	1.8 kb (FAC)	2	2.35	
	1.8 kb (FAC)	3	2.34	
	pGL4.10 (dH_2O)	1	0.00 *	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	0.92	1.00
	2 kb (dH_2O)	2	0.89	
	2 kb (dH_2O)	3	0.69 *	
	2 kb (FAC)	1	0.99	1.11
	2 kb (FAC)	2	0.80 *	
	2 kb (FAC)	3	1.02	
	1.8 kb (dH_2O)	1	2.52	2.79
	1.8 kb (dH_2O)	2	2.51	
	1.8 kb (dH_2O)	3	2.51	
	1.8 kb (FAC)	1	1.86 *	2.71
	1.8 kb (FAC)	2	2.44	
	1.8 kb (FAC)	3	2.44	
	pGL4.10 (dH_2O)	1	0.00 *	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00 *	
	pGL4.10 (FAC)	3	0.00	
3	2 kb (dH_2O)	1	0.82	1.00
	2 kb (dH_2O)	2	1.18 *	
	2 kb (dH_2O)	3	0.75	
	2 kb (FAC)	1	1.18	1.35
	2 kb (FAC)	2	0.96	
	2 kb (FAC)	3	1.03	
	1.8 kb (dH_2O)	1	3.35	4.18
	1.8 kb (dH_2O)	2	3.15	
	1.8 kb (dH_2O)	3	3.34	
	1.8 kb (FAC)	1	2.40	3.33
	1.8 kb (FAC)	2	2.83	
	1.8 kb (FAC)	3	1.97 *	
	pGL4.10 (dH_2O)	1	0.00	0.00
	pGL4.10 (dH_2O)	2	0.00 *	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00	

Table S2.1 Normalized luciferase data for *FTH1* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	78.00	8.73
	140 bp (dH_2O)	2	73.24	
	140 bp (dH_2O)	3	73.78	
	140 bp (FAC)	1	93.87	10.96
	140 bp (FAC)	2	94.25	
	140 bp (FAC)	3	94.23	
	pGL4.23 (dH_2O)	1	8.55	1.00
	pGL4.23 (dH_2O)	2	8.63	
	pGL4.23 (dH_2O)	3	1118.48 *	
	pGL4.23 (FAC)	1	7.48	0.95
	pGL4.23 (FAC)	2	8.63	
	pGL4.23 (FAC)	3	8.34	
2	140 bp (dH_2O)	1	21.79	8.78
	140 bp (dH_2O)	2	24.12	
	140 bp (dH_2O)	3	22.66	
	140 bp (FAC)	1	28.59	11.10
	140 bp (FAC)	2	30.16	
	140 bp (FAC)	3	27.94	
	pGL4.23 (dH_2O)	1	2.49	1.00
	pGL4.23 (dH_2O)	2	2.84	
	pGL4.23 (dH_2O)	3	2.48	
	pGL4.23 (FAC)	1	3.02	1.15
	pGL4.23 (FAC)	2	3.04	
	pGL4.23 (FAC)	3	2.91	
3	140 bp (dH_2O)	1	6.59	9.51
	140 bp (dH_2O)	2	6.90	
	140 bp (dH_2O)	3	7.40	
	140 bp (FAC)	1	9.49	12.61
	140 bp (FAC)	2	8.97	
	140 bp (FAC)	3	5.79 *	
	pGL4.23 (dH_2O)	1	0.71	1.00
	pGL4.23 (dH_2O)	2	0.75	
	pGL4.23 (dH_2O)	3	1.07 *	
	pGL4.23 (FAC)	1	0.61	0.82
	pGL4.23 (FAC)	2	0.63	
	pGL4.23 (FAC)	3	0.57	

Table S2.2 Normalized luciferase data for *FTH1* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	2.48	
	2 kb (dH_2O)	2	2.19	1.00
	2 kb (dH_2O)	3	1.79 *	
	2 kb (FAC)	1	3.37	
	2 kb (FAC)	2	3.83	1.54
	2 kb (FAC)	3	2.34 *	
	1.8 kb (dH_2O)	1	0.87	
	1.8 kb (dH_2O)	2	0.90	0.37
	1.8 kb (dH_2O)	3	0.85	
	1.8 kb (FAC)	1	1.75	
	1.8 kb (FAC)	2	1.77	0.75
	1.8 kb (FAC)	3	1.76	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.01	
2	2 kb (dH_2O)	1	2.54	
	2 kb (dH_2O)	2	2.90	1.00
	2 kb (dH_2O)	3	2.52	
	2 kb (FAC)	1	3.86	
	2 kb (FAC)	2	3.99	1.44
	2 kb (FAC)	3	3.60	
	1.8 kb (dH_2O)	1	0.57	
	1.8 kb (dH_2O)	2	0.57	0.22
	1.8 kb (dH_2O)	3	1.18 *	
	1.8 kb (FAC)	1	1.19	
	1.8 kb (FAC)	2	1.15	0.42
	1.8 kb (FAC)	3	1.04	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.00	
3	2 kb (dH_2O)	1	2.12	
	2 kb (dH_2O)	2	1.97	1.00
	2 kb (dH_2O)	3	2.03	
	2 kb (FAC)	1	4.64 *	
	2 kb (FAC)	2	3.18	1.50
	2 kb (FAC)	3	2.93	
	1.8 kb (dH_2O)	1	0.91	
	1.8 kb (dH_2O)	2	0.89	0.44
	1.8 kb (dH_2O)	3	1.33 *	
	1.8 kb (FAC)	1	1.38	
	1.8 kb (FAC)	2	1.41	0.67
	1.8 kb (FAC)	3	1.34	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01 *	0.00
	pGL4.10 (FAC)	3	0.01	

Table S2.3 Normalized luciferase data for *FTH1* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	2.05	4.88
	140 bp (dH_2O)	2	1.67	
	140 bp (dH_2O)	3	2.32	
	140 bp (FAC)	1	2.24	5.15
	140 bp (FAC)	2	2.15	
	140 bp (FAC)	3	1.98	
	pGL4.23 (dH_2O)	1	0.40	1.00
	pGL4.23 (dH_2O)	2	0.42	
	pGL4.23 (dH_2O)	3	0.51 *	
	pGL4.23 (FAC)	1	0.58	1.34
	pGL4.23 (FAC)	2	0.43 *	
	pGL4.23 (FAC)	3	0.52	
2	140 bp (dH_2O)	1	1.72	4.05
	140 bp (dH_2O)	2	1.87	
	140 bp (dH_2O)	3	2.07	
	140 bp (FAC)	1	2.60	5.82
	140 bp (FAC)	2	2.65	
	140 bp (FAC)	3	2.89	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.57	1.23
	pGL4.23 (FAC)	2	0.62	
	pGL4.23 (FAC)	3	0.52	
3	140 bp (dH_2O)	1	1.76	4.38
	140 bp (dH_2O)	2	1.94	
	140 bp (dH_2O)	3	1.88	
	140 bp (FAC)	1	2.29	5.38
	140 bp (FAC)	2	2.29	
	140 bp (FAC)	3	2.28	
	pGL4.23 (dH_2O)	1	0.39	1.00
	pGL4.23 (dH_2O)	2	0.46	
	pGL4.23 (dH_2O)	3	0.43	
	pGL4.23 (FAC)	1	0.54	1.26
	pGL4.23 (FAC)	2	0.60	
	pGL4.23 (FAC)	3	0.47	

Table S2.4 Normalized luciferase data for *FTH1* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	2.69	
	2 kb (dH_2O)	2	2.68	1.00
	2 kb (dH_2O)	3	2.91	
	2 kb (FAC)	1	4.66	
	2 kb (FAC)	2	4.64	1.68
	2 kb (FAC)	3	3.42 *	
	1.8 kb (dH_2O)	1	0.83	
	1.8 kb (dH_2O)	2	0.84	0.30
	1.8 kb (dH_2O)	3	1.30 *	
	1.8 kb (FAC)	1	1.40	
	1.8 kb (FAC)	2	1.39	0.50
	1.8 kb (FAC)	3	1.35	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	2.64	
	2 kb (dH_2O)	2	2.50	1.00
	2 kb (dH_2O)	3	2.67	
	2 kb (FAC)	1	3.38	
	2 kb (FAC)	2	3.03	1.26
	2 kb (FAC)	3	3.45	
	1.8 kb (dH_2O)	1	1.03	
	1.8 kb (dH_2O)	2	1.02	0.39
	1.8 kb (dH_2O)	3	1.02	
	1.8 kb (FAC)	1	1.35	
	1.8 kb (FAC)	2	1.37	0.52
	1.8 kb (FAC)	3	1.33	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	
3	2 kb (dH_2O)	1	2.38	
	2 kb (dH_2O)	2	2.35	1.00
	2 kb (dH_2O)	3	2.32	
	2 kb (FAC)	1	3.07	
	2 kb (FAC)	2	3.10	1.31
	2 kb (FAC)	3	2.40 *	
	1.8 kb (dH_2O)	1	0.75	
	1.8 kb (dH_2O)	2	0.71	0.33
	1.8 kb (dH_2O)	3	0.84	
	1.8 kb (FAC)	1	1.07	
	1.8 kb (FAC)	2	1.07	0.44
	1.8 kb (FAC)	3	1.00	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	

Table S3.1 Normalized luciferase data for *HAMP* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold change
1	140 bp (dH_2O)	1	1.28	0.47
	140 bp (dH_2O)	2	1.25	
	140 bp (dH_2O)	3	1.24	
	140 bp (FAC)	1	1.37	0.49
	140 bp (FAC)	2	1.31	
	140 bp (FAC)	3	1.30	
	pGL4.23 (dH_2O)	1	2.76	1.00
	pGL4.23 (dH_2O)	2	2.78	
	pGL4.23 (dH_2O)	3	2.53	
	pGL4.23 (FAC)	1	3.67	1.24
	pGL4.23 (FAC)	2	3.03	
	pGL4.23 (FAC)	3	1.85 *	
2	140 bp (dH_2O)	1	4.46	0.51
	140 bp (dH_2O)	2	4.44	
	140 bp (dH_2O)	3	4.12	
	140 bp (FAC)	1	4.67	0.58
	140 bp (FAC)	2	4.98	
	140 bp (FAC)	3	5.21	
	pGL4.23 (dH_2O)	1	8.55	1.00
	pGL4.23 (dH_2O)	2	8.63	
	pGL4.23 (dH_2O)	3	1118.48 *	
	pGL4.23 (FAC)	1	7.48	0.95
	pGL4.23 (FAC)	2	8.63	
	pGL4.23 (FAC)	3	8.34	
3	140 bp (dH_2O)	1	1.38	0.54
	140 bp (dH_2O)	2	1.45	
	140 bp (dH_2O)	3	1.37	
	140 bp (FAC)	1	1.53	0.63
	140 bp (FAC)	2	1.67	
	140 bp (FAC)	3	1.69	
	pGL4.23 (dH_2O)	1	2.49	1.00
	pGL4.23 (dH_2O)	2	2.84	
	pGL4.23 (dH_2O)	3	2.48	
	pGL4.23 (FAC)	1	3.02	1.15
	pGL4.23 (FAC)	2	3.04	
	pGL4.23 (FAC)	3	2.91	

Table S3.2 Normalized luciferase data for *HAMP* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	5.12 *	
	2 kb (dH_2O)	2	1.98	1.00
	2 kb (dH_2O)	3	2.16	
	2 kb (FAC)	1	2.39	
	2 kb (FAC)	2	1.95	1.07
	2 kb (FAC)	3	2.33	
	1.8 kb (dH_2O)	1	1.95	
	1.8 kb (dH_2O)	2	1.92	0.92
	1.8 kb (dH_2O)	3	1.80	
	1.8 kb (FAC)	1	2.15	
	1.8 kb (FAC)	2	1.68 *	1.02
	1.8 kb (FAC)	3	2.06	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00 *	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00 *	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	1.72	
	2 kb (dH_2O)	2	2.66 *	1.00
	2 kb (dH_2O)	3	1.99	
	2 kb (FAC)	1	2.46	
	2 kb (FAC)	2	2.05	1.16
	2 kb (FAC)	3	1.97	
	1.8 kb (dH_2O)	1	1.87	
	1.8 kb (dH_2O)	2	1.94	1.04
	1.8 kb (dH_2O)	3	1.98	
	1.8 kb (FAC)	1	2.27	
	1.8 kb (FAC)	2	2.77 *	1.14
	1.8 kb (FAC)	3	1.97	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.01	
3	2 kb (dH_2O)	1	1.68	
	2 kb (dH_2O)	2	1.66	1.00
	2 kb (dH_2O)	3	1.70	
	2 kb (FAC)	1	2.72	
	2 kb (FAC)	2	2.81	1.69
	2 kb (FAC)	3	2.96	
	1.8 kb (dH_2O)	1	1.95	
	1.8 kb (dH_2O)	2	1.17 *	1.16
	1.8 kb (dH_2O)	3	1.94	
	1.8 kb (FAC)	1	1.98	
	1.8 kb (FAC)	2	1.80 *	1.18
	1.8 kb (FAC)	3	1.98	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.00	

Table S3.3 Normalized luciferase data for *HAMP* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold change
1	140 bp (dH_2O)	1	0.48	0.97
	140 bp (dH_2O)	2	0.43	
	140 bp (dH_2O)	3	0.76 *	
	140 bp (FAC)	1	0.52	1.05
	140 bp (FAC)	2	0.45	
	140 bp (FAC)	3	0.48	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.57	1.23
	pGL4.23 (FAC)	2	0.62	
	pGL4.23 (FAC)	3	0.52	
2	140 bp (dH_2O)	1	0.57	0.77
	140 bp (dH_2O)	2	0.51	
	140 bp (dH_2O)	3	0.61	
	140 bp (FAC)	1	0.64	0.86
	140 bp (FAC)	2	0.80 *	
	140 bp (FAC)	3	0.62	
	pGL4.23 (dH_2O)	1	0.71	1.00
	pGL4.23 (dH_2O)	2	0.76	
	pGL4.23 (dH_2O)	3	0.72	
	pGL4.23 (FAC)	1	0.72	0.94
	pGL4.23 (FAC)	2	0.67	
	pGL4.23 (FAC)	3	0.68	
3	140 bp (dH_2O)	1	0.36 *	0.65
	140 bp (dH_2O)	2	0.29	
	140 bp (dH_2O)	3	0.26	
	140 bp (FAC)	1	0.28	0.66
	140 bp (FAC)	2	0.35 *	
	140 bp (FAC)	3	0.29	
	pGL4.23 (dH_2O)	1	0.39	1.00
	pGL4.23 (dH_2O)	2	0.46	
	pGL4.23 (dH_2O)	3	0.43	
	pGL4.23 (FAC)	1	0.54	1.19
	pGL4.23 (FAC)	2	0.50	
	pGL4.23 (FAC)	3	0.47	

Table S3.4 Normalized luciferase data for *HAMP* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	1.43	
	2 kb (dH_2O)	2	1.56	1.00
	2 kb (dH_2O)	3	1.24	
	2 kb (FAC)	1	1.35 *	
	2 kb (FAC)	2	1.55	1.13
	2 kb (FAC)	3	1.62	
	1.8 kb (dH_2O)	1	0.85	
	1.8 kb (dH_2O)	2	1.06 *	0.61
	1.8 kb (dH_2O)	3	0.86	
	1.8 kb (FAC)	1	1.10	
	1.8 kb (FAC)	2	1.10	0.78
	1.8 kb (FAC)	3	1.08	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00 *	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00 *	
2	2 kb (dH_2O)	1	1.62	
	2 kb (dH_2O)	2	1.06 *	1.00
	2 kb (dH_2O)	3	1.37	
	2 kb (FAC)	1	1.51	
	2 kb (FAC)	2	1.50	1.00
	2 kb (FAC)	3	1.49	
	1.8 kb (dH_2O)	1	0.86	
	1.8 kb (dH_2O)	2	0.83	0.57
	1.8 kb (dH_2O)	3	0.85	
	1.8 kb (FAC)	1	1.22 *	
	1.8 kb (FAC)	2	1.07	0.68
	1.8 kb (FAC)	3	0.95	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00 *	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00 *	
3	2 kb (dH_2O)	1	1.37	
	2 kb (dH_2O)	2	1.40	1.00
	2 kb (dH_2O)	3	1.36	
	2 kb (FAC)	1	1.38	
	2 kb (FAC)	2	1.40	1.01
	2 kb (FAC)	3	1.39	
	1.8 kb (dH_2O)	1	0.78	
	1.8 kb (dH_2O)	2	0.85	0.61
	1.8 kb (dH_2O)	3	0.89	
	1.8 kb (FAC)	1	1.04	
	1.8 kb (FAC)	2	1.04	0.76
	1.8 kb (FAC)	3	1.08	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	

Table S4.1 Normalized luciferase data for *HFE* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	0.01 *	0.00
	140 bp (dH_2O)	2	0.01	
	140 bp (dH_2O)	3	0.01	
	140 bp (FAC)	1	0.01 *	0.01
	140 bp (FAC)	2	0.01	
	140 bp (FAC)	3	0.01	
	pGL4.23 (dH_2O)	1	2.76	1.00
	pGL4.23 (dH_2O)	2	2.78	
	pGL4.23 (dH_2O)	3	2.53	
	pGL4.23 (FAC)	1	3.67	1.24
	pGL4.23 (FAC)	2	3.03	
	pGL4.23 (FAC)	3	1.85 *	
2	140 bp (dH_2O)	1	0.01	0.00
	140 bp (dH_2O)	2	0.01	
	140 bp (dH_2O)	3	22.35 *	
	140 bp (FAC)	1	0.02	0.01
	140 bp (FAC)	2	0.02	
	140 bp (FAC)	3	0.02	
	pGL4.23 (dH_2O)	1	2.49	1.00
	pGL4.23 (dH_2O)	2	2.84	
	pGL4.23 (dH_2O)	3	2.48	
	pGL4.23 (FAC)	1	3.02	1.15
	pGL4.23 (FAC)	2	3.04	
	pGL4.23 (FAC)	3	2.91	
3	140 bp (dH_2O)	1	0.05 *	0.01
	140 bp (dH_2O)	2	0.01	
	140 bp (dH_2O)	3	0.01	
	140 bp (FAC)	1	0.01 *	0.01
	140 bp (FAC)	2	0.02	
	140 bp (FAC)	3	0.02	
	pGL4.23 (dH_2O)	1	1.74	1.00
	pGL4.23 (dH_2O)	2	1.83	
	pGL4.23 (dH_2O)	3	1.79	
	pGL4.23 (FAC)	1	1.93	1.08
	pGL4.23 (FAC)	2	1.44 *	
	pGL4.23 (FAC)	3	1.91	

Table S4.2 Normalized luciferase data for *HFE* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	1.25	
	2 kb (dH_2O)	2	1.20	1.00
	2 kb (dH_2O)	3	0.99	
	2 kb (FAC)	1	3.71 *	
	2 kb (FAC)	2	1.45	1.25
	2 kb (FAC)	3	1.42	
	1.8 kb (dH_2O)	1	3.20 *	
	1.8 kb (dH_2O)	2	2.52	2.17
	1.8 kb (dH_2O)	3	2.45	
	1.8 kb (FAC)	1	2.22	
	1.8 kb (FAC)	2	2.43	2.10
	1.8 kb (FAC)	3	2.59	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	1.02	
	2 kb (dH_2O)	2	0.86 *	1.00
	2 kb (dH_2O)	3	1.18	
	2 kb (FAC)	1	1.72	
	2 kb (FAC)	2	2.02 *	1.55
	2 kb (FAC)	3	1.70	
	1.8 kb (dH_2O)	1	2.11	
	1.8 kb (dH_2O)	2	2.25	1.98
	1.8 kb (dH_2O)	3	2.18	
	1.8 kb (FAC)	1	1.74	
	1.8 kb (FAC)	2	1.75	1.61
	1.8 kb (FAC)	3	1.81	
	pGL4.10 (dH_2O)	1	0.01	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01	0.01
	pGL4.10 (FAC)	3	0.01	
3	2 kb (dH_2O)	1	2.30	
	2 kb (dH_2O)	2	1.70 *	1.00
	2 kb (dH_2O)	3	2.63	
	2 kb (FAC)	1	2.60	
	2 kb (FAC)	2	2.77	1.10
	2 kb (FAC)	3	2.80	
	1.8 kb (dH_2O)	1	3.38	
	1.8 kb (dH_2O)	2	3.43	1.38
	1.8 kb (dH_2O)	3	4.30 *	
	1.8 kb (FAC)	1	3.51	
	1.8 kb (FAC)	2	3.60	1.49
	1.8 kb (FAC)	3	3.92	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01 *	0.00
	pGL4.10 (FAC)	3	0.01	

Table S4.3 Normalized luciferase data for *HFE* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	0.01	0.01
	140 bp (dH_2O)	2	0.00 *	
	140 bp (dH_2O)	3	0.01	
	140 bp (FAC)	1	0.01	0.02
	140 bp (FAC)	2	0.01	
	140 bp (FAC)	3	0.01 *	
	pGL4.23 (dH_2O)	1	0.40	1.00
	pGL4.23 (dH_2O)	2	0.42	
	pGL4.23 (dH_2O)	3	0.51 *	
	pGL4.23 (FAC)	1	0.58	1.34
	pGL4.23 (FAC)	2	0.43 *	
	pGL4.23 (FAC)	3	0.52	
2	140 bp (dH_2O)	1	0.01	0.01
	140 bp (dH_2O)	2	0.01	
	140 bp (dH_2O)	3	0.01	
	140 bp (FAC)	1	0.01	0.02
	140 bp (FAC)	2	0.01	
	140 bp (FAC)	3	0.01 *	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.57	1.23
	pGL4.23 (FAC)	2	0.62	
	pGL4.23 (FAC)	3	0.52	
3	140 bp (dH_2O)	1	0.01 *	0.01
	140 bp (dH_2O)	2	0.01	
	140 bp (dH_2O)	3	0.01	
	140 bp (FAC)	1	0.01	0.02
	140 bp (FAC)	2	0.01	
	140 bp (FAC)	3	0.01	
	pGL4.23 (dH_2O)	1	0.71	1.00
	pGL4.23 (dH_2O)	2	0.76	
	pGL4.23 (dH_2O)	3	0.72	
	pGL4.23 (FAC)	1	0.72	0.94
	pGL4.23 (FAC)	2	0.67	
	pGL4.23 (FAC)	3	0.68	

Table S4.4 Normalized luciferase data for *HFE* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	0.90 *	
	2 kb (dH_2O)	2	0.75	1.00
	2 kb (dH_2O)	3	0.71	
	2 kb (FAC)	1	0.75 *	
	2 kb (FAC)	2	1.14	1.43
	2 kb (FAC)	3	0.94	
	1.8 kb (dH_2O)	1	0.83	
	1.8 kb (dH_2O)	2	0.85	1.11
	1.8 kb (dH_2O)	3	0.75	
	1.8 kb (FAC)	1	0.79	
	1.8 kb (FAC)	2	0.82	1.08
	1.8 kb (FAC)	3	0.75	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00 *	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00 *	
2	2 kb (dH_2O)	1	0.54	
	2 kb (dH_2O)	2	0.59	1.00
	2 kb (dH_2O)	3	0.56	
	2 kb (FAC)	1	0.62 *	
	2 kb (FAC)	2	0.83	1.41
	2 kb (FAC)	3	0.76	
	1.8 kb (dH_2O)	1	0.54	
	1.8 kb (dH_2O)	2	0.59	1.02
	1.8 kb (dH_2O)	3	0.59	
	1.8 kb (FAC)	1	0.55	
	1.8 kb (FAC)	2	0.58	1.01
	1.8 kb (FAC)	3	0.59	
	pGL4.10 (dH_2O)	1	0.00 *	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00 *	0.00
	pGL4.10 (FAC)	3	0.00	
3	2 kb (dH_2O)	1	0.51	
	2 kb (dH_2O)	2	0.71 *	1.00
	2 kb (dH_2O)	3	0.54	
	2 kb (FAC)	1	0.60	
	2 kb (FAC)	2	0.65	1.19
	2 kb (FAC)	3	0.63	
	1.8 kb (dH_2O)	1	0.62	
	1.8 kb (dH_2O)	2	0.72	1.30
	1.8 kb (dH_2O)	3	0.71	
	1.8 kb (FAC)	1	0.68	
	1.8 kb (FAC)	2	0.67	1.27
	1.8 kb (FAC)	3	0.66	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00 *	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	

Table S5.1 Normalized luciferase data for *HFE2* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	10.96	4.52
	140 bp (dH_2O)	2	12.47	
	140 bp (dH_2O)	3	13.05	
	140 bp (FAC)	1	15.04	6.00
	140 bp (FAC)	2	15.69	
	140 bp (FAC)	3	17.70	
	pGL4.23 (dH_2O)	1	2.76	1.00
	pGL4.23 (dH_2O)	2	2.78	
	pGL4.23 (dH_2O)	3	2.53	
	pGL4.23 (FAC)	1	3.67	1.06
	pGL4.23 (FAC)	2	3.03	
	pGL4.23 (FAC)	3	1.85 *	
2	140 bp (dH_2O)	1	10.20	4.27
	140 bp (dH_2O)	2	12.16	
	140 bp (dH_2O)	3	10.79	
	140 bp (FAC)	1	18.62	7.68
	140 bp (FAC)	2	20.29	
	140 bp (FAC)	3	20.71	
	pGL4.23 (dH_2O)	1	2.55	1.00
	pGL4.23 (dH_2O)	2	2.63	
	pGL4.23 (dH_2O)	3	18.48 *	
	pGL4.23 (FAC)	1	2.48	0.96
	pGL4.23 (FAC)	2	2.63	
	pGL4.23 (FAC)	3	2.34	
3	140 bp (dH_2O)	1	10.21	4.08
	140 bp (dH_2O)	2	11.63	
	140 bp (dH_2O)	3	10.02	
	140 bp (FAC)	1	12.46	5.34
	140 bp (FAC)	2	14.67	
	140 bp (FAC)	3	14.61	
	pGL4.23 (dH_2O)	1	2.49	1.00
	pGL4.23 (dH_2O)	2	2.84	
	pGL4.23 (dH_2O)	3	2.48	
	pGL4.23 (FAC)	1	3.02	1.15
	pGL4.23 (FAC)	2	3.04	
	pGL4.23 (FAC)	3	2.91	

Table S5.2 Normalized luciferase data for *HFE2* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	0.99 *	
	2 kb (dH_2O)	2	0.84	1.00
	2 kb (dH_2O)	3	0.80	
	2 kb (FAC)	1	0.89	
	2 kb (FAC)	2	0.87	1.12
	2 kb (FAC)	3	0.99	
	1.8 kb (dH_2O)	1	0.78	
	1.8 kb (dH_2O)	2	0.73	0.93
	1.8 kb (dH_2O)	3	0.79	
	1.8 kb (FAC)	1	1.49 *	
	1.8 kb (FAC)	2	1.12	1.38
	1.8 kb (FAC)	3	1.15	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	0.59 *	
	2 kb (dH_2O)	2	0.79	1.00
	2 kb (dH_2O)	3	0.81	
	2 kb (FAC)	1	1.08	
	2 kb (FAC)	2	1.63 *	1.39
	2 kb (FAC)	3	1.14	
	1.8 kb (dH_2O)	1	0.69	
	1.8 kb (dH_2O)	2	0.87 *	0.86
	1.8 kb (dH_2O)	3	0.67	
	1.8 kb (FAC)	1	0.96	
	1.8 kb (FAC)	2	0.99	0.80
	1.8 kb (FAC)	3	1.13 *	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.02 *	
3	2 kb (dH_2O)	1	1.02	
	2 kb (dH_2O)	2	1.12	1.00
	2 kb (dH_2O)	3	1.05	
	2 kb (FAC)	1	1.31	
	2 kb (FAC)	2	1.31	1.23
	2 kb (FAC)	3	1.53 *	
	1.8 kb (dH_2O)	1	1.05	
	1.8 kb (dH_2O)	2	1.09	1.01
	1.8 kb (dH_2O)	3	1.08	
	1.8 kb (FAC)	1	1.28	
	1.8 kb (FAC)	2	2.01 *	1.20
	1.8 kb (FAC)	3	1.27	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	

Table S5.3 Normalized luciferase data for *HFE2* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold change
1	140 bp (dH_2O)	1	1.02	2.79
	140 bp (dH_2O)	2	1.28	
	140 bp (dH_2O)	3	1.43 *	
	140 bp (FAC)	1	1.06	2.44
	140 bp (FAC)	2	0.96	
	140 bp (FAC)	3	0.99	
	pGL4.23 (dH_2O)	1	0.40	1.00
	pGL4.23 (dH_2O)	2	0.42	
	pGL4.23 (dH_2O)	3	0.51 *	
	pGL4.23 (FAC)	1	0.58	1.34
	pGL4.23 (FAC)	2	0.43 *	
	pGL4.23 (FAC)	3	0.52	
2	140 bp (dH_2O)	1	1.26	2.56
	140 bp (dH_2O)	2	1.18	
	140 bp (dH_2O)	3	1.14	
	140 bp (FAC)	1	0.99	2.20
	140 bp (FAC)	2	1.08	
	140 bp (FAC)	3	1.00	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.57	1.23
	pGL4.23 (FAC)	2	0.62	
	pGL4.23 (FAC)	3	0.52	
3	140 bp (dH_2O)	1	1.55	2.02
	140 bp (dH_2O)	2	2.41 *	
	140 bp (dH_2O)	3	1.39	
	140 bp (FAC)	1	1.29	1.64
	140 bp (FAC)	2	1.14	
	140 bp (FAC)	3	1.17	
	pGL4.23 (dH_2O)	1	0.71	1.00
	pGL4.23 (dH_2O)	2	0.76	
	pGL4.23 (dH_2O)	3	0.72	
	pGL4.23 (FAC)	1	0.72	0.94
	pGL4.23 (FAC)	2	0.67	
	pGL4.23 (FAC)	3	0.68	

Table S5.4 Normalized luciferase data for *HFE2* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	2.27	1.00
	2 kb (dH_2O)	2	2.18	
	2 kb (dH_2O)	3	2.10	
	2 kb (FAC)	1	3.96	1.78
	2 kb (FAC)	2	3.80	
	2 kb (FAC)	3	3.88	
	1.8 kb (dH_2O)	1	2.40	1.11
	1.8 kb (dH_2O)	2	2.38	
	1.8 kb (dH_2O)	3	2.49	
	1.8 kb (FAC)	1	2.64	1.27
	1.8 kb (FAC)	2	2.65	
	1.8 kb (FAC)	3	3.02	
	pGL4.10 (dH_2O)	1	0.00	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	3.05	1.00
	2 kb (dH_2O)	2	2.71	
	2 kb (dH_2O)	3	2.83	
	2 kb (FAC)	1	3.63	1.18
	2 kb (FAC)	2	3.31	
	2 kb (FAC)	3	3.22	
	1.8 kb (dH_2O)	1	3.19	1.03
	1.8 kb (dH_2O)	2	2.82	
	1.8 kb (dH_2O)	3	2.83	
	1.8 kb (FAC)	1	3.03	1.07
	1.8 kb (FAC)	2	3.27	
	1.8 kb (FAC)	3	2.91	
	pGL4.10 (dH_2O)	1	0.00	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00 *	
	pGL4.10 (FAC)	3	0.00	
3	2 kb (dH_2O)	1	2.36	1.00
	2 kb (dH_2O)	2	2.56	
	2 kb (dH_2O)	3	2.45	
	2 kb (FAC)	1	2.09 *	1.08
	2 kb (FAC)	2	2.63	
	2 kb (FAC)	3	2.66	
	1.8 kb (dH_2O)	1	2.13	0.92
	1.8 kb (dH_2O)	2	2.54	
	1.8 kb (dH_2O)	3	2.09	
	1.8 kb (FAC)	1	2.60	1.02
	1.8 kb (FAC)	2	2.23	
	1.8 kb (FAC)	3	2.71	
	pGL4.10 (dH_2O)	1	0.00	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00 *	

Table S6.1 Normalized luciferase data for *HMOX1* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold change
1	140 bp (dH_2O)	1	1.10	0.40
	140 bp (dH_2O)	2	1.03	
	140 bp (dH_2O)	3	1.08	
	140 bp (FAC)	1	1.29	0.49
	140 bp (FAC)	2	1.26	
	140 bp (FAC)	3	1.42	
	pGL4.23 (dH_2O)	1	2.76	1.00
	pGL4.23 (dH_2O)	2	2.78	
	pGL4.23 (dH_2O)	3	2.53	
	pGL4.23 (FAC)	1	3.67	1.24
	pGL4.23 (FAC)	2	3.03	
	pGL4.23 (FAC)	3	1.85 *	
2	140 bp (dH_2O)	1	3.71	0.43
	140 bp (dH_2O)	2	3.74	
	140 bp (dH_2O)	3	7.06 *	
	140 bp (FAC)	1	4.41	0.52
	140 bp (FAC)	2	4.25	
	140 bp (FAC)	3	4.84	
	pGL4.23 (dH_2O)	1	8.55	1.00
	pGL4.23 (dH_2O)	2	8.63	
	pGL4.23 (dH_2O)	3	1118.48 *	
	pGL4.23 (FAC)	1	7.48	0.95
	pGL4.23 (FAC)	2	8.63	
	pGL4.23 (FAC)	3	8.34	
3	140 bp (dH_2O)	1	1.42	0.55
	140 bp (dH_2O)	2	1.39	
	140 bp (dH_2O)	3	1.51	
	140 bp (FAC)	1	1.56	0.62
	140 bp (FAC)	2	1.78	
	140 bp (FAC)	3	1.52	
	pGL4.23 (dH_2O)	1	2.49	1.00
	pGL4.23 (dH_2O)	2	2.84	
	pGL4.23 (dH_2O)	3	2.48	
	pGL4.23 (FAC)	1	3.02	1.15
	pGL4.23 (FAC)	2	3.04	
	pGL4.23 (FAC)	3	2.91	

Table S6.2 Normalized luciferase data for *HMOX1* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	38.10	
	2 kb (dH_2O)	2	23.97 *	1.00
	2 kb (dH_2O)	3	32.86	
	2 kb (FAC)	1	52.59	
	2 kb (FAC)	2	62.75	1.64
	2 kb (FAC)	3	59.61	
	1.8 kb (dH_2O)	1	249.62	
	1.8 kb (dH_2O)	2	209.62	6.98
	1.8 kb (dH_2O)	3	283.78	
	1.8 kb (FAC)	1	314.96	
	1.8 kb (FAC)	2	364.40	9.57
	1.8 kb (FAC)	3	123.36 *	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00 *	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00 *	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	25.26	
	2 kb (dH_2O)	2	28.43	1.00
	2 kb (dH_2O)	3	28.60	
	2 kb (FAC)	1	34.46	
	2 kb (FAC)	2	53.55 *	1.31
	2 kb (FAC)	3	37.66	
	1.8 kb (dH_2O)	1	199.34 *	
	1.8 kb (dH_2O)	2	157.54	6.00
	1.8 kb (dH_2O)	3	171.69	
	1.8 kb (FAC)	1	216.37	
	1.8 kb (FAC)	2	191.27	7.51
	1.8 kb (FAC)	3	210.36	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.01	
3	2 kb (dH_2O)	1	32.95	
	2 kb (dH_2O)	2	36.58	1.00
	2 kb (dH_2O)	3	42.02 *	
	2 kb (FAC)	1	55.63	
	2 kb (FAC)	2	57.60	1.62
	2 kb (FAC)	3	55.61	
	1.8 kb (dH_2O)	1	176.62	
	1.8 kb (dH_2O)	2	203.82	5.12
	1.8 kb (dH_2O)	3	153.76	
	1.8 kb (FAC)	1	263.40	
	1.8 kb (FAC)	2	373.85 *	7.14
	1.8 kb (FAC)	3	233.39	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.00	

Table S6.3 Normalized luciferase data for *HMOX1* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold change
1	140 bp (dH_2O)	1	0.48	1.05
	140 bp (dH_2O)	2	0.51	
	140 bp (dH_2O)	3	0.48	
	140 bp (FAC)	1	0.54	1.20
	140 bp (FAC)	2	0.54	
	140 bp (FAC)	3	0.59	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.57	1.23
	pGL4.23 (FAC)	2	0.62	
	pGL4.23 (FAC)	3	0.52	
2	140 bp (dH_2O)	1	0.41	0.76
	140 bp (dH_2O)	2	0.47	
	140 bp (dH_2O)	3	0.46	
	140 bp (FAC)	1	0.53	0.81
	140 bp (FAC)	2	0.44	
	140 bp (FAC)	3	0.46	
	pGL4.23 (dH_2O)	1	0.66 *	1.00
	pGL4.23 (dH_2O)	2	0.52	
	pGL4.23 (dH_2O)	3	0.57	
	pGL4.23 (FAC)	1	0.72	1.21
	pGL4.23 (FAC)	2	0.71	
	pGL4.23 (FAC)	3	0.70	
3	140 bp (dH_2O)	1	0.41	1.01
	140 bp (dH_2O)	2	0.44	
	140 bp (dH_2O)	3	0.44	
	140 bp (FAC)	1	0.50	1.12
	140 bp (FAC)	2	0.46	
	140 bp (FAC)	3	0.47	
	pGL4.23 (dH_2O)	1	0.39	1.00
	pGL4.23 (dH_2O)	2	0.46	
	pGL4.23 (dH_2O)	3	0.43	
	pGL4.23 (FAC)	1	0.54	1.26
	pGL4.23 (FAC)	2	0.60	
	pGL4.23 (FAC)	3	0.47	

Table S6.4 Normalized luciferase data for *HMOX1* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	16.26	1.00
	2 kb (dH_2O)	2	16.76	
	2 kb (dH_2O)	3	17.13	
	2 kb (FAC)	1	20.54	1.20
	2 kb (FAC)	2	15.47 *	
	2 kb (FAC)	3	19.66	
	1.8 kb (dH_2O)	1	30.47	1.88
	1.8 kb (dH_2O)	2	30.26	
	1.8 kb (dH_2O)	3	33.62	
	1.8 kb (FAC)	1	36.08	2.24
	1.8 kb (FAC)	2	37.88	
	1.8 kb (FAC)	3	38.44	
	pGL4.10 (dH_2O)	1	0.00	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00 *	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00 *	
2	2 kb (dH_2O)	1	15.24	1.00
	2 kb (dH_2O)	2	18.37	
	2 kb (dH_2O)	3	18.24	
	2 kb (FAC)	1	24.95	1.48
	2 kb (FAC)	2	24.55	
	2 kb (FAC)	3	27.13	
	1.8 kb (dH_2O)	1	42.59	2.42
	1.8 kb (dH_2O)	2	43.51	
	1.8 kb (dH_2O)	3	39.38	
	1.8 kb (FAC)	1	45.94	2.58
	1.8 kb (FAC)	2	46.32	
	1.8 kb (FAC)	3	41.74	
	pGL4.10 (dH_2O)	1	0.00	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00 *	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00 *	
3	2 kb (dH_2O)	1	10.42	1.00
	2 kb (dH_2O)	2	9.97	
	2 kb (dH_2O)	3	11.38	
	2 kb (FAC)	1	21.16	1.97
	2 kb (FAC)	2	20.64	
	2 kb (FAC)	3	20.64	
	1.8 kb (dH_2O)	1	27.93	2.68
	1.8 kb (dH_2O)	2	28.78	
	1.8 kb (dH_2O)	3	19.35 *	
	1.8 kb (FAC)	1	29.42	2.74
	1.8 kb (FAC)	2	29.34	
	1.8 kb (FAC)	3	28.37	
	pGL4.10 (dH_2O)	1	0.00	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00	

Table S7.1 Normalized luciferase data for *IREB2* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	0.25	0.25
	140 bp (dH_2O)	2	0.27	
	140 bp (dH_2O)	3	0.24	
	140 bp (FAC)	1	0.28	0.29
	140 bp (FAC)	2	0.29	
	140 bp (FAC)	3	0.29	
	pGL4.23 (dH_2O)	1	0.65	0.64
	pGL4.23 (dH_2O)	2	0.63	
	pGL4.23 (dH_2O)	3	1.85 *	
	pGL4.23 (FAC)	1	0.79	0.83
	pGL4.23 (FAC)	2	0.86	
	pGL4.23 (FAC)	3	0.83	
2	140 bp (dH_2O)	1	0.25	0.25
	140 bp (dH_2O)	2	0.26	
	140 bp (dH_2O)	3	0.25	
	140 bp (FAC)	1	0.29	0.29
	140 bp (FAC)	2	0.44 *	
	140 bp (FAC)	3	0.30	
	pGL4.23 (dH_2O)	1	0.74	0.79
	pGL4.23 (dH_2O)	2	0.83	
	pGL4.23 (dH_2O)	3	0.79	
	pGL4.23 (FAC)	1	0.93	0.92
	pGL4.23 (FAC)	2	0.44 *	
	pGL4.23 (FAC)	3	0.91	
3	140 bp (dH_2O)	1	0.31 *	0.22
	140 bp (dH_2O)	2	0.22	
	140 bp (dH_2O)	3	0.22	
	140 bp (FAC)	1	0.31	0.27
	140 bp (FAC)	2	0.26	
	140 bp (FAC)	3	0.23	
	pGL4.23 (dH_2O)	1	0.61	0.81
	pGL4.23 (dH_2O)	2	0.75	
	pGL4.23 (dH_2O)	3	1.07	
	pGL4.23 (FAC)	1	0.61	0.60
	pGL4.23 (FAC)	2	0.63	
	pGL4.23 (FAC)	3	0.57	

Table S7.2 Normalized luciferase data for *IREB2* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	1022.94	
	2 kb (dH_2O)	2	1048.32	1.00
	2 kb (dH_2O)	3	934.16	
	2 kb (FAC)	1	954.04	
	2 kb (FAC)	2	1089.60	1.03
	2 kb (FAC)	3	1051.40	
	1.8 kb (dH_2O)	1	1133.88	
	1.8 kb (dH_2O)	2	952.08	0.99
	1.8 kb (dH_2O)	3	878.94	
	1.8 kb (FAC)	1	1055.33	
	1.8 kb (FAC)	2	1048.34	1.04
	1.8 kb (FAC)	3	1025.92	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.02 *	
2	2 kb (dH_2O)	1	1446.29	
	2 kb (dH_2O)	2	1463.71	1.00
	2 kb (dH_2O)	3	1396.98	
	2 kb (FAC)	1	1369.75	
	2 kb (FAC)	2	1664.37	1.09
	2 kb (FAC)	3	1652.96	
	1.8 kb (dH_2O)	1	1514.93	
	1.8 kb (dH_2O)	2	1414.18	0.98
	1.8 kb (dH_2O)	3	1308.59	
	1.8 kb (FAC)	1	1715.78	
	1.8 kb (FAC)	2	1726.83	1.17
	1.8 kb (FAC)	3	1602.60	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	
3	2 kb (dH_2O)	1	1092.35	
	2 kb (dH_2O)	2	1007.84	1.00
	2 kb (dH_2O)	3	985.80	
	2 kb (FAC)	1	1029.80	
	2 kb (FAC)	2	1202.11	1.12
	2 kb (FAC)	3	1212.50	
	1.8 kb (dH_2O)	1	1231.66	
	1.8 kb (dH_2O)	2	1096.89	1.14
	1.8 kb (dH_2O)	3	1194.75	
	1.8 kb (FAC)	1	1267.58	
	1.8 kb (FAC)	2	1337.91	1.27
	1.8 kb (FAC)	3	1025.94 *	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	

Table S7.3 Normalized luciferase data for *IREB2* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	0.27	0.61
	140 bp (dH_2O)	2	0.23	
	140 bp (dH_2O)	3	1.31 *	
	140 bp (FAC)	1	0.29	0.66
	140 bp (FAC)	2	0.54 *	
	140 bp (FAC)	3	0.26	
	pGL4.23 (dH_2O)	1	0.40	1.00
	pGL4.23 (dH_2O)	2	0.42	
	pGL4.23 (dH_2O)	3	0.51 *	
	pGL4.23 (FAC)	1	0.58	1.34
	pGL4.23 (FAC)	2	0.43 *	
	pGL4.23 (FAC)	3	0.52	
2	140 bp (dH_2O)	1	0.26	0.57
	140 bp (dH_2O)	2	0.25	
	140 bp (dH_2O)	3	0.29	
	140 bp (FAC)	1	0.40	0.79
	140 bp (FAC)	2	0.34	
	140 bp (FAC)	3	0.36	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.57	1.23
	pGL4.23 (FAC)	2	0.62	
	pGL4.23 (FAC)	3	0.52	
3	140 bp (dH_2O)	1	0.43	0.58
	140 bp (dH_2O)	2	0.43	
	140 bp (dH_2O)	3	0.42	
	140 bp (FAC)	1	0.51	0.70
	140 bp (FAC)	2	0.51	
	140 bp (FAC)	3	0.95 *	
	pGL4.23 (dH_2O)	1	0.71	1.00
	pGL4.23 (dH_2O)	2	0.76	
	pGL4.23 (dH_2O)	3	0.72	
	pGL4.23 (FAC)	1	0.72	0.94
	pGL4.23 (FAC)	2	0.67	
	pGL4.23 (FAC)	3	0.68	

Table S7.4 Normalized luciferase data for *IREB2* constructs transfected into COS-1 HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	873.15	
	2 kb (dH_2O)	2	920.37	1.00
	2 kb (dH_2O)	3	908.69	
	2 kb (FAC)	1	855.62	
	2 kb (FAC)	2	585.75 *	0.92
	2 kb (FAC)	3	802.89	
	1.8 kb (dH_2O)	1	822.70	
	1.8 kb (dH_2O)	2	787.78	0.87
	1.8 kb (dH_2O)	3	738.47	
	1.8 kb (FAC)	1	680.80	
	1.8 kb (FAC)	2	698.90	0.80
	1.8 kb (FAC)	3	770.22	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00 *	0.00
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	1025.68	
	2 kb (dH_2O)	2	983.37	1.00
	2 kb (dH_2O)	3	956.62	
	2 kb (FAC)	1	939.79	
	2 kb (FAC)	2	976.14	0.96
	2 kb (FAC)	3	943.19	
	1.8 kb (dH_2O)	1	862.67	
	1.8 kb (dH_2O)	2	827.19	0.85
	1.8 kb (dH_2O)	3	2.00 *	
	1.8 kb (FAC)	1	847.27	
	1.8 kb (FAC)	2	829.95	0.83
	1.8 kb (FAC)	3	796.62	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00 *	
3	2 kb (dH_2O)	1	1241.81	
	2 kb (dH_2O)	2	1059.02	1.00
	2 kb (dH_2O)	3	1182.93	
	2 kb (FAC)	1	1126.22	
	2 kb (FAC)	2	1111.77	0.98
	2 kb (FAC)	3	1168.62	
	1.8 kb (dH_2O)	1	992.71	
	1.8 kb (dH_2O)	2	1019.39	0.88
	1.8 kb (dH_2O)	3	1045.29	
	1.8 kb (FAC)	1	951.64	
	1.8 kb (FAC)	2	973.72	0.82
	1.8 kb (FAC)	3	926.35	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00 *	

Table S8.1 Normalized luciferase data for *LTF* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	0.68	0.27
	140 bp (dH_2O)	2	0.77	
	140 bp (dH_2O)	3	0.72	
	140 bp (FAC)	1	0.93	0.36
	140 bp (FAC)	2	1.03	
	140 bp (FAC)	3	0.98	
	pGL4.23 (dH_2O)	1	2.76	1.00
	pGL4.23 (dH_2O)	2	2.78	
	pGL4.23 (dH_2O)	3	2.53	
	pGL4.23 (FAC)	1	3.67	1.24
	pGL4.23 (FAC)	2	3.03	
	pGL4.23 (FAC)	3	1.85 *	
2	140 bp (dH_2O)	1	1.92	0.24
	140 bp (dH_2O)	2	2.34	
	140 bp (dH_2O)	3	1.98	
	140 bp (FAC)	1	2.35	0.30
	140 bp (FAC)	2	2.73	
	140 bp (FAC)	3	2.58	
	pGL4.23 (dH_2O)	1	8.55	1.00
	pGL4.23 (dH_2O)	2	8.63	
	pGL4.23 (dH_2O)	3	1118.48 *	
	pGL4.23 (FAC)	1	7.48	0.95
	pGL4.23 (FAC)	2	8.63	
	pGL4.23 (FAC)	3	8.34	
3	140 bp (dH_2O)	1	0.90	0.33
	140 bp (dH_2O)	2	0.79	
	140 bp (dH_2O)	3	0.90	
	140 bp (FAC)	1	1.05	0.39
	140 bp (FAC)	2	1.04	
	140 bp (FAC)	3	0.94	
	pGL4.23 (dH_2O)	1	2.49	1.00
	pGL4.23 (dH_2O)	2	2.84	
	pGL4.23 (dH_2O)	3	2.48	
	pGL4.23 (FAC)	1	3.02	1.15
	pGL4.23 (FAC)	2	3.04	
	pGL4.23 (FAC)	3	2.91	

Table S8.2 Normalized luciferase data for *LTF* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	4.31 *	
	2 kb (dH_2O)	2	0.87	1.00
	2 kb (dH_2O)	3	0.90	
	2 kb (FAC)	1	0.37	
	2 kb (FAC)	2	0.48	0.48
	2 kb (FAC)	3	0.72 *	
	1.8 kb (dH_2O)	1	0.55	
	1.8 kb (dH_2O)	2	1.29 *	0.63
	1.8 kb (dH_2O)	3	0.56	
	1.8 kb (FAC)	1	0.16 *	
	1.8 kb (FAC)	2	0.46	0.49
	1.8 kb (FAC)	3	0.41	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00 *	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.00 *	0.01
	pGL4.10 (FAC)	3	0.01	
2	2 kb (dH_2O)	1	0.64	
	2 kb (dH_2O)	2	0.58	1.00
	2 kb (dH_2O)	3	0.73	
	2 kb (FAC)	1	0.36	
	2 kb (FAC)	2	0.42	0.62
	2 kb (FAC)	3	0.42	
	1.8 kb (dH_2O)	1	0.41	
	1.8 kb (dH_2O)	2	0.48	0.70
	1.8 kb (dH_2O)	3	0.49	
	1.8 kb (FAC)	1	0.53 *	
	1.8 kb (FAC)	2	0.42	0.61
	1.8 kb (FAC)	3	0.37	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.01
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01 *	0.01
	pGL4.10 (FAC)	3	0.01	
3	2 kb (dH_2O)	1	0.95	
	2 kb (dH_2O)	2	0.68 *	1.00
	2 kb (dH_2O)	3	0.83	
	2 kb (FAC)	1	0.28 *	
	2 kb (FAC)	2	0.39	0.47
	2 kb (FAC)	3	0.45	
	1.8 kb (dH_2O)	1	0.45	
	1.8 kb (dH_2O)	2	1.07 *	0.52
	1.8 kb (dH_2O)	3	0.47	
	1.8 kb (FAC)	1	0.39	
	1.8 kb (FAC)	2	0.38	0.44
	1.8 kb (FAC)	3	0.57 *	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.01
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01	0.01
	pGL4.10 (FAC)	3	0.01	

Table S8.3 Normalized luciferase data for *LTF* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	0.22	0.54
	140 bp (dH_2O)	2	0.22	
	140 bp (dH_2O)	3	0.22	
	140 bp (FAC)	1	0.64 *	0.83
	140 bp (FAC)	2	0.36	
	140 bp (FAC)	3	0.33	
	pGL4.23 (dH_2O)	1	0.40	1.00
	pGL4.23 (dH_2O)	2	0.42	
	pGL4.23 (dH_2O)	3	0.51 *	
	pGL4.23 (FAC)	1	0.58	1.34
	pGL4.23 (FAC)	2	0.43 *	
	pGL4.23 (FAC)	3	0.52	
2	140 bp (dH_2O)	1	0.28	0.62
	140 bp (dH_2O)	2	0.32	
	140 bp (dH_2O)	3	0.27	
	140 bp (FAC)	1	0.36	0.74
	140 bp (FAC)	2	0.33	
	140 bp (FAC)	3	0.36	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.57	1.23
	pGL4.23 (FAC)	2	0.62	
	pGL4.23 (FAC)	3	0.52	
3	140 bp (dH_2O)	1	0.28	0.64
	140 bp (dH_2O)	2	0.30	
	140 bp (dH_2O)	3	0.24	
	140 bp (FAC)	1	0.34	0.75
	140 bp (FAC)	2	0.19	
	140 bp (FAC)	3	0.30	
	pGL4.23 (dH_2O)	1	0.39	1.00
	pGL4.23 (dH_2O)	2	0.46	
	pGL4.23 (dH_2O)	3	0.43	
	pGL4.23 (FAC)	1	0.54	1.26
	pGL4.23 (FAC)	2	0.60	
	pGL4.23 (FAC)	3	0.47	

Table S8.4 Normalized luciferase data for *LTF* constructs transfected into COS-1 cell line.

1	2 kb (dH_2O)	1	1.12	1.00
	2 kb (dH_2O)	2	0.96	
	2 kb (dH_2O)	3	0.77 *	
	2 kb (FAC)	1	1.16	1.01
	2 kb (FAC)	2	0.93	
	2 kb (FAC)	3	1.81 *	
	1.8 kb (dH_2O)	1	1.10	0.92
	1.8 kb (dH_2O)	2	0.75 *	
	1.8 kb (dH_2O)	3	1.00	
	1.8 kb (FAC)	1	0.86	0.83
	1.8 kb (FAC)	2	0.86	
	1.8 kb (FAC)	3	0.67 *	
	pGL4.10 (dH_2O)	1	0.00	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	0.77	1.00
	2 kb (dH_2O)	2	0.80	
	2 kb (dH_2O)	3	0.68	
	2 kb (FAC)	1	0.83	1.14
	2 kb (FAC)	2	0.80	
	2 kb (FAC)	3	0.93	
	1.8 kb (dH_2O)	1	0.44	0.53
	1.8 kb (dH_2O)	2	0.32 *	
	1.8 kb (dH_2O)	3	0.40	
	1.8 kb (FAC)	1	0.42	0.51
	1.8 kb (FAC)	2	0.36	
	1.8 kb (FAC)	3	0.37	
	pGL4.10 (dH_2O)	1	0.00 *	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00	
	pGL4.10 (FAC)	3	0.00 *	
3	2 kb (dH_2O)	1	0.90 *	1.00
	2 kb (dH_2O)	2	0.79	
	2 kb (dH_2O)	3	0.77	
	2 kb (FAC)	1	0.75 *	1.15
	2 kb (FAC)	2	0.93	
	2 kb (FAC)	3	0.87	
	1.8 kb (dH_2O)	1	0.51 *	0.52
	1.8 kb (dH_2O)	2	0.41	
	1.8 kb (dH_2O)	3	0.40	
	1.8 kb (FAC)	1	0.46	0.60
	1.8 kb (FAC)	2	0.43	
	1.8 kb (FAC)	3	0.51	
	pGL4.10 (dH_2O)	1	0.00 *	0.00
	pGL4.10 (dH_2O)	2	0.00	
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	0.00
	pGL4.10 (FAC)	2	0.00 *	
	pGL4.10 (FAC)	3	0.00	

Table S9.1 Normalized luciferase data for *TFRC* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	1.15	0.50
	140 bp (dH_2O)	2	1.46	
	140 bp (dH_2O)	3	1.45	
	140 bp (FAC)	1	1.85	0.71
	140 bp (FAC)	2	1.77	
	140 bp (FAC)	3	2.12	
	pGL4.23 (dH_2O)	1	2.76	1.00
	pGL4.23 (dH_2O)	2	2.78	
	pGL4.23 (dH_2O)	3	2.53	
	pGL4.23 (FAC)	1	3.67	1.24
	pGL4.23 (FAC)	2	3.03	
	pGL4.23 (FAC)	3	1.85 *	
2	140 bp (dH_2O)	1	2.93	0.36
	140 bp (dH_2O)	2	3.05	
	140 bp (dH_2O)	3	3.28	
	140 bp (FAC)	1	3.67	0.47
	140 bp (FAC)	2	4.16	
	140 bp (FAC)	3	4.35	
	pGL4.23 (dH_2O)	1	8.55	1.00
	pGL4.23 (dH_2O)	2	8.63	
	pGL4.23 (dH_2O)	3	1118.48 *	
	pGL4.23 (FAC)	1	7.48	0.95
	pGL4.23 (FAC)	2	8.63	
	pGL4.23 (FAC)	3	8.34	
3	140 bp (dH_2O)	1	0.33	0.44
	140 bp (dH_2O)	2	0.37	
	140 bp (dH_2O)	3	0.39	
	140 bp (FAC)	1	0.40 *	0.59
	140 bp (FAC)	2	0.46	
	140 bp (FAC)	3	0.46	
	pGL4.23 (dH_2O)	1	0.74	1.00
	pGL4.23 (dH_2O)	2	0.83	
	pGL4.23 (dH_2O)	3	0.79	
	pGL4.23 (FAC)	1	0.93	1.17
	pGL4.23 (FAC)	2	0.44 *	
	pGL4.23 (FAC)	3	0.91	

Table S9.2 Normalized luciferase data for *TFRC* constructs transfected into HepG2 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	1561.48 *	
	2 kb (dH_2O)	2	1959.32	1.00
	2 kb (dH_2O)	3	2220.00	
	2 kb (FAC)	1	2402.91	
	2 kb (FAC)	2	2358.86	1.14
	2 kb (FAC)	3	2396.49	
	1.8 kb (dH_2O)	1	3394.73	
	1.8 kb (dH_2O)	2	3869.85	1.74
	1.8 kb (dH_2O)	3	4578.00 *	
	1.8 kb (FAC)	1	3950.74	
	1.8 kb (FAC)	2	4080.39	1.90
	1.8 kb (FAC)	3	3869.76	
	pGL4.10 (dH_2O)	1	0.01	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.01	
2	2 kb (dH_2O)	1	2383.80	
	2 kb (dH_2O)	2	2605.69	1.00
	2 kb (dH_2O)	3	2345.92	
	2 kb (FAC)	1	2752.55	
	2 kb (FAC)	2	2836.53	1.17
	2 kb (FAC)	3	3026.49	
	1.8 kb (dH_2O)	1	3369.28	
	1.8 kb (dH_2O)	2	2711.79 *	1.43
	1.8 kb (dH_2O)	3	3631.84	
	1.8 kb (FAC)	1	3957.70	
	1.8 kb (FAC)	2	3234.97	1.47
	1.8 kb (FAC)	3	3610.48	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01 *	0.00
	pGL4.10 (FAC)	3	0.01	
3	2 kb (dH_2O)	1	3052.37	
	2 kb (dH_2O)	2	3780.56 *	1.00
	2 kb (dH_2O)	3	3003.57	
	2 kb (FAC)	1	3591.05	
	2 kb (FAC)	2	3127.73	1.15
	2 kb (FAC)	3	3683.32	
	1.8 kb (dH_2O)	1	3712.74	
	1.8 kb (dH_2O)	2	3420.48	1.18
	1.8 kb (dH_2O)	3	3593.06	
	1.8 kb (FAC)	1	3643.50	
	1.8 kb (FAC)	2	3720.85	1.18
	1.8 kb (FAC)	3	3366.18	
	pGL4.10 (dH_2O)	1	0.01 *	
	pGL4.10 (dH_2O)	2	0.01	0.00
	pGL4.10 (dH_2O)	3	0.01	
	pGL4.10 (FAC)	1	0.01	
	pGL4.10 (FAC)	2	0.01	0.00
	pGL4.10 (FAC)	3	0.01	

Table S9.3 Normalized luciferase data for *TFRC* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	140 bp (dH_2O)	1	0.60 *	0.98
	140 bp (dH_2O)	2	0.39	
	140 bp (dH_2O)	3	0.42	
	140 bp (FAC)	1	0.50	1.24
	140 bp (FAC)	2	0.48	
	140 bp (FAC)	3	0.56	
	pGL4.23 (dH_2O)	1	0.40	1.00
	pGL4.23 (dH_2O)	2	0.42	
	pGL4.23 (dH_2O)	3	0.51 *	
	pGL4.23 (FAC)	1	0.48	1.08
	pGL4.23 (FAC)	2	0.43	
	pGL4.23 (FAC)	3	0.42	
2	140 bp (dH_2O)	1	0.41	0.94
	140 bp (dH_2O)	2	0.49	
	140 bp (dH_2O)	3	0.42	
	140 bp (FAC)	1	0.54	1.18
	140 bp (FAC)	2	0.56	
	140 bp (FAC)	3	1.05 *	
	pGL4.23 (dH_2O)	1	0.49	1.00
	pGL4.23 (dH_2O)	2	0.49	
	pGL4.23 (dH_2O)	3	0.42	
	pGL4.23 (FAC)	1	0.47	1.08
	pGL4.23 (FAC)	2	0.52	
	pGL4.23 (FAC)	3	0.52	
3	140 bp (dH_2O)	1	0.37	0.86
	140 bp (dH_2O)	2	0.36	
	140 bp (dH_2O)	3	0.28 *	
	140 bp (FAC)	1	0.49	1.10
	140 bp (FAC)	2	0.49	
	140 bp (FAC)	3	0.42	
	pGL4.23 (dH_2O)	1	0.39	1.00
	pGL4.23 (dH_2O)	2	0.46	
	pGL4.23 (dH_2O)	3	0.43	
	pGL4.23 (FAC)	1	0.44	1.03
	pGL4.23 (FAC)	2	0.40	
	pGL4.23 (FAC)	3	0.47	

Table S9.4 Normalized luciferase data for *TFRC* constructs transfected into COS-1 cell line.

Experiment	Construct	Repeat	Normalized Value	Fold Change
1	2 kb (dH_2O)	1	418.77	
	2 kb (dH_2O)	2	435.12	1.00
	2 kb (dH_2O)	3	495.22	
	2 kb (FAC)	1	575.02	
	2 kb (FAC)	2	608.51	1.34
	2 kb (FAC)	3	627.90	
	1.8 kb (dH_2O)	1	726.82	
	1.8 kb (dH_2O)	2	756.36	1.65
	1.8 kb (dH_2O)	3	736.42	
	1.8 kb (FAC)	1	880.86	
	1.8 kb (FAC)	2	808.45	1.85
	1.8 kb (FAC)	3	801.80	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	
2	2 kb (dH_2O)	1	619.53	
	2 kb (dH_2O)	2	572.35	1.00
	2 kb (dH_2O)	3	615.83	
	2 kb (FAC)	1	719.25	
	2 kb (FAC)	2	744.23	1.20
	2 kb (FAC)	3	708.99	
	1.8 kb (dH_2O)	1	787.37	
	1.8 kb (dH_2O)	2	806.31	1.34
	1.8 kb (dH_2O)	3	835.16	
	1.8 kb (FAC)	1	831.51	
	1.8 kb (FAC)	2	899.91	1.42
	1.8 kb (FAC)	3	841.06	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	
3	2 kb (dH_2O)	1	658.70	
	2 kb (dH_2O)	2	655.34	1.00
	2 kb (dH_2O)	3	624.76	
	2 kb (FAC)	1	699.74	
	2 kb (FAC)	2	683.25	1.06
	2 kb (FAC)	3	666.69	
	1.8 kb (dH_2O)	1	740.93	
	1.8 kb (dH_2O)	2	763.77	1.18
	1.8 kb (dH_2O)	3	777.99	
	1.8 kb (FAC)	1	766.19	
	1.8 kb (FAC)	2	755.45	1.19
	1.8 kb (FAC)	3	781.03	
	pGL4.10 (dH_2O)	1	0.00	
	pGL4.10 (dH_2O)	2	0.00 *	0.00
	pGL4.10 (dH_2O)	3	0.00	
	pGL4.10 (FAC)	1	0.00	
	pGL4.10 (FAC)	2	0.00	0.00
	pGL4.10 (FAC)	3	0.00	