Analysis of Genetic Variants in the 5' Regulatory Region of the *ALAS1* Gene in South African Patients with Variegate Porphyria (VP)

By

Nelita du Plessis

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science (MSc) in Genetics at the University of Stellenbosch.

Supervisor: Prof L Warnich

Co-Supervisor: Dr MG Zaahl

March 2007

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature: _____ Date: _____

ABSTRACT

The porphyrias are a group of genetic disorders arising from mutations in either one of the final seven genes encoding the haeme synthesis enzymes. These disease-causing mutations lead to an enzyme deficiency that disrupts normal haeme production, resulting in clinical features due to the subsequent accumulation of porphyrin precursors. Like most of the porphyrias, variegate porphyria (VP) is characterized by high inter- and intra- familial clinical variability, with no apparent genotype-phenotype correlation. The delta-aminolevulinate synthase-1 gene (*ALAS1*) is an apparent candidate gene to explain the variable clinical expression observed in VP, since it encodes the first and rate-determining enzyme of haeme synthesis. Several studies have defined important regulatory elements for the human-, rat- and chicken *ALAS1* gene that regulate expression patterns of this gene. It was hypothesized that in VP individuals, variants within/near critical regulatory sites might alter the transcription rate of this gene, and consequently increase/decrease the amount of haeme precursors accumulating as a result of the defective haeme synthesis enzyme.

The aim of this study was to identify genetic variants that could influence gene expression in the proximal promoter area of the *ALAS1* gene, as well as the two *ALAS1*-drug responsive enhancer sequences (ADRES) located further upstream. DNA (2133 bp per patient) of 19 clinically defined VP patients was analysed by polymerase chain reaction (PCR) and semi-automated DNA sequencing. Subsequently, *in silico* analyses using appropriate software programs, and *in vitro* studies using the luciferase reporter system, were performed to investigate the functionality of the identified variants on *ALAS1* gene transcription.

Two novel single nucleotide polymorphisms (SNPs), a C>T transition at nucleotide -853 and a T>A transversion at nucleotide position -1253, were identified in three VP patients from a single family. No correlation between the sequence variants and clinical symptoms could be observed, suggesting that these two variants are not definitive factors in determining the variable expression of VP. However, due to the small sample size of the VP cohort used, the possibility of variation in *ALAS1* gene expression contributing to the complex process of phenotypic presentation in VP and other porphyrias could not be excluded.

In silico analysis revealed that the -853C/T variant is located within three bp from a potential half-palindromic estrogen response element (ERE), the estrogen receptor (ERα) binding site, while the -1253T/A variant is located 3' to a putative non-consensus stimulatory protein-1 (Sp1) site. For *in vitro* analyses, *ALAS1* promoter mutation constructs were generated and co-transfection experiments performed with an ERα-expression vector in HepG2 cells. Results revealed that the *ALAS1* wild-type promoter is transcriptionally up-regulated in response to

estrogen stimulation and that ERα is required for the observed enhanced activity. Co-transfections with the promoter mutation constructs indicated that the -853T allele amplifies the effect of estrogen, whereas the -1253A allele reduced transcriptional activity irrespective of the presence of estrogen.

This study provides evidence for a novel mechanism of haeme regulation via *ALAS1* in response to estrogen. These *in vivo* and *in silico* findings may also contribute to a better understanding of the variable sympomatic expression of the porphyrias, such as VP, with possible alleviating and aggravating effects mediated by the -1253A and -853T variants respectively.

OPSOMMING

Die porfirieë is `n groep genetiese siektes wat ontstaan het as gevolg van mutasies in enige een van die finale sewe gene wat kodeer vir die heemsintese ensieme. Hierdie siekte-veroorsakende mutasies gee aanleiding tot `n ensiem tekort wat normale heemsintese ontwrig en kliniese simptome veroorsaak as gevolg van `n opeenhoping van heemvoorlopers. Soos meeste ander porfirieë, word variegate porfirie (VP) gekenmerk deur aansienlike inter- en intra-familiële kliniese verskille, met geen ooglopende genotipe-fenotipe korrelasie nie. Die delta-aminolevuliniese suur sintase-I geen (*ALAS1*) is `n potensiële kandidaat geen wat die variërende kliniese ekspressie van VP kan verduidelik, aangesien dit vir die eerste en tempo-bepalende ensiem van die heempadweg kodeer. `n Aantal studies het belangrike regulatoriese elemente, wat die uitdrukkings-patrone van *ALAS1* reguleer, in die mens-, rot- en hoender *ALAS1* geen beskryf. Daar word dus veronderstel dat in VP individue, variante in/naby kritiese regulatoriese areas die transkripsie vlakke van hierdie geen sal verander, en vervolgens die aantal heemvoorlopers, wat ophoop as gevolg van die defektiewe heemsintese ensiem, sal verhoog/verlaag.

Die doelwit van die huidige studie was om genetiese variante wat geen uitdrukking beïnvloed, in die proksimale promotor area van die *ALAS1* geen, asook in die twee *ALAS1*-middel reagerend-verhogende volgordes (ADRES) verder stroomop, te identifiseer. Die DNS (2133 bp per pasiënt) van 19 klinies gedefinieerde VP pasiënte is geanaliseer deur polimerase ketting reaksie (PKR) en semi-outomatiese DNS volgorde bepaling. Gevolglik is *in silico* analises, deur gebruik te maak van toepaslike sagteware programme, en *in vitro* analises met die lusiferase verklikker sisteem uitgevoer, om sodoende die funksionaliteit van hierdie variante op *ALAS1* transkripsie te ondersoek.

Twee nuwe puntmutasies is geïdentifiseer, naamlik, `n C>T transisie by nukleotied posisie -853, en `n T>A transversie by nukleotied posisie -1253, in drie VP pasiënte wat afkomstig is van dieselfde familie. Geen verband kon gevind word tussen die betrokke variante en simptome nie, wat `n aanduiding is dat hierdie variante nie `n bepalende faktor is in die variërende uitdrukking van VP nie. As gevolg van die beperkte aantal VP pasiënt-monsters wat beskikbaar was, kan die moontlikheid dat variasie in *ALAS1* geenuitdrukking bydra tot die komplekse proses van die fenotipiese uitdrukking in VP en ander porfirieë, egter nie uitgeskakel word nie.

In silico analise het aangedui dat die -853C/T variant binne drie bp van `n half-palindromiese estrogeen reaktiewe element (ERE), die bindings setel vir die estrogeen reseptor (ER$\alpha$), geposisioneer is, terwyl die -1253T/A variant aan die 3' kant van `n moontlike nie-konsensus stimulerende proteïen-1 (Sp1) setel geleë is. Vir *in vitro* analises is *ALAS1* promotor-mutasie

konstrukte geskep, en ko-transfeksie eksperimente uitgevoer met `n ERα-ekspressie vektor in HepG2 selle. Resultate het aangedui dat die *ALAS1* wilde-tipe promoter, in reaksie tot estrogeen, transkripsioneel verhoog word, en dat ERα noodsaaklik is vir die waargenome verhoging in aktiwiteit. Ko-transfeksies met die promoter-mutasie konstrukte het getoon dat die -853T alleel die effek van estrogeen amplifiseer, terwyl die -1253A alleel transkripsie verlaag, ongeag van die teenwoordigheid van estrogeen.

Hierdie studie verskaf bewyse vir `n nuwe meganisme van heem regulering, via *ALAS1* in reaksie op estrogeen. Die *in vivo* en *in silico* bevindinge mag ook moontlik bydra tot `n beter begrip van die variërende simptomatiese uitdrukking van die porfirieë, soos VP, met moontlike verbeterde en verswarende effekte veroorsaak deur onderskeidelik die -1253A en -853T variante.

ACKNOWLEDGEMENTS

I would like to express appreciation towards the following individuals and institutions:

The MRC and the NRF for the funding of this study.

The University of Stellenbosch and the Department of Genetics for providing the infrastructure and facilities to complete this study.

My supervisor, Prof L Warnich, for giving me the opportunity to carry out my research in her research laboratory, for her encouragement, the excellent and critical scientific guidance and for reading various drafts of this thesis.

My co-supervisor, Dr MG Zaahl, for teaching me tissue culture- and luciferase reporter techniques.

Nicky Verhoog from the Biochemistry Department at the University of Stellenbosch for advice on transfection experiments.

Dr MG Zaahl and Dr KJH Robson (Oxford University) for providing the pSVβ-gal plasmid and HepG2 cell line.

Dr A Louw for the pCDNA/ER expression vector and interesting discussions.

Mrs A Sadie and Mr W Botes for help with statistical analyses.

Ms L van der Merwe and my other colleagues in lab 231 for their enthusiasm and positive spirit.

My mentor, Mr M Kimberg, for his support and help throughout this study. Thanks for your patience, for keeping me positive during difficult times and for always believing in me.

My friends, Ferdi and the others, for your support.

My father, mother and brother, for your interest and emotional support during this study. Without your love and advice I would never have come this far in life.

To my heavenly Father, in whom I find my strength, providing me with courage, endurance and the ability to complete this study.

TABLE OF CONTENTS

List of Figures

CHAPTER 2: LITERATURE REVIEW

CHAPTER 3:

CHAPTER 4: RESULTS AND DISCUSSION

List of Tables

CHAPTER 2: LITERATURE REVIEW

CHAPTER 3: MATERIALS AND METHODS

CHAPTER 4: RESULTS AND DISCUSSION

List of Abbreviations and Symbols

α	alpha
~	approximately
β-Gal	beta-galactosidase
χ^2	chi-square
°C	degrees celsius
δ	delta
µ	micro (10^{-6})
µg	microgram
µl	microlitre
µM	micro molar
%	percentage
®	registered trademark
™	trademark
3’	3-prime
5’	5-prime
A	adenine
AA	acrylamide; C_3H_5NO
AclI	*Acinetobacter calcoaceticus* M4 (S.K. Degtyarev), 1st enzyme
ADRES	*ALAS1*-drug-responsive enhancer sequences
ALA	aminolevulinic acid
ALAS1	delta-aminolevulinate synthase-1 gene
ALAS2	delta-aminolevulinate synthase-2 gene
Amp	ampicillin
ANOVA	Analysis of Variance
AP1	activator protein 1
APS	ammonium persulphate: $(NH_4)_2S_2O_8$
ATCC	American Type Culture Collection
BAA	bis-acrylamide; N,N’-methylene-bis-acrylamide: $C_7H_{10}O_2N_2$
Bgl II	*Bacillus globigii*; 2nd enzyme
bp	base pair
BSA	bovine serum albumin
BsaI	*Bacillus stearothermophilus* 6-55 (Z. Chen), 1st enzyme
C	cytosine
$CaCl_2$	calcium chloride
CAR	constitutive androstane receptor
cDNA	complementary deoxyribonucleic acid

CFTR	transmembrane conductance regulator gene
ChIP	Chromatin immunoprecipitation
cis	a regulatory element located on the same strand
CNSs	conserved non-coding sequences
CO_2	carbon dioxide
COUP-TF	chicken ovalbumin upstream promoter transcription factor
CPO	coproporphyrinogen oxidase
CXR	chicken xenobiotic receptor
CYP	cytochrome P450 superfamily
dATP	2'-deoxyadenosine-5'-triphosphate
dH_2O	distilled water
DMEM	Dulbecco's Modified Eagles Medium
DNA	deoxyribonucleic acid
dNTP	2'-deoxy-nucleotide-5'-triphosphate
DR4	direct repeat 4
DR5	direct repreat 5
DTT	dithiothreitol
E. coli	*Escherichia coli*
E_2	17β-estradiol: $C_{18}H_{24}O_2$
EDTA	ethylenediaminetetraacetic acid: $C_{10}H_{16}N_2O_8$
EMSA	electrophoretic mobility shift assay
EPI	European Porphyria Initiative
ERα	estrogen receptor-alpha
ERβ	estrogen receptor-beta
ER(s)	estrogen receptor(s)
ERE(s)	estrogen responsive element(s)
et al.	and others
EtBr	ethidium bromide: $C_{21}H_{20}BrN_3$
EtOH	ethanol: CH_3CH_2OH
F	forward primer
FCS	fetal calf serum
Fe^{2+}	ferrous iron
FECH	ferrochelatase
fSNP(s)	functional single nucleotide polymorphism(s)
g	gram
G	guanine
gDNA	genomic deoxyribonucleic acid
h	hours

HCl	hydrochloric acid
HepG2	human hepatocellular carcinoma
HMOX1	heme oxygenase-1 gene
HNF-3	hepatic nuclear factor-3
i.e.	that is
i*n silico*	refers to research conducted with computers only
in vitro	Latin phrase for “in a test tube”
in vivo	Latin phrase for “in the living organism”
Inr	initiator element
IPTG	isopropyl β-D-thiogalactopyranoside
IRE	insulin-responsive element
KpnI	*Klebsiella pneumoniae* OK8, 1st enzyme
kb	kilo base
L	litre
LB	Luria-Bertani medium
Luc	luciferase
m	milli (10^{-3})
M	moles per litre/ molar
MCS	multiple cloning site
mg	milligram
$MgCl_2$	magnesium chloride
ml	millilitre
mM	millimoles per litre/ millimolar
mRNA	messenger ribonucleic acid
MW	molecular weight
n	nano (10^{-9})
NaCl	sodium chloride
NaOH	sodium hydroxide
NFκB	nuclear factor kappa-beta
ng	nanogram
ng/µl	nanogram per microlitre
NH_4Oac	ammonium acetate
NheI	*Neisseria mucosa heidelbergensis*, 1st enzyme
NR	nuclear receptors
NRF-1	nuclear respiratory factor-1
nt	nucleotide
OD	optical density
p	short arm of chromosome

p	probability
PB	phenobarbital
PBG	porphobilinogen
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PGC-1α	proliferator-activated coactivator 1-alpha
PIA	propylisopropylacetamide
PKA	protein kinase A
PPIX	protoporphyrinogen
PPOX	protoporphyrinogen oxidase
pmol	pico mole
PWMs	position weight matrices
PXR	pregnane X receptor
R	reverse primer
RACE	rapid amplification of cDNA ends
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SAP	shrimp alkaline phosphatase
S.D	standard deviation
S.E	standard error
SNP(s)	single nucleotide polymorphism(s)
Sp1	stimulatory protein 1
T	thymine
Taq	*Thermus aquaticus*
TBE	tris-borate-EDTA buffer
TBP	TATA-binding protein
TEMED	N,N,N',N'-tetramethylethylenediamine
Temp	temperature
TF(s)	transcription factor(s)
TFBS(s)	transcription factor binding site(s)
Tm	melting temperature
trans	a regulatory element located on another molecule
Tris-HCl	tris hydrochloride
TSS(s)	transcriptional start site(s)
U	enzyme activity unit
UPG-III	uroporphyrinogen-III
US	University of Stellenbosch

UTR	untranslated region
UV	ultraviolet
V	volt
v/v	volume per volume
VP	variegate porphyria
vs.	versus
WIMM	Weatherall Institute of Molecular Medicine
w/v	weight per volume
wt	wild-type [nucleotide(s) as indicated by the reference sequence used]
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside: $C_{14}H_{15}Br\ Cl\ N\ O_6$
x g	times gravity

CHAPTER 1:

INTRODUCTION AND OVERVIEW

1. INTRODUCTION AND OVERVIEW

In 1909, the Danish botanist Wilhelm Ludwig Johannsen formulated the word *genes* as the units of heredity and shortly afterwards he established the distinction between genotype and phenotype (Johannsen 1909; Wanscher 1975). Today, a century later, continued effort is still being devoted to the gathering of data that will relate the events occurring at molecular level (Pearson 2006), with the clinical results seen in patients with heritable diseases.

Several lines of evidence indicate that the correlation between genotype and phenotype is incomplete and that it is rather the complete subset of all the mutations present, that predicts phenotype (Dipple and McCabe 2000; Knight 2003). In complex diseases it is the interplay between numerous genetic loci, genetic susceptibility loci, quantitative trait loci and environmental factors, that seem to contribute to the variable disease outcome. It is now evident that even in monogenic diseases, where the causative mutation is identified, relatives can differ in phenotypic disease severity (Weatherall 2001). Clearly, additional genetic or environmental factors must be involved. As a result, it is apparent why one of the firmly held theories of human genetics is that only the complete comprehension of the cumulative effect of all these disease modifiers, will provide the capability to more efficiently correlate genotype with phenotype. Understanding the multiple interactions underlying the phenotype of disorders is, however, a difficult task since each genetic locus and environmental factor may contribute only a small part.

The genetic components of these disease modifiers are variations in our DNA, also termed modifier loci (discussed in section 2.2.2). These modifier loci can occur in *cis* (on the same allele) or in *trans* (on another allele or gene) to the disease-causing locus and may alter the protein production and/or functionality of the disease gene or disease pathway. Such alterations may go unnoticed in healthy individuals, but may influence the disease phenotype in conjunction with the disease causing mutation. Classic examples include variants of the thalassaemias (Antonarakis *et al.* 1984), hypercholesterolemia (Koivisto *et al.* 1994), Alzheimer's disease (Kamboh *et al*. 1995; 2006), Hirschsprung's disease (Fitze *et al.* 2002; de Pontual *et al*. 2006), Fragile X syndrome (Garber *et al.* 2006), autosomal dominant polycystic kidney disease (Devuyst *et al*. 2003), Haemochromatosis (Jacolot *et al*. 2004), Cystic fibrosis (Drumm *et al*. 2005), heart failure (Le Corvoisier *et al.* 2003) and deafness (Yan *et al*. 2006).

1.1 BACKGROUND

A good example of a group of diseases where the genotype frequently does not reflect the phenotype, is the porphyrias. A classic model of a porphyria with numerous additive factors

contributing to the phenotypic outcome, is variegate porphyria (VP). VP is an autosomal dominant genetic condition, caused by the inactivation of one allele of the protoporphyrinogen-IX oxidase gene (*PPOX*, EC 1.3.3.4) that encodes for the seventh enzyme in the haeme biosynthetic pathway (discussed in section 2.1). VP is a rare disease globally, but occurs at a high incidence in South Africa. This is due to the founder effect of the R59W mutation in exon 3 of *PPOX*, present in more than 90% of South African VP patients (Meissner *et al*. 1996, Warnich *et al*. 1996b).

Although a single mutation in *PPOX* may cause VP, incomplete penetrance and variable clinical expression are recurrently observed between individuals harbouring the identical *PPOX* mutation (Hift *et al.* 2004a; Hift 2005; Hift and Meissner 2005). The typical clinical appearance of VP include skin symptoms and/or acute attacks that are proposed to be directly attributable to the build-up of neurotoxic haeme precursors (Kauppinen 2005). Although only a handful of individuals with PPOX mutations ever develop symptoms, the heterogeneous disease expression causes even asymptomatic patients to follow severe restrictions with regard to their daily life and medication. These restrictions include the avoidance or monitoring of precipitating factors such as alcohol, sun exposure, smoking, and physiological changes such as infections, emotional stress, hormones and certain medications. A hormonal influence on clinical expression is clearly visible in the high prevalence of acute attacks in women and the presentation of symptoms in both sexes primarily after puberty (Hift *et al.* 1997; 2004a; Anderson *et al*. 2005). The influence of these environmental/exogenous factors surely explains some aspects of this phenotypic heterogeneity, but is not sufficient to explain the low penetrance of a dominant mutation.

To date, no genotype-phenotype correlation has been made that may provide insight into the inconsistent disease expression observed in VP (Kirsch *et al*. 1998). The high occurrence of VP and the prevalence of the R59W mutation make South African VP patients an ideal cohort to study the consequence of other genetic factors in conjunction with R59W, which may possibly influence the severity or outcome of the disease symptoms. Assessing the involvement of modifier factors in the pathogenesis of a disease, particularly when the contribution of other factors is suspected, requires the molecular analysis of a wide region covering both the coding- and non-coding sequences of the gene of interest.

Investigations into possible modifying variants within the exon- and intron regions of *PPOX* itself revealed an abundance of variants. However, none of these could explain the variance in disease expression observed in VP (Whatley *et al.* 1999; Warnich *et al*. 2002a). Studies investigating the promoter region of *PPOX* have been performed to determine whether these variants could have an effect on gene expression, and possibly explain differences in disease expression. To date, none of these studies provided substantial evidence for such modifying loci

in *PPOX* (Warnich *et al*. 2002a, Gouya *et al.* 2004). These studies proposed that the contributions of other candidate trans-acting loci should be investigated.

Since it has been established that disease phenotype may be influenced by variations distinct from the disease-causing locus, the influence of other gene/s that act *via* the same biological and molecular pathway as *PPOX*, appear to be an attractive possibility (Badminton and Elder 2005). These polymorphisms will not cause porphyria and may be tolerated in individuals without *PPOX* mutations, but they may alter haeme homeostasis. This will consequently influence the downstream steps in the haeme pathway and generate an augmented build-up of haeme precursors in the presence of an already defective PPOX enzyme. For this reason, subsequent investigations performed mutation analysis of the uroporphyrinogen III synthase (*UROS*) and the haemochromatosis (*HFE)* genes, in order to explain the severe phenotype of an R59W heterozygous South African patient. However, no disease-modifying alterations could be observed in this patient (de Villiers *et al*. 2005).

Another obvious candidate gene was the first and rate-determining gene of the haeme synthesis pathway, δ-aminolevulinate synthase-*1 (ALAS1)*. A previous study conducted by our research group investigated the existence of modifying variants within the coding- and splice-site regions of *ALAS1*, but only one variant was detected and no correlation was observed between this previously identified silent mutation (4713C>T) and the variable VP symptoms (Steyn 2002). In the current study, focus was directed to the 5' regulatory region of *ALAS1*, that has the potential to modify haeme homeostasis.

1.2 HYPOTHESIS

Given the fact that variations in regulatory regions of genes can have a functional effect by influencing gene expression, the hypothesis for the current project was formulated as follows:
Sequence variants in the 5' non-coding region of *ALAS1*, which alter the regulation of gene expression under drug induced or basal conditions, will alter the quantity of the enzyme produced. This event will influence the rate of haeme synthesis and in conjunction with a defective *PPOX* gene, may contribute to the phenotypic outcome of a R59W positive individual, by causing an altered amount of haeme precursors. Since *ALAS1* is a strong candidate for such a mechanism of phenotypic presentation, this study aims to identify genetic variants in the *ALAS1* 5' region, to evaluate their influence on gene expression and investigate their association with VP clinical expression. This study is of clinical interest since early diagnosis of the VP status of an individual, together with the existence of modifier loci, could calculate the potential to develop symptoms and also the level of phenotypic expression, contributing to an improved disease management and genetic counselling. These findings may contribute to the achievement

of a major goal in modern biological research: the improved prediction, understanding and assessment of the clinical symptoms of not only VP, the other porphyrias and disorders related to the haeme pathway, but may also provide information applicable to other complex diseases.

1.3 SPECIFIC AIMS

1) Amplify 1.3 kb of the *ALAS1* gene promoter and the upstream *ALAS1*-drug responsive enhancer sequence (ADRES) elements of VP patients by polymerase chain reaction (PCR) amplification.
2) Analyse the amplified genomic regions for DNA sequence variants by DNA sequencing analysis.
3) Determine the allelic frequencies of the identified sequence variants in a population matched control group, and investigate the existence of a trend between these variants and a specific clinical phenotype.
4) Examine the isolated genomic regions for putative regulatory elements by bioinformatic analyses and determine the level of conservation of these areas by phylogenetic footprinting.
5) Determine the functionality of each identified variant, under basal- or drug-induced conditions by an *in vitro* functional assay.

CHAPTER 2:

LITERATURE REVIEW

2. LITERATURE REVIEW

2.1 INHERITED DISORDERS OF HAEME SYNTHESIS: THE PORPHYRIAS

In 1874, a German medical student, JH Schultz, described a clinical malaise in his doctoral thesis which he termed “porphyria” (Schultz 1874; With 1980). Today the porphyrias are known as a group of heterogeneous disorders caused by abnormalities in the specific enzymes needed to complete each chemical step of the haeme synthesis pathway (Brodie *et al*. 1977). These enzymatic defects subsequently lead to the accumulation and excessive excretion of haeme precursors (porphyrins and/or their precursors). There exists a vital requirement for haeme by nearly all living cells, as it serves as a prosthetic group to a large number of haeme-containing proteins (haemoproteins), responsible for important and diverse biological processes such as cell respiration, drug metabolism, oxygen transport and cell growth (Tsiftsoglou *et al*. 2006). In mammals the two major body compartments that constantly requires newly synthesized haeme, are the bone marrow and the liver (reviewed in Taketani 2005). Although porphyrin synthesis occurs in virtually all mitochondrial containing cells, haeme synthesis is hence most active in erythropoietic tissues where hemoglobin is produced, and in the hepatic tissues where haeme is used for the synthesis of haemoproteins such as cytochrome (CYP) type of proteins.

Haeme synthesis is performed by eight enzymes (reviewed in: Ponka 1997;1999; Ajioka *et al.* 2006). The rate of haeme synthesis is regulated by the first enzyme in the haeme pathway, delta-aminolevulinic acid synthase (ALAS). This regulatory role is evident in the manner in which the gene coding for the ALAS isozyme primarily in hepatic tissue, *ALAS1*, is distinctly upregulated in response to depleted haeme levels (May *et al.* 1995). The second isozyme (ALAS2) is an erythroid cell-specific enzyme, of which the synthesis is developmentally regulated and is increased during erythropoiesis to meet the demand for haeme during haemoglobin production (May *et al*. 1995; Sadlon *et al.* 1999).

ALAS is responsible for the condensation of glycine and succinyl CoA in the mitochondria, to produce coenzyme A, carbon dioxide, and delta-aminolevulinic acid (ALA) (Jordan and Shemin 1972). ALA is exported from the mitochondria to the cytoplasm where the next four steps of the synthesis process take place (refer to Figure 2.1). At the end of these four cytosolic reactions, uroporphyrinogen-III (UPG-III) enters the mitochondria for the sixth step in haeme synthesis where coproporphyrinogen oxidase (CPO) converts coproporphyrinogen-III to protoporphyrinogen-IX (PPIX). Protoporphyrinogen oxidase (PPOX), located in the inner mitochondrial membrane, is the seventh enzyme and catalyzes the oxidation of PPIX to protoporphyrin-IX. The concluding step is the formation of ferroprotoporphyrin-IX (protohaeme)

by the incorporation of ferrous iron (Fe^{2+}) into protoporphyrin-IX by the last enzyme, ferrochelatase (FECH). Although haeme is linked to various proteins, a small amount remains uncommitted in a "free" haeme pool. This "free" haeme regulates the pathway by inhibiting *ALAS1 via* a negative feedback-loop (discussed in section 2.3.1.2).

Because haeme is such an essential component for normal cellular function, but cytotoxic when produced in excess (Jeney *et al.* 2003), defects in haeme metabolism have far-reaching biochemical and pathological consequences (Moore 1998). With the exception of the first enzyme, defects in any of the other seven enzymes in the haeme pathway, each cause a different type of porphyria. Even though *ALAS1* has not been linked to any disease, a defect in *ALAS2* leads to X-linked sideroblastic anemia (Cotter *et al.* 1992; Edgar *et al.* 1997).

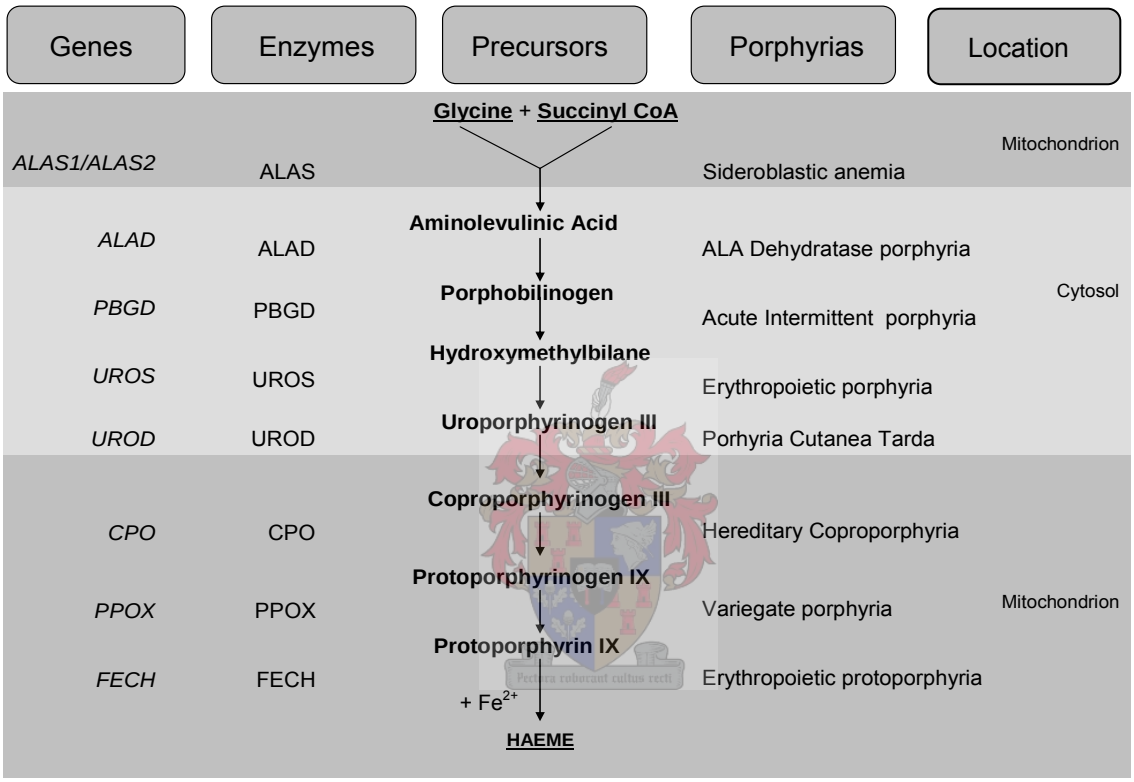
The porphyrias have several different classification systems (James and Hift 2000; Nordmann and Puy 2002). Each type of porphyria can be classified by its specific enzyme deficiency, defined by a unique pattern of accumulation and excretion of haeme precursors, as well as a reduction in the relevant enzyme activity (Figure 2.1). A second, more general, classification system is based on the location where excess precursors originate: primarily in the liver (hepatic porphyrias) or primarily in the bone marrow (erythropoietic porphyrias). Another classification system distinguishes porphyrias that cause neurological symptoms (acute porphyrias) from those that cause photosensitivity (cutaneous porphyrias). Some porphyrias can, however, present as both cutaneous lesions and acute attacks (reviewed in: Gordon 1999). Acute porphyria occurs in all races (Elder *et al*. 1997; Elder 1998), with acute attacks about five times more common in women than in men and more frequent in the second to fourth decades (Elder *et al.* 1997; Liu *et al*. 2005).

2.1.1 Variegate porphyria (VP)

Variegate porphyria (VP, OMIM 176200) is an example of an acute hepatic porphyria with cutaneous symptoms, that occurs due to partial deficiency of the enzyme protoporphyrinogen oxidase (*PPOX*, E.C.1.3.3.4) in the liver. To date, more than 130 sequence variants, in different regions of the *PPOX* gene, have been identified in VP patients world wide [Deybach *et al*. 1996; Meissner *et al*. 1996; Warnich *et al.* 1996a, 1996b; also see online: Human Gene Mutation Database (December 2006): http://www.hgmd.org]. While extensive population heterogeneity is observed in *PPOX* disease mutations (Wiman *et al*. 2003; Schneider-Yin and Minder 2006; Lecha *et al*. 2006) and polymorphisms (Warnich *et al*. 2002b), the R59W mutation in exon 3 of the *PPOX* gene is the primary cause for VP in the South African Caucasian population (Meissner *et al*. 1996; Warnich *et al*. 1996b). As mentioned in section 1.1, this founder mutation leads to a 50% reduced activity of the PPOX enzyme and accounts for the unusually high prevalence of VP in South Africa (Dean *et al*. 1972, Meissner *et al*.

1996; Warnich *et al*. 1996b; Groenewald *et al*. 1998). It is suggested that the homozygous state of the R59W mutation is lethal since it will abolish *PPOX* activity. Rare cases of compound homozygous VP have been detected where the patients carry a disease-causing mutation and a less severe *PPOX* mutation (Hift *et al.* 1993; Frank *et al*. 1998; Corrigal *et al.* 2000; Kauppinen *et al*. 2001).

Figure 2.1 Haeme synthesis and the main porphyria types with their associated defective genes and enzymes.



ALAS: delta aminolevulinic acid synthase. ALAD: delta-aminolevulinate dehydratase. PBGD: porphobilinogen deaminase. UROS: uroporphyrin III synthase. UROD: uroporphyrinogen decarboxylase. CPO: coproporphyrinogen oxidase. PPOX: protoporphyrin oxidase. FECH: ferrochelatase.

2.1.1.1 VP phenotype

VP is classified as a low-penetrance disease with variable clinical expression. Many patients heterozygous for a mutation associated with VP remain clinically and biochemically latent (Hift *et al*. 1997; Kirsch *et al*. 1998; Hift *et al*. 2004a; 2004b). It has been proposed that the diminishing occurrence of acute attacks and skin lesions in VP patients may be attributed to increased diagnosis before disease onset and a greater awareness of precipitating factors (Hift *et al.* 1997; von und zu Fraunberg *et al.* 2002; Hift *et al.* 2004a; 2005). However, even sibs with the same disease causing mutation, often display dissimilar disease phenotypes (Whatley *et al*. 1999; Frank *et al*. 2001). Clinical

manifestation of VP may therefore require environmental and/or genetic factors in conjuction with a single VP causative mutation (Poh-Fitzpatrick 1998; Sassa and Kappas 2000).

Classical VP symptoms are characterized by photosensitivity and acute attacks which can occur separately or concurrently, with symptoms developing over hours or days. Dermatological changes are caused by the photosensitizing properties of accumulated porphyrins, causing the skin to become fragile and easily develop blisters and sores (Day 1986; Kauppinen and Mustajoki 1992; Kauppinen *et al.* 1997). The cutaneous lesions observed in VP patients are triggered by sun-exposure and treated by avoidance of sunlight and attention to skin care.

Acute attacks are typically associated with abdominal pain, constipation, nausea, hypertension and insomnia (Kirsch *et al.* 1998), with extreme cases experiencing heart palpitations, a rapid heartbeat, high blood pressure, anxiety, confusion, seizures and paralysis (Kauppinen and Mustajoki 1992; Thadani *et al.* 2000; Hift *et al.* 2004a). The severity and frequency of acute attacks fluctuate considerably, but usually they do not manifest before puberty (Hift *et al.* 2004a).

Carbohydrate intake, antioxidants and haeme administration in the form of hematin or haeme arginate, during an acute attack has proven to be helpful in alleviating some of the clinical symptoms (Watson *et al*. 1977; Anderson *et al*. 2005). The mechanism through which haeme treatment functions, is by restoring the intracellular haeme pool, thereby reducing the activity of the first and rate-determining enzyme, *ALAS1*, with the subsequent reduction of the porphyrin precursors upstream of the defective enzyme (Mustajoki and Nordmann 1993; Tenhunen and Mustajoki 1998; Fukada *et al*. 2005; Bonkovsky 2005; discussed in section 2.3.1.2). Carbohydrates act through a mechanism known as the “glucose-effect”, also inhibiting *ALAS1* induction through the insulin-responsive element (IRE) in the *ALAS1* gene (Scassa *et al*. 2004).

2.1.1.2 Biochemical profile

Important biochemical characteristics of VP include the increased biliary excretion of coproporphyrinogen and protoporphyrinogen and their corresponding porphyrins (Bloomer *et al*. 1990), which are commonly measured by fecal porphyrin analysis. Depending on the type of porphyria, porphyrin precursors accumulate at specific stages in the haeme pathway during an acute attack. In VP an acute attack is biochemically characterized by ALA and porphobilinogen (PBG) in the urine because of the induction of haeme synthesis and the presence of an impediment in the pathway (Meyer *et al*. 1998). It has been

proposed that these haeme precursors are neurotoxic and are consequently dangerous to the areas of the brain without the blood-brain barrier, as well as to the nervous system (Shanley *et al*. 1975; Lindberg *et al.* 1999). Another indicator to the diagnosis of VP is dark colored urine, due to polymerization of PBG to porphyrins. Increased plasma fluorescence is usually observed in VP and is an important diagnostic tool (Poh-Fitzpatick 1980). However, between attacks, concentrations of urinary PBG and ALA are often normal (Elder 1980).

2.1.1.3 Precipitating agents of acute attacks

Acute attacks are precipitated by various exogenous or endogenous factors. The substances that are known to trigger acute attacks have in common the capicity to initiate the activity of the haeme pathway (De Matteis 1972; Kauppinen and Mustajoki *et al.* 1992, Sassa and Nagai 1996). The presumed mechanism postulates that these substances require hepatic haeme-mediated metabolism that instigates a subsequent increased requirement for haeme for enzymes such as the cytochrome P450 enzymes (P450s). Several mechanisms functioning *via ALAS1*, have indicated to assist in the up-regulation of haeme synthesis during these situations (Hift 2005; discussed in section 2.3.1.2). Porphyrinogenic substances are likely to be drugs such as barbiturates, anaesthetic agents, tranquilizers and a number of antibiotics. As briefly mentioned in section 1.1, alcohol, smoking, infections, fasting and endogenous changes in sex-hormone balance (i.e menstruation or pregnancy) are common causes of acute attacks (Thunnel *et al.* 1992; Kauppinen and Mustajoki 1992). One of the most common precipitating factors is the contraceptive pill due to its high content of sex hormones. It is known to provoke attacks of acute porphyria either alone or by interacting with other factors (Gross *et al*. 1995; Hift 2005).

Several helpful databases have been launched with useful information regarding porphyria in order to assist patients and clinicians in the usage/prescription of medication. These online websites contain general information relating to porphyria and a list of drugs presently considered to be safe or unsafe, e.g. The European Porphyria Initiative (EPI) (http://www.porphyria-europe.com); American Porphyria Foundation (http://www.porphyriafoundation.com) and the Cape Town Porphyria service (http://web.uct.ac.za/depts/porphyria), to name a few.

2.2 DISEASE MODIFIERS

The variable expression seen in VP, such as interfamilial symptom variability, is one example of the increasing number of genetic disorders in which the genotype does not always correlate with the phenotype. This 'phenotype gap' that exists between the mutant DNA sequence of a gene, and a trait, remains largely unanswered.

2.2.1 The genotype-phenotype link

The use of molecular genetic analysis to interpret and predict the phenotype in genetic diseases initially appeared promising. However, for many human diseases, no clear relationship of phenotype to genotype has been demonstrated (reviewed in: Botstein and Risch 2003; Silverman and Mifflin 2005). Subsequently, the classification of genetic disorders as either chromosomal, monogenic, polygenic or multifactorial became problematic (Romeo and McKusick 1994; Dipple and McCabe 2000). Although an organism's genotype is the largest influencing factor in the development of its phenotype, it is not the only one (reviewed in: Jura *et al.* 2006). Instead, there are a myriad of processes involved in the attainment of an organism's phenotype: epigenetic mechanisms, influences from the environment and the complete genetic profile which includes disease-causing variants and modifier loci (reviewed in: Lopez 1998; Oliver 2004).

2.2.2 Modifier Loci

As mentioned above, phenotypic variation can be explained by aspects such as external influences and/or additional genetic factors. These additional genetic influences are universally termed modifier loci. Modifier loci do not only signify the presence of additional *cis*-acting modifying factors (alternative disease alleles); it can also include *trans*-acting factors such as polymorphisms in genes other than the disease gene that influence disease expression, irrespective of environmental influences (Nadeau 2001; Agarwal and Moorchung 2005). These loci are not sufficient to cause a disease, but it may alter the risk of developing the disease phenotype. This modifying effect, caused by the interaction between a disease causing mutation and variant(s) within the disease allele or other gene(s), can be either positive or negative. That is, cause a more extreme- (enhanced), a less extreme- (reduced), a novel-, or a wild-type (normal) phenotype. Genetic variants can cause this modifying effect by changing not only gene expression, but also disease penetrance, dominance and pleiotropy (phenotypic effect of a single-gene mutation) (Nadeau 2001).

When a phenotypic modifier variant is located in a gene other than the disease causing gene, the term modifier gene is commonly used (Nadeau 2001; Nadeau 2003a; 2003b). Recently, the concept of oligogenic diseases with modifier genes has been developed (Agarwal and

Moorchung 2005). Oligogenic disorders are primarily genetic in origin, in contrast to complex diseases which occur because of the complex interaction between various genes and the environment. Oligogenic disorders are based on the concept that a small number of mutant alleles at a small number of loci are required to modify the effect of a dominant gene. Modifier genes are the basis of oligogenic disorders.

The outcome of modifier genes can be described as additive: the possibility of a particular disease phenotype, brought about by sequence variations in alleles at two or more different loci, is the sum of the risk of each individual allele (Slavotinek and Biesecker 2003, and references therein). The phenotypic variation is therefore due to the interaction of the disease gene with other genes, impacting on the final disease presentation. As a result, phenotypic changes also occur when the expression of one gene alters the expression of another gene. Products of modifier genes can affect splicing, transcription, translation, posttranslational processes, and also protein trafficking, degradation and/or secretion (Persu *et al*. 2002; Buchner *et al*. 2003; Robledo *et al*. 2003). The effect of modifying genes could therefore involve changes in the disease causing gene at transcriptional level through to intermediate phenotypes at the molecular and cellular level, or even end-point phenotypes at the organ, system or organismal level (Peltonen *et al*. 2001; reviewed in: Levine and Tjian 2003).

Cystic fibrosis (CF [OMIM 219700]) provides a striking example of a metabolic disease with various modifying loci that alter phenotypic expression. The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene encodes a chloride channel in the membrane of epithelial cells. The absence or abnormal function of this protein leads to the recessively inherited disease CF. Environmental factors and additional variations in this disease gene that alters CF phenotype, have subsequently been identified (Kerem *et al.* 1989; Salvatore *et al.* 2002). However, the discordant phenotype observed in CF siblings argued against a major role of external factors and pointed towards the involvement of additional modifying factors that prompted an abundance of genotype-phenotype correlation studies (Estivill *et al.* 1995; Mercier *et al.* 1995; Hubert *et al.* 1997; de Gracia *et al.* 2006). Copious numbers of studies therefore focused on the effect of non-*CFTR* genetic polymorphisms affecting CF phenotype (Hull 1998; Drumm *et al.* 2001; 2005; Salvatore *et al.* 2002; Grasemann *et al*. 2003; Slieker *et al.* 2005). These examples reinforce the concept that, apart from environmental influences, independent genetic loci can modify a disease phenotype.

2.3 DELTA AMINOLEVULINATE SYNTHASE-1 *(ALAS1)* AS CANDIDATE MODIFIER GENE TO INFLUENCE PORPHYRIA PHENOTYPE

The various clinical problems associated with defective haeme metabolism (e.g. porphyrias, X-linked anemia), together with its tissue specific expression patterns and the fact that intracellular haeme can generate reactive oxygen species which cause DNA breaks and protein degradation, highlight the importance of the firm regulation of the haeme synthesis- and degradation pathways to ensure haeme homeostasis (reviewed in: Ponka 1997; Nastiti *et al*. 2004; Taketani 2005). Understanding the mechanisms influencing haeme synthesis is therefore of clinical interest, particularly in diseases associated with defective haeme metabolism. Haeme synthesis is primarily regulated by factors influencing *ALAS1* expression. Given the important role of *ALAS1* in haeme synthesis, changes in the regulation of *ALAS1* itself, will have monumental effects on haeme synthesis, and subsequently, on the hepatic porphyrias.

2.3.1 *ALAS1*

Delta-aminolevulinic acid (ALAS), also known as 5-aminolevulinic acid, delta-aminolevulinate or 5-amino-4-oxopentanoic acid, is an aminoketo acid with MW = 167.6 (Fukada *et al*. 2005). ALAS is found in a variety of species ranging from protist, bacteria, fungi, plants to mammals. Studies by Duncan *et al*. (1999) showed that the carboxy terminal two-thirds of the unprocessed ALAS shows 40-90% amino acid conservation in protein alignments across all taxa, with ALAS in bacteria the most highly conserved. Phylogenetic trees indicated that a gene duplication event apparantly led to the formation of the ubiquitous- (*ALAS1*) and erythroid (*ALAS2*) form of the gene (Duncan *et al*. 1999 and references therein). *ALAS1* appears to be an attractive candidate gene to influence haeme metabolism due to the fact that it codes for the first and rate-limiting liver-specific enzyme for the synthesis of tetrapyrroles in the form of haeme.

2.3.1.1 Structure and organization

Delta-aminolevulinate synthase is encoded by two separate genes producing two mitochondrial enzymes. The *ALAS1* gene (EC 2.3.1.37) produces the housekeeping isoform of the enzyme which is ubiquitously expressed, with the highest levels occuring in the liver where it catalyzes the first step in haeme synthesis, the C-C bond condensation of glycine and succinyl coenzyme-A with pyridoxal 5'-phosphate (PLP) as cofactor (Ponka 1997). The 17kb *Homo sapiens ALAS1* gene is assigned to chromosome 3p21.1 while the gene for the other isoform, *ALAS2* (EC 2.3.1.37), is mapped to the X-chromosome (Bishop 1990). The expression of *ALAS2* is restricted to erythroid cells and is critical for erythropoiesis. Interestingly, *ALAS2* cannot be compensated for by the expression of

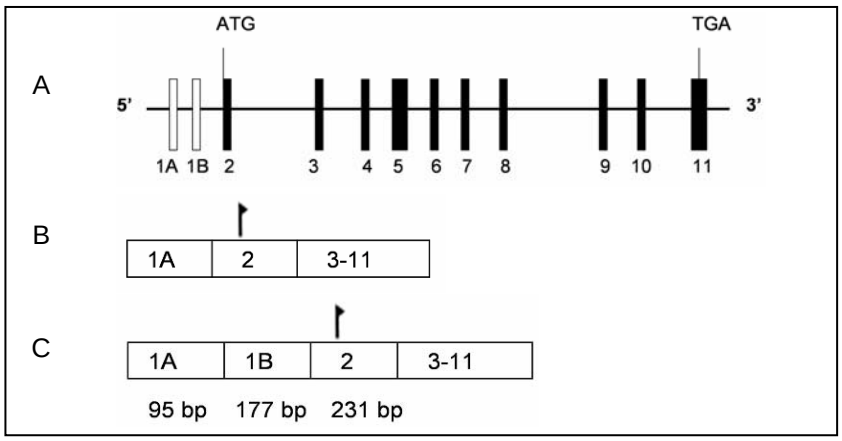
ALAS1, suggesting a very fine method of regulation of ALAS synthesis in a tissue-specific manner.

Dailey *et al.* (2004) illustrated that *ALAS1* and *ALAS2* have the same intron/exon organization. Both ALAS1 and ALAS2 isozymes are described as having three distinct regions: region I the mitochondrial targeting sequence, region II the amino-terminal portion of the mature protein and region III the catalytic region of the enzyme (Dailey *et al.* 1995; Sadlon *et al.* 1999).

Human *ALAS1* consists of 10 coding exons, and two additional non-coding exons (1A and 1B) in the 5'-untranslated region (UTR) which is alternatively spliced to produce a major (exon 1B omitted) and a minor (containing 1A and 1B) mRNA transcript (Figure 2.2). The alternatively spliced transcript was found not to have an important role in the basal expression of *ALAS1* as the amount of the minor transcript was the same in all the tissues tested (Roberts and Elder 2001). In the same study Roberts and Elder (2001) used the 5'-RACE technique to illustrate the presence of multiple TSSs that are used in a tissue specific manner to produce transcripts of different lengths in the liver and brain. The study showed that transcripts in the liver predominantly start from the TSS downstream from the TATA box (-30 bp), numbered nt +1 (Figure 2.3). Their results indicated that in HeLa cells, transcription can be initiated at up to five additional upstream sites (such as a putative initiator-like element) within the core promoter of both the human and rat *ALAS1* genes (Roberts and Elder 2001).

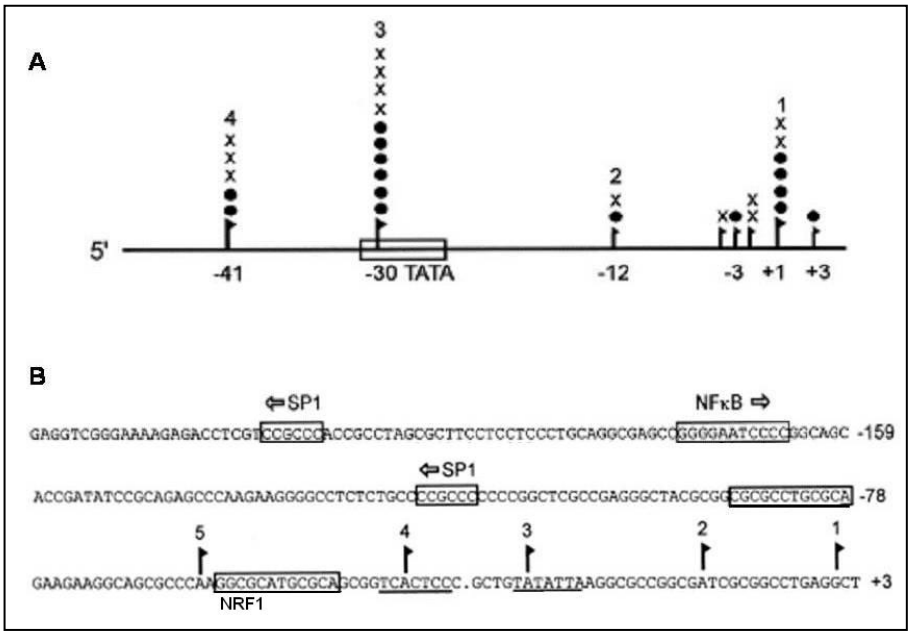
The organization of the orthologous *ALAS1* rat, mouse and chicken genes are similar to human *ALAS1*, except that the rat gene contains one non-coding exon in its 5-UTR (Yomogida *et al.* 1993), while the chicken has none (Maguire *et al.* 1986).

Figure 2.2 Alternatively spiced human *ALAS1* gene produce two mRNA transcripts.



(A) Structural organisation of the human *ALAS1* gene. Shaded rectangles indicate coding exons; unshaded rectangles indicate 5' UTR.
ALAS1 major (B) and (C) minor mRNA transcripts. Positions of translation initiation codons are flagged. (Data obtained from Roberts and Elder 2001).

Figure 2.3 The human *ALAS1* gene contains multiple transcription initiation sites (Adapted from Roberts and Elder 2001).



(A) Illustration of the *ALAS1* core promoter to show transcription start sites (flags) identified by cloning and sequencing 5'RACE products from total RNA from HeLa cells. The TATA element is boxed, start sites are numbered and individual clones are indicated by symbols (●, major transcript; x, minor transcript). (B) Sequences with homology to consensus Sp1, NFκB and NRF1 binding sites are boxed and their directions indicated by arrows. Major TSSs are flagged.

2.3.1.2 Regulation

The mechanisms for the basal and drug-induced regulation of *ALAS1* are currently popular topics under investigation, as the regulation of *ALAS1* activity appears to be more complex than that of the other isoform. Initially it was believed that *ALAS1* enzyme production was exclusively controlled by haeme. Indeed, mammalian *ALAS1* expression in the liver has shown to be under negative feedback control of haeme in various species including rat and human hepatoma cells and chick hepatocytes (Drew *et al.* 1989; Ponka 1999; May *et al.* 1995; Scassa *et al.* 1998; Cable *et al.* 2000). It is inhibited by the regulatory pool of "free-haeme" that influences the ALAS1 isozyme expression by blocking mitochondrial import of the precursor protein (Lathrop and Timko 1993), destabilizing the mRNA (Hamilton *et al.* 1991) and repressing mRNA transcription (May *et al.* 1995). However, reports indicate that the direct inhibition of *ALAS1* transcription by haeme, appears to be unlikey at physiological haeme concentrations (Hamilton *et al.* 1991) and the precise molecular mechanism by which haeme regulates *ALAS1* mRNA levels remains poorly characterized and controversial.

Wolfson *et al.* (1979) suggested that the direct inhibition of *ALAS1 in vivo* by haeme at physiological concentrations appears not to be an important mechanism for the regulation of hepatic haeme biosynthesis, and that *ALAS1* activity does not decrease as the rate of haeme formation increases. Even at a rate of haeme generation at 75 times the rate occurring *in vivo*, *ALAS1* activity remained unchanged (Wolfson *et al.* 1979). Conversely,

when Kolluri *et al.* (2005) investigated the haeme-dependent down regulation of the avian *ALAS1* gene under basal and drug-induced conditions, they identified separate haeme- and drug-reponsive enhancer regions in the promoter region. The haeme-responsive regions were shown to suppress the drug-induced upregulation of *ALAS1* transcription. Experiments with expression constructs containing the human *ALAS1* coding region were performed by Dailey *et al.* (2005) and demonstrated that *ALAS1* possesses haeme-responsive elements close to the amino-terminus and in the leader sequence of the mature protein that inhibit translocation into the mitochondrion. None of the other haeme synthesis pathway proteins examined, were sensitive to the same levels of haeme as was *ALAS1*. These results demonstrate that at a physiologically relevant concentration of haeme, only *ALAS1* translocation is important in regulation of porhyrin synthesis by haeme in non-erythroid cells (Dailey *et al.* 2005). Many other reports verified that *ALAS1* is indeed under negative feedback control of haeme, and that the primary method is by decreasing mRNA stability and blocking mitochondrial import (Cable *et al.* 2000; Munakata *et al*. 2004). In a recent study it was established that the alternatively spliced exon 1B in the 5' UTR of the major *ALAS1* mRNA transcript renders it more resistant to haeme mediated decay than the minor mRNA transcript containing the exon 1B. It was found that the altered 5' UTR of the major transcript inhibits translation, and that the haeme-mediated destabilization requires *ALAS1* mRNA translation (Roberts *et al.* 2005).

Roberts and Elder (2001) also demonstrated the previously unrecognised complexity in the transcriptional regulation of the mammalian *ALAS1* gene. Their analyses of mRNA transcripts in human and rat tissues revealed tissue-specific differences in the use of TSSs by similar *ALAS1* core promoters. They found that the TATA element is dominant and drives transcription from a single downstream site, while a mutation in the TATA box leaves transcription unimpaired because of activation of the upstream TSSs. Their findings indicate that the conformation of the core *ALAS1* promoter that directs assembly of the transcription pre-initiation complex, may vary between tissues and have implications for the tissue-specific regulated expression of this gene. In liver, initiation was TATA-driven from a single downstream site that appeared to be used exclusively for induction by drugs. From these findings they proposed that the induction of *ALAS1* by drugs requires the core promoter to be in a conformation that favours TATA-driven initiation. Additionally, they proposed that variations in the way in which the transcription initiation complex is assembled on the core promoter, may confer a particular promoter conformation, required for activation of *ALAS1* transcription by factors interacting with the more distant motifs that may be involved in *ALAS1* response to barbiturates and related drugs (Roberts and Elder 2001).

Braidotti *et al.* (1993) demonstrated by deletion analysis of promoter constructs that the promoter of rat *ALAS1* is located between nt +1 to -3160 (Braidotti *et al*. 1993). Giono *et al*. (2001) identified binding sites for activator protein-1 (AP-1) (-261 bp) and cAMP-responsive element-binding protein (CREB) (-149 and -145 bp) in the rat *ALAS1* promoter. Also, binding sites for hepatic nuclear factor-3 (HNF3) and nuclear factor-1 (NF1) between –459 bp and –354 bp, which is reported to function in the basal expression of rat *ALAS1*, was decsribed (Scassa *et al.* 2004). Within this region, the human and rat sequences are 63% identical overall and almost completely identical around the initiation start sites. Both contain perfectly matched TATA boxes and elements that are identical or close to the consensus Inr sequence (Roberts and Elder 2001).

Transcription of *ALAS1* is also enhanced by the peroxisome proliferator-activated coactivator 1α (PGC-1α), a co-factor that activates transcription factors (TFs) such as nuclear receptors (NR) (Handschin *et al*. 2005). PGC-1α mediates this effect by interacting with NRF-1 (nuclear respiratory factor 1) and FOX01 (a fork head family transcription factor member) which associate with the *ALAS1* promoter (Virbasius and Scarpulla 1994; Handschin *et al.* 2005). Transcription of PCG-1α is upregulated by a decrease in cellular glucose concentrations (Scassa *et al*. 2001; 2004). Consequently, when cellular glucose levels are low, increased PGC-1α expression is observed, which leads to the transcriptional activation of *ALAS1*. This may create conditions capable of triggering an acute attack of porphyria and provides a possible explanation to the clinical observation that fasting precipitate attacks whereas glucose infusions may attenuate the severity of attacks (Thadani *et al*. 2000; Anderson *et al*. 2005).

Unequivocal evidence indicates that transcription of this housekeeping gene is also directly increased by phorbol esters (Guberman *et al.* 2003) and by drugs which also induce cytochrome P450 proteins (Jover *et al.* 1996). Although exogenous drugs (xenosensors) exert little stimulatory effect on haeme production in the bone marrow, the liver P450 proteins are of critical importance in the oxidative metabolism of drugs (Jover *et al*. 1996). P450s are particularly prominent in the liver where they are induced by their substrates, including numerous drugs, environmental chemicals, hormones, steroids and endogenous/exogenous stimuli (Gonzalez 1988; Nelson *et al.* 1996; Guengerich 2003), resulting in enhanced metabolism and elimination from the body (Honkakoski and Negishi 2000; Podvinec *et al.* 2004). The rate of haeme biosynthesis must be reactive to increased demands, for instance, during the generation of these haeme-containing P450 enzymes and steroids, resulting in the phenomenon that certain types of drugs induce haeme synthesis (Jover *et al.* 1996). Upon exposure to drugs that induce P450s and other drug-metabolizing enzymes, more haeme is utilized for the production of these haeme-containing

enzymes. Generation of P450s therefore necessitates continuous high rates of haeme synthesis, along with the fact that P450s have estimated half-lives of only 90-180 minutes (Bickers 1982). It was illustrated that the continued usage of haeme for the synthesis of an increasing amount of P450s, causes a depletion of the available "free-haeme pool", which in turn stimulates the synthesis of new haeme by de-repressing the haeme-mediated inhibition of *ALAS1* (Granick 1966).

Podvinec *et al.* (2004) inserted a human *ALAS1* promoter fragment, spanning -1249 bp to -1 bp, into the pGL3-Basic reporter vector and assayed gene activity in chicken hepatoma cells (LMH). The promoter fragment exhibited basal promoter activity, but did not cause an increase in transcription in response to prototypic inducers of *ALAS1*. They subsequently used an *in silico* approach to identify sequences further upstream in the *ALAS1* 5' region and described two sequence elements in the 30 kb 5' flanking region of the gene that directly activate human *ALAS1* gene transcription in response to drugs. These two delta-aminolevulinate acid drug-responsive enhancer sequences (ADRES) are located approximately 16 and 20 kb upstream of the transcription initiation site and was tested *in vitro* to be responsive to compounds which also induce hepatic cytochrome P450, such as phenobarbital (PB), metyrapone, propylisopropylacetamide (PIA) and glutethimide (Podvinec *et al.* 2004). They concluded that drug induction of *ALAS1* is due to a direct effect of the inducing compound and not a consequence of the increased synthesis and accumulation of cytochrome P450 (Jover *et al.* 1996; Podvinec *et al.* 2004).

Drug-responsive enhancer elements have also been identified for the 5' flanking regions of the chicken and mouse *ALAS1* gene, and studies determined that the ADRES elements work independently of each other and contain functional direct repeat 4 (DR4) and 5 (DR5) sequences (Fraser *et al*. 2002; 2003). These sequences act as binding sites for NR such as chicken xenobiotic receptor (CXR) in chicken, pregnane X receptor NR1I2 (PXR), and constitutive androstane receptor NR1I3 (CAR) in humans (Fraser *et al.* 2003). These findings propose that the mechanism of gene activation by drugs is the same for cytochrome P450 and *ALAS1* (Podvinec *et al.* 2004), confirming that the same substances that induce *ALAS1* also induce the synthesis of several members of the *CYP* genes to produce P450 enzymes (Nelson *et al.* 1996).

i) Hormonal regulation

The endocrine system exerts a multitude of hormonal effects on haeme metabolism and haeme utilization in the liver (Galbraith 1987). As described above, *ALAS1* is markedly increased by conditions demanding more haeme such as certain drugs and hormones, whereas the expression of the other enzymes in the pathway remains relatively unchanged

(May *et al.* 1995). The presence of these substances increases hepatic haeme by derepression of the negative feedback-loop, that increases haeme synthesis in order to augment P450 levels to accelerate drug biotransformation.

One such an endogenous/exogenous substance metabolised by P450s, is the hormone estrogen. Shou *et al.* (1997) provided support that the CYP1A1 and CYP3A4 isoforms are responsible for the hydroxylation of estrogen. The metabolism of estrogen occurs mainly in the liver where most P450 enzymes are abundantly expressed (Schneider *et al*. 1983; Tsuchiya *et al*. 2005). Early studies reported on induction of the ALAS enzyme upon stimulation with sex steroids and their metabolites (Granick 1966; Granick and Kappas 1967a; 1967b; Kappas and Granick 1968). Tschudy *et al*. (1967) also observed a series of oscillations of the level of ALAS enzyme, instigated by a single intraveneous injection of a minute dose of estrogen. They speculated that estrogen does not directly induce *ALAS1*, but that it is rather as a result of perturbation of the closed negative feedback loop which controls hepatic ALAS synthesis.

Giger and Meyer (1981) were the first to report the inhibitory effect of insulin on *ALAS1* (also refer to section 2.1.1.2). Scassa *et al.* (1998) confirmed this inhibitory effect of insulin in human HepG2 cells and provided evidence that one or more sequences present in the 876-bp promoter region of the rat *ALAS1* gene could be responsible for the insulin inhibitory effect. Their transient transfection experiments with rat *ALAS1* promoter constructs demonstrated that the promoter is involved in the insulin repression of *ALAS1* and that insulin overrides the HNF-3 or HNF-3 plus NF1-mediated stimulation of *ALAS1* transcriptional activity. Their data indicate that these cis-elements are required for effective insulin repression of basal rat *ALAS1* gene expression (Scassa *et al.* 1998). Further studies suggested that the phosphatidylinositol 3-kinase (PI3-kinase) pathway and the Ras/MAPK pathway are jointly required for complete insulin inhibition of *ALAS1* transcription (Scassa *et al.* 2001). Computational analysis identified a putative insulin response element (IRE) located at position -383 to -389 and two consensus hepatocyte nuclear factor-3 (HNF-3) motifs at positions -397 to -387 and -383 to -373 in the promoter region (Scassa *et al.* 2004). They postulate that insulin prevents the binding of HNF-3 to the promoter and allows the binding of a complex that recognizes the IRE. The IRE was found to repress *ALAS1* expression in HepG2 cells, but not in HeLa cells (Scassa *et al.* 2004). It was also found that rat *ALAS1* expression is induced in tissues (especially in the liver) by cAMP (Giono *et al.* 2001).

2.3.1.3 Previously identified polymorphisms

Collective information gathered from single nucleotide polymorphism (SNP) databases from the National Centre for Biotechnology Information (NCBI), Ensembl and HAPMAP indicate that approximately 30 SNPs have been identified in *ALAS1* and the 1.5 kb DNA region flanking the gene at its 3' and 5' ends (NCBI: http://www.ncbi.nlm.nih.gov; Ensembl: http://www.ensembl.org; HAPMAP: http://www.hapmap.org). Only two of these variants are located in coding exons, three in the 5' non-coding region, two in the 3' non-coding region with the rest located in introns and non-coding exons. Only one of these SNPs, located in exon 9 (rs17052017) is an non-synonymous SNP, while the SNP located in exon 4 (rs352168), is synonymous. This synonymous SNP in exon 4 have previously been identified in our study population of VP patients (Steyn 2002). The absence of copious numbers of sequence variants in the *ALAS1* coding region is another indication of the conservation and importance of this gene in the regulation of haeme synthesis.

While the previous sections discussed several factors mediating a regulatory role in *ALAS1* expression, various other, yet unidentified regulatory influences may exist (Podvinec *et al*. 2004). Numerous factors participate to regulate gene expression, each of which may contribute to *ALAS1* gene expression.

2.4 MOLECULAR REGULATION OF GENE EXPRESSION

The ability to orchestrate the synchronized regulation of the expression of the gene complement, is a process crucial to the survival of each living cell. The intricate network of molecular interactions functioning in eukaryotic gene regulation is optimally adjusted to provide complex patterns of gene expression, thereby ensuring the exact levels of gene products specific for every cell type, while still maintaining cellular homeostasis (Tupler *et al*. 2001).

In eukaryotes, gene expression entails a range of processes such as the transcription of DNA into RNA, translation, protein folding, post-translational modifications and epigenetic mechanisms such as DNA methylation, polyadenylation and packaging of chromatin (Grunstein 1997; Fickett and Hatzigeorgiou 1997; Pedersen *et al.* 1999; Jaenisch and Bird 2003; Lareau *et al.* 2004).

Several mechanisms mediating the regulation of gene expression have been established. During alternative splicing, different combinations of exons and introns are included or not included in the mRNA transcript, resulting in polypeptides with varying biological functions and properties (Lejeune *et al.* 2001; reviewed in Matlin *et al.* 2005). Several human genetic diseases and

developmental defects have been associated with disruptions of alternative splicing control in genes (Cáceres and Kornblihtt *et al*. 2002; reviewed in Kornblihtt *et al*. 2004; Zhang *et al.* 2004). The use of alternative promoters and different transcription start sites (TSSs), is another way in which one gene can encode for various polypeptides. Regulatory effects can also arise from mutations in the transcribed sequence that affects the transcription rate or the half-life of a mRNA transcript (Carter *et al.* 2002; Wray *et al.* 2003). Recently, a research group also discovered the important role of non-coding RNA molecules, originating from non-coding DNA regions called intergenic spacers (IGS), in a number of control processes including the regulation of ribosomal RNA (rRNA) production (Wang *et al.* 2006). Additionally, various post-translational modifications exist that contribute to the regulation of gene expression (Kahn *et al*. 1975). Alternative regulation of gene expression indicates that each gene can be used in a variety of different ways depending on how it is regulated. Regulation can occur at any of these stages but the primary regulation takes place during transcription (Wright 1993; Lemon and Tjian 2000; Levine and Tjian 2003). Non-coding DNA regions have demonstrated to be fundemental in transcriptional regulation.

Every functional gene is flanked by these non-coding regulatory sequences that were once considered as irrelevant genetic waste and incorrectly termed as “junk DNA”. However, the “garbage” of DNA has proved to be imperative in transcriptional regulation. Regulatory regions in non-coding DNA, together with the activity of proteins termed transcription factors (TFs), which are encoded elsewhere, regulate when expression occurs, at what level, under what environmental conditions, and in which cells or tissues (Warren 2002). The non-coding DNA region upstream of the protein-coding sequence of genes, binds to and directs RNA polymerase II to the TSS of the gene promoter. Promoters are required for transcription initiation and are therefore among the most important regions driving gene expression. The exact promoter region of a gene is difficult to delineate, but can roughly be characterized into a core-, proximal-, and distal promoter. The core/basal promoter typically encompass the region ~40 bp upstream from the TSS, followed by the proximal promoter, defined as the region ~250 bp from the TSS, while regulatory sites further upstream are classified as the distal promoter (Tjian 1995).

Regardless of these abundant reports on genes and their regulation by relevant DNA sequences, certain aspects of promoter mechanisms remain incompletely understood or appreciated.

2.4.1 Transcriptional regulation

The environment of each cell is subjected to constant change. The ability of cells to adapt to this constantly changing environment, is mediated primarily by transcriptional regulation. Genes have the capability to integrate the multitude of stimuli, and regulate transcriptional expression accordingly. These numerous inputs reach the cell’s “machinery” in the form of

TFs and cellular proteins that bind in a sequence specific manner to regulatory elements, thereby affecting the rate of transcription (Ogbourne and Antalis 1998; Butler and Kadonaga 2002). Transcriptional regulation of protein-coding genes is among the most sophisticated regulatory processes and is mediated by an arrangement of cis- and trans-acting factors.

2.4.1.1 Cis-acting factors: promoter regulatory elements

Transcriptional control can be mediated by regulatory sequence elements located in the coding- and/or the non-coding gene regions or regions far from the transcription initiation site.

i) Core promoter

A core promoter is defined as the minimal stretch of contiguous DNA sequence that is sufficient to direct the accurate initiation of transcription by the RNA polymerase II machinery (reviewed in: Smale 2001; Butler and Kadonga 2002). In humans the core promoter of a gene is recognised by the TSS and the TATA box, the first eukaryotic core promoter motif to be identified (reviewed in: Lee and Young 2000; Butler and Kadonaga 2002). The core promoter of some genes can also contain an initiator element (Inr) that binds the transcriptional apparatus (Smale and Baltimore 1989; Javahery *et al.* 1994; reviewed in: Smale and Kadonga 2003). The transcriptional apparatus generally consists of the RNA polymerase II enzyme and auxiliary factors, commonly termed the "basal" or "general" TFs (Orphanides *et al*. 1996). The binding of basal TF, TFIID, comprised of the TATA-binding protein (TBP) and associated factors (TAFs), and the rest of the transcription initiation complex to the TATA element, activate transcription (reviewed in: Smale and Kadonga 2003).

Although the action of this basal transcriptional machinery is sufficient to initiate transcription, the core promoter drives transcription at a basal rate when unaided by additional factors (Ogborne and Antalis 1998; Lemon and Tjian 2000). Sufficient *in vivo* regulation therefore requires the action of additional regulatory elements (Paranjape *et al.* 1994).

ii) Enhancers, silencers and insulators

The tempo of transcription can be greatly increased by the binding of additional TFs and cellular proteins to cis-acting regulatory regions, collectively termed transcription factor binding sites (TFBSs), which interact to alter the transcription rate of a particular gene. The regulation of TBP binding depends on these upstream regulatory sequences that are located at various distances from the TSS: some regulatory elements are located adjacent

to the core promoter while others are positioned several tens of kilobases upstream or downstream from the promoter.

TFBSs typically consists of a 6-25 bp sequence region that are surprisingly conserved in eukaryotes (reviewed in: Blackwood and Kadonaga 1998; Lettice *et al.* 2002; Cooper and Sidow 2003). TFBSs regulate promoters by their recognition sites, each tailored for a sequence-specific TF (reviewed in: Blackwood and Kadonga 1998; Davidson 2001; de Bruin *et al.* 2001). The distribution of TFBSs in gene promoter regions is not uniform and their frequencies tend to increase in the regions closer to the TSS. Results showed that the number of TFBSs is two to three times higher in the -100 bp to -50 bp region which suggests that most TFBSs occur within 250 bp of the transcription initiation site. Beyond (downstream of) the TSS, the frequency of binding sites decreases dramatically, with the frequency value dropping below that of the entire upstream sequence (Guo and Jamison 2005).

When the binding of a TF to a TFBS initiates or activates transcription, the element is termed an enhancer, while elements bound by TFs mediating the opposite effect, are designated as silencers (Wang and Giaever 1988; reviewed in: Blackwood and Kadonaga 1998). Classical silencers and enhancers are thus characterized as sequence elements that can respectively repress or enhance promoter activity in an orientation- and position-independent manner with regard to the particular gene (reviewed in: Lee and Young 2000).

Enhancers appear to act by increasing RNA polymerase II binding and initiation. Most models describe the mechanism of transcription regulation executed by enhancer elements as an event that includes some type of DNA looping. A DNA loop physically brings activator proteins bound to distant enhancer elements, into proximity with complexes associated with the basal promoter transcriptional initiation complexes (Szutorisz *et al.* 2005). The model proposes that enhancer-bound activators have a high affinity for the basal transcription initiation complex and assist their binding to DNA (Hochschild and Dove 1998; Barberis and Petrascheck 2003; Szutorisz *et al.* 2005).

Different types of silencers exist that can affect transcription by interfering with positive acting TFs or by affecting chromatin structure and intron splicing to ultimately down-regulate gene expression (Clark and Docherty 1993; Davies *et al.* 1998; reviewed in: Ogbourne and Antalis 1998; Monsalve *et al.* 2000). The precise mechanism of silencer action also appears to involve DNA looping/bending, enabling the bound repressor to be in the correct orientation to interact and obstruct downstream positive regulators. (Ryder *et al.*

1986; Spana and Corces 1990; Arnold *et al.* 1996; reviewed in: Ogbourne and Antalis 1998)

Another type of transcriptional element is a boundary element (or insulator). These DNA elements function as transcriptionally neutral segments that block the spreading of the influence of enhancers and silencers, to nearby genes (Bell and Felsenfeld 1999; Gerasimova and Corces 2001).

A set of genes may each contain identical sequences for a particular transcriptional regulatory element, but its importance in each specific gene may differ according to the tissue-type and physiological circumstances. This indicates that a gene may contain many enhancers and silencers, but that their presence does not guarantee that it will bind its cognate TF *in vivo* and consequently, may have no influence on transcription (Mitchell and Tjian 1989; Li and Johnston 2001).

The contribution of numerous regulatory elements on the expression of a single gene can be described as cumulative to the overall spatial and temporal regulation of a gene (Blackwood and Kadonaga 1998). Although our knowledge of all these regulatory elements functioning in human genes is not complete, it is evident that these TFBSs play strategic roles in the regulation of gene expression, thereby directing intracellular homeostasis, cell growth, differentiation and specification of development (Adams and Timberlake 1990; Gray and Levine 1996; Belting *et al.* 1998; Cho *et al*. 200; Chen *et al.* 2005).

2.4.1.2 Trans-acting factors: Transcription factors (TFs) - activators and repressors

TFs are defined as proteins that either positively (activators) or negatively (repressors) affect the rate of transcription *via* interaction with regulatory DNA sequences (enhancers and repressors respectively) and/or other regulatory proteins. In human cells there exists a large variety of TFs and the most recent literature indicate that more than 1400 known human TFs have been described (Lander *et al.* 2001; references cited in: Itzkovits *et al*. 2006). However, because of their sequence specificity and binding kinetics in different tissue-types and conditions, not all are operational in each gene's transcriptional control (Pabo and Sauer 1992; Dröge and Müller-Hill 2001).

Activators typically consist of a DNA binding domain and an activation domain. The binding of the activator to DNA serves to locate it to a position from which the activation domain can interact with components of the RNA polymerase II complex and recruit the transcription complex in the vicinity of a gene promoter, thereby increasing its local concentration and facilitating binding to the promoter (Saltzman and Weinmann 1989;

Jones 1990; reviewed in: Barberis and Petrascheck 2003). It has been shown that some activators can also interact with other cellular proteins (co-factors) and/or lead to a conformational change in chromatin structure (reviewed in: Clark and Docherty 1993). In addition to this, some activators have ligand-binding domains, such as the hormone receptors that have hormone-binding domains that are essential for controlling its activity (Villard 2004). Although activators bind to their own consensus DNA sequence, most bind to a range of degenerate motifs (Pabo and Sauer 1992) and in higher eukaryotes, genes are activated not by just one but by several activators that act in concert (Ptashne and Gann 2002 as cited in Barberis and Petrascheck 2002; Green 2005).

TFs can also act as repressors by negatively regulating transcription. Since repressors are difficult to define, Gaston and Jayaraman (2003) categorized them into three major classes. Members of each class can, however, repress *via* multiple mechanisms and some repressors fall into more than one category. Class I repressors are defined as DNA-binding proteins that negatively regulate the transcription of specific genes. Class II repressors are TFs that do not bind directly to DNA, but instead, are recruited to promoters by other proteins. Class III repressors are proteins that do not necessarily bind to DNA directly or indirectly but target activators, co-activators or components of the transcription initiation complex and often reduce the amount of functional protein available to regulate transcription (Gaston and Jayaraman 2003). Class II and III repressors are generally considered to be co-repressors. Once a co-repressor is recruited to a gene by interaction with a repressor, it could be transferred to a nearby DNA-bound activator protein where it could then serve to block activation, perhaps by obstructing an interaction between the enhancer's activation domain or the general machinery (Latchman 1996a). This interaction with an enhancer to block the activity of its activation domain is a phenomenon known as "quenching" (Gray and Levine 1996). Repressors can also bind to a sequence adjacent to, or overlapping the binding site for an activator, thus preventing the binding of the activator by steric hindrance (Clark and Docherty 1993). Repression can therefore be achieved by forming a non-DNA-binding complex with the activator or by organizing the DNA into an inactive chromatin structure preventing the activator from binding.

The activities of many TFs are context dependent and are often modulated by transcriptional co-regulators (co-activators and co-repressors; Shibata *et al.* 1997). These co-regulators (or co-factors) typically lack a DNA binding domain, but contain a domain that mediates protein-protein interactions, thereby strongly influencing the binding specificity of TFs (Knoepfler and Kamps 1995). Various TFs, such as some NR, can affect transcription even in the absence of ligands. This effect is mediated by co-regulator proteins that bind to the unliganded receptors. Upon ligand binding, co-regulator complexes are dissociated and

various regulatory complexes are recruited to the liganded receptor (Love *et al.* 2000; Rosenfeld and Glass 2001; Privalsky 2004). From several studies, it is now also evident that a given TF can function with several types of co-activators or co-repressors and vice versa (Karin 1990; Onate *et al.* 1995; Horwitz *et al.* 1996; Liu *et al.* 2002; Privalsky *et al.* 2004; Spiegelman and Heinrich 2004; Goodson *et al.* 2005). Thus, a single activated TF can induce transcription of one gene while repressing that of another. For this reason, TFs are grouped together into 5 superclasses (Appendix 5), not on the basis of their influence on transcription, but on similarities in their DNA-binding domains. These superclasses are each subdivided into classes (Appendix 5) and the classes subdivided into families and subfamilies (Coulson and Ouzounis 2003; Stegmaier *et al*. 2004).

Due to the focus of this study, the NR transcription factor class warrants further discussion. As described earlier (section 2.3.1.2), NR are known to interact with distant motifs in the *ALAS1* promoter, influencing transcription in response to various drugs.

i) Nuclear Receptors (NR)

NR are a superfamily of ligand-activated TFs that work in concert with co-activators and co-repressors to mediate specific gene expression (Escriva *et al.* 2000; Handschin *et al.* 2004). NR are involved in a wide range of physiological processes including development, metabolism, energy homeostasis and apoptosis (Kastner *et al.* 1995; Beato *et al.* 1995; reviewed in: Olefsky 2001). This class of receptors constitutes an approximate variety of between 50 and 60 different members of functional diverse TFs (For a detailed description of the different NR types, see Nuclear Receptors Nomenclature Committee 1999; The International Human Genome Sequencing Consortium 1999; Robinson-Rechavi *et al.* 2001; Duarte *et al*. 2002; Ruau *et al*. 2004; online at http://www. ens-lyon.fr/LBMC/laudet/nurebase/nurebase.html.). Molecular cloning and structural/functional analyses have revealed that the members of the NR superfamily share a remarkable structural and functional similarity in that they have a common functional domain structure (Kumar and Thompson 1999). Upon binding of the particular ligand to the receptor it undergoes conformational changes which mediate binding of the receptor to a specific DNA recognition sequence in the target gene (Wurtz *et al.* 1996; Kumar and Thompson 1999; reviewed in: Germain *et al.* 2003). Three different classes of NRs have been identified as indicated in Table 2.1. (Mangelsdorf *et al.* 1995).

Table 2.1 Examples of the nuclear receptor (NR) superfamily.

Nuclear Receptors (NR)		
Class I (Binds steroids)	**Class II (Binds other ligands)**	**Orphan Receptors (Currently being classified)**
glucocorticoid receptor (GR)	retinoid X receptor (RXR)	pregnane X receptor (PXR)
mineralocorticoid receptor (MR)	retinoic acid receptor (RAR)	constitutive activated receptor (CAR)
androgen receptor (AR)	thyroid hormone receptor (TR)	
estrogen receptor (ER)	vitamin D receptor (VDR)	

ii) Example of NR: Estrogen Receptor (ER)

The estrogen receptor (ER) is perhaps the most well defined NR system regarding biologic responses and clinical implications. The ER is a classical model of a Class I NR (Evans 1988). Two types of estrogen receptors, ERα and ERβ, have been identified (Green *et al.* 1986; Moore *et al.* 1998). In humans, both receptors bind to the same DNA response elements (Kuiper *et al.* 1997) and usually form homodimers, although ERα and ERβ heterodimers have been reported (Kumar and Chambon 1988; Ogawa *et al.* 1998; Cowley *et al.* 1997). These two receptors exhibit distinct transcriptional properties (Hall and Mcdonnell 1999) but exhibit similar, although not identical, ligand binding characteristics. Hall and McDonnell also specified that the one role of ERß is to modulate ERα transcriptional activity (Hall and Mcdonnel 1999). Even though estrogen generally regulates gene expression by activation of transcription (Beato 1989; Cho and Katzenellenbogen 1993), a growing number of investigations have demonstrated that a ligand bound ER can also mediate transcriptional repression (Ray *et al*. 1997; Chen *et al*. 1999; Jones *et al*. 2002). Their distribution varies between different tissues or organ types and in many tissues both subtypes are expressed, while in the liver, ERα is the predominant type (Porter *et al.* 1983; Couse *et al.* 1997; Taylor and Al-Azzawi 2000; Pelletier 2001).

ERα and ERβ have six domains of which the four major functional domains are: a N-terminus that modulates transcription in a gene and cell specific manner; the C-domain that interacts directly with the DNA; the D-domain that binds the ligand; and the F-domain that recognizes estrogen agonists versus antagonists (Montano *et al.* 1995; Kuiper *et al.* 1998;

Pike *et al.* 1999). The differential activities of the two ERs arise from functional variations in the receptor's N-termini (Huang *et al*. 2005).

The mechanism of ERα action has been widely studied. According to literature, *in vivo*, in the absence of the hormone estrogen, the ERα remains in an inactive state in the nucleus. When the receptor interacts with estrogen or estrogen-like ligands, it is phosphorylated an undergoes a conformational change, termed activation, and forms a homodimer (Nardulli *et al.* 1995). However, the belief of the strict requirement of estrogen for the activation of estrogen receptors have been challenged by evidence that ERα is also activated by estrogen-independent mechanisms. These include protein kinase A (PKA), protein kinase C, cytokines, peptide growth factors, neurotransmitters or cell cycle regulators (Nilsson *et al*. 2001, and references therein). Even though ERα is phosphorylated in the absence of estrogen in some cells (Le Goff *et al*. 1994), enhanced phosphorylation is observed in response to estrogen stimulation (Ali *et al*. 1993; Le Goff *et al*. 1994).

ERα activation instigates a cascade of events that lead to target gene expression. ERα can mediate its effect through two methods: (i) direct binding of ERα to the estrogen response element (ERE) located within the regulatory regions of the target gene, resulting in direct interaction with the basal transcription machinery, ensuing an altered transcription rate (Klinge 2000); or (ii) "tethering", a mechanism where the ERα binds to other TFs to stabilize the complex, enabling these TFs to bind the DNA (Kushner *et al.* 1998 and references therein). Of note, around one third of known human ERα target genes associate only indirectly with ERα through intermediate transcription factor(s) (Jakacka *et al.* 2001; reviewed in O'Lone *et al.* 2004). Abundant reports of tethering include the interaction of ERα with stimulatory protein 1 (Sp1) (Wang *et al.* 1998; Scholz *et al*. 1998; Qin *et al.* 1999; Petz *et al.* 2000; Vyhlidal *et al.* 2000; Li *et al.* 2001; Safe 2001) and the interaction of ERα/β with activator protein 1 (AP1) (Kushner *et al.* 2000; Barkhem *et al.* 2002; Jones *et al.* 2002; Garcia-Arencibia *et al.* 2005).

The minimal ERE that confers estrogen inducibilty has been reported to consist of two palindromic inverted repeats separated by three bp to form the consensus sequence: 5'-GGTCAnnnTGACC-3' (Klein-Hitpass *et al.* 1988). In a successive study it was determined that the optimal ERE sequence is a 17 bp inverted repeat of 5'-TGACCTG-3' (separated by three bp) with an AT-rich flanking region (Peale *et al*. 1988). However, it is surprising that to date, only a few near-consensus EREs have been characterized in the promoters of estrogen-regulated human or mouse genes (reviewed in: Sanchez *et al.* 2002). Alternatively, it has been found that many estrogen responsive genes contain EREs and

half-palindromic EREs (halfsites) that differ from the consensus by one or more nucleotides and are subsequently designated as imperfect elements (Berry *et al.* 1989; Anolik *et al.* 1993; Klinge *et al.* 1997; Driscoll *et al.* 1998; Klinge 2001). Kato *et al.* (1992) defined the binding capabilities of an ERE halfsite, 5'-TGACC-3', and determined that it binds with up to 100-fold less affinity than a perfect or imperfect palindromic ERE. They, and several other authors, confirmed the ability of an ERα to bind to widely separated consensus and half-palindromic ERE sites, but with reduced affinity, to synergistically activate expression (Kato *et al.* 1992; Driscoll *et al.* 1996; Nardulli *et al.* 1996; Sathya *et al.* 1997; Anderson and Gorski 2000).

Abundant reports have also commented on the importance of the nucleotides directly flanking the ERE sites, impacting on ERα-ERE binding (Berry *et al.* 1989; Martinez and Wahli 1989; Anolik *et al.* 1993; 1996; Xing and Archer 1998). Harendza *et al.* (2005) eloquently illustrated this. They established the SNP distribution in the regulatory regions of the Gelatinase A (matrix metalloproteinase-2) gene promoter region by means of sequence analysis. In the 1665 bp upstream region a SNP, located at -1575 bp, was found to be situated 1 bp from a possible half-palindromic ERE site. The functionality of this SNP was examined by transient transfection studies in an ERα-positive cell line and results indicated that the -1575G allele functioned as an enhancer, whereas the -1575A allele reduced transcription activity significantly. Results were confirmed by co-transfection experiments with an ERα-expression vector in cells that do not constitutively express an ER, and revealed that ERα is absolutely required for transcriptional enhancement. Moreover, allelic distribution results indicated that a C→T transition within a Sp1 binding site at -1306 in the Gelatinase A promoter, was in linkage disequilibrium with the -1575G→A transition. Reporter studies with a luciferase construct with the -1575A/-1306T allele, versus a construct with the wild type -1575G/-1306C allele, demonstrated an additive reduction in response to ERα- and Sp1-expression plasmid transfections, suggesting that the Sp1 and ERα proteins act independently, but with a calculated effect. In conclusion, the evidence indicated that the -1575G→A variant, linked to the -1306C→T variant, impairs the transcriptional response of Gelatinase A to ERα binding and is associated with a significant decrease in genetic fitness. This study demonstrated the importance of a promoter variant in a regulatory site, the repercussions of a variant in the nucleotides directly flanking a ERE half-site as well as the involvement of other cis-regulatory sites on ERα-ERE action. (Harendza *et al.* 2003).

In a study by Bourdeau *et al*. (2004), *in silico*, *in vitro* and *in vivo* experiments were performed which displayed that many near-consensus EREs previously identified in human genes, often have counterparts at similar positions in their mouse orthologs. They identified approximately 70 000 motifs in the human and mouse genomes and demonstrated that approximately 1% of the elements appear to be conserved in the flanking regions (-10 kb to -5 kb) of the two species. Results revealed that near-consensus EREs occur frequently in both genomes and that chromatin structure are likely to modulate access to binding sites, while many far upstream elements are evolutionarily conserved and bind ERs *in vivo* (Bourdeau *et al.* 2004).

2.4.1.3 Epigenetic factors

Besides the regulation mediated by regulatory elements bound by TFs, epigenetic control incorporates the genetic phenomena that affect gene expression without any nucleotide changes. These heritable factors include DNA methylation, genomic imprinting, histone modifications and X-inactivation to name a few (Pastinen *et al.* 2004; Pastinen and Hudson 2004). Gene regulation by epigenetic factors are, however, not within the scope of this study, and therefore will not be discussed in this review.

Ultimately it is the combination and specific rearrangement of all these regulatory factors and the interaction between positive and negative regulators *in vivo*, that determine the individualized spatial transcriptional program of each gene.

2.4.2 Role of transcriptional control in phenotype

Many reviews argue that differences in the expression of protein-coding genes are responsible for the phenotypic distinction between the various kinds of cells in higher eukaryotes, as well as their evolution, response and adaptation to the environment (Carroll 2000; Lemon and Tjian 2000; Pastinen and Hudson 2004; Pastinen *et al.* 2004). Numerous authors have also evaluated the direct correlation between changes in how, where or when a gene is expressed, and the phenotypic consequences (Wray *et al.* 2003, and references therein; Knight 2005; Wittkopp 2005; Donaldson and Gottgens 2006). The destruction or alteration of a regulatory element may disrupt the use of its target gene, or directly lead to altered gene expression. It is also possible that a genetic variant, even if it is not located in a known TFBS, can modulate transcription by e.g. altering the DNA structure (reviewed in: Maston *et al*. 2006). These changes in transcription rate will ultimately cause an increase or decrease in the amount of protein expressed by the target gene. Given the fact that protein levels regulate various biological processes, genetic variants in regulatory DNA sequences can therefore influence a multitude of pathways (Hoogendoorn *et al*. 2003; Knight 2003). These can include developmental- and physiological pathways, critical to fundamental

biological processes which may be involved in disease mechanisms. Their influence on these important physiological processes makes evident their important function in the genetic basis of disease phenotypes (Villard 2004; Nadeau 2005). This is apparent from the growing list of human diseases and disease penetrance that are influenced by genetic defects in transcriptional regulatory regions (Crossley and Brownlee 1990; Bullido *et al.* 1998; Martin *et al.* 1998; Yan *et al.* 2002; Arbini *et al.* 2006; De Gobbi *et al.* 2006). Several other examples are available where changes in promoter sequences of genes are directly responsible for altered or reduced gene expression which in turn influence the phenotypic expression of disease (Hegele 1997; van't Hooft *et al.* 1999; Mitchison 2001; Lettice *et al.* 2002; Simeonia *et al.* 2004; Villard 2004; Buckland 2004; reviewed in Knight 2005).

In the past few years, a growing number of diseases have also shown to be associated with mutations in genes coding for TFs and co-factors. This is to be anticipated when recognizing the vital role of TFs and the fact that a single TF can influence expression of numerous genes. Mutations in genes coding for general TFs or activators and repressors have been well documented as a source for differences in organismal phenotype, causes for disease or phenotypic modifiers of human diseases such as leukaemias, xeroderma pigmentosum, mental retardation with combined pituitary hormone deficiency, maturity-onset diabetes of the young, Rubinstein-Taby syndrome, heart disease, etc. (Latchman 1996b; Yamagata *et al.* 1996; Giles *et al.* 1998; Schott *et al.* 1998; reviewed in Lehmann 2001; Winter 2003; Garg *et al.* 2005; Nikolova and Vilain 2006).

To summarize, genetic variants alone in cis- or trans-acting modulators, or their combination in certain haplotypes may consequently be associated with increased risk of disease phenotype (Botstein and Risch 2003). A fascinating area of inquiry, in which the clinical repercussions of a non-coding genetic variant are explored, makes available information applicable to inter-individual variation and disease variation which includes variable clinical expression such as diseases with incomplete penetrance. This may hopefully provide insights into, as well as novel targets for, therapeutic interventions in human diseases (Villard 2004).

2.4.3 Distrubution of sequence variants in regulatory regions

To investigate more efficiently the genetic variants with significant biological consequences, a practical approach is to concentrate efforts on the variants located in genomic regions with important functions (Rasmuson 2002). Previous sections indicate that non-coding regions are central to the regulation of gene expression. Comparison of long, non-coding genomic DNA sequences, involved in the regulation of gene expression, have revealed that these sequences are highly conserved between species, and have consequently been designated,

conserved non-coding sequences (CNSs) (Hardison 2000 and references therein). Early comparative studies by Duret *et al.* (1993) uncovered a surprisingly strong conservation of the flanking- and UTRs of orthologous genes among widely diverged vertebrate groups. These conserved blocks of non-coding sequences have, in many cases, been found to correspond to regulatory regions (Hardison 2000). The work of Dermitzakis *et al.* (2002) and others supported these early findings and suggested that the conservation of non-coding DNA is not just a feature of a few mammalian species (Dermitzakis *et al.* 2002; Gaffney and Keightley 2004; Ganley *et al*. 2005).

One of the most common methods for studying genetic variation is the identification of SNPs. Together with tandemly repeated segments and small insertions and deletions, SNPs form part of the 0.1% of the divergent DNA between any two random individuals. SNPs contributing to phenotypic variation are referred to as functional SNP (fSNPs) (Macdonald and Long 2005). fSNPs can be located in the coding regions of protein-coding genes and affect the protein structure/function, or can be found in introns where they interfere with intron-exon splice sites. However, fSNPs altering the expression of complex diseases are generally those located in the promoter regions of genes that affect the transcription rate (Buckland 2004).

As a result of high-throughput genotyping methods, millions of SNPs in the human genome have been reported in recent years. It is estimated that there exist more than 10 million SNPs throughout the genome, but that the number varies within genes and gene regions [Smigielski *et al.* 2000; Zhao and Zhang 2006; http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=snp (January 2007)].

In a current research article, Guo and Jamison (2005) investigated the occurrence of SNPs in gene regulatory regions. In the study, they retrieved promoter sequences from the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) database and mapped promoter SNPs using information from the NCBI dbSNP database (Sherry *et al.* 1999; http://www.ncbi.nlm.nih.gov/SNP). Consensus sequences of TFBSs were retrieved from the TRANSFAC database 6.0 (Wingender *et al.* 2001). Their analyses showed that in gene promoter regions, more SNPs occur in proximity to the TSSs, than in regions further upstream. Unexpectedly, it was found that the number of SNPs in the predicted TFBSs, is higher than in non-binding site sequences. This is contradictory to the hypothesis that TFBSs are highly conserved because they are important to gene regulation (Schena 1989). They postulated that the high frequency of SNPs in TFBSs might demonstrate diverse requirements for differential gene expression under different conditions. Alternatively, their findings could be ascribed to the fact that the identified putative TFBSs might not be functional binding sites *in*

vivo in the specific context, or, that to date, research has largely been focused on the regions in/closely flanking known genes.

On the other hand, it is to be expected that gene regulation has the ability to evolve if DNA changes occur in the functional non-coding regions, resulting in new gene regulatory programs. Even though the vast majority of SNPs are located in non-coding regions (Shastry 2002), their significance is determined by their effect on gene regulation. Regulatory SNPs differ in importance; while some binding site sequence alterations caused by SNPs may totally interrupt gene expression, others may only influence the level of expression slightly (Fried *et al*. 2004; Zhang *et al*. 2006). This is apparent from the degenerative nature of most TFBSs (Duret and Bucher 1997; Dermitzakis and Clark 2002) and implies that a system exist for maintaining a robust transcriptional response, independent of binding specificity (Zhang *et al*. 2006). These findings emphasize the importance of promoter analysis and subsequent verification experiments.

2.4.4 Promoter analysis

In the recent past, the focus of genomics has shifted from sequencing, to processing and interpreting the sequence data. In particular, the identification the factors determining the regulation of gene expression has attracted much attention. Various methods are used to analyze promoter regions and to identify, and locate regulatory elements (such as TFBSs) in non-coding DNA regions.

2.4.4.1 *In silico* assays

Since the availablity of the human genome sequence, *in silico* science has stepped into a new era of increasingly overloaded database-sequence information and high-throughput genomic analyses. However, at present, we still have an incomplete picture of how the genome regulates gene expression and ultimately gene function, due to the limited laboratory investigations on the DNA sequences directly flanking genes. Current methodology is inefficient to investigate the vast majority of the human genome sequence and regions flanking genes (Pastinen and Hudson 2004). Finding regulatory elements are key to understanding how these regions influence the cell ultimately, but it remains a challenge to locate these short, variable and usually degenerate, stretches in the genome. One approach to keep up with the dramatic increase in sequence information and find regulatory regions, is to utilize biocomputational (*in silico*) approaches.

Computational search for regulatory sites in promoter regions is an effective tool to identify possible sites involved in transcriptional regulation. However, searching for these short patterns in large promoter sequences frequently produces exceedingly high numbers of

results. Consequently, it is still almost impossible to predict the precise expression of a given gene by simply analyzing its promoter sequence with software programs (Bucher 1999). Regardless of this, the bioinformatics approach remains a popular method to uncover potential TFBSs and direct experimental strategies. Bioinformatics apply techniques derived from disciplines such as mathematics, computer science and statistics to try and make sense of, and organize non-coding DNA sequences. Given the degeneracy of TFBSs, the most popular method used to describe these elements, is by using position weight matrices (PWMs) (Quandt *et al*. 1995; Stormo 2000). Various software programs rely on these computationally intense algorithms, of which a few are (1) *TRES*, a program that identifies regulatory elements by comparative promoter analysis using matrices described in TRANSFAC® 7.0 (Heinemeyer *et al.* 1999); (2) *Matinspector*, a software program that searches for the most true positive binding site matches according to a library of PWMs (Quandt *et al*. 1995; Cartharius *et al.* 2005); (3) *AliBaba2*, a program that scans the query DNA against a TF database (TRANSFAC 3.5 public; Heinemeyer *et al.* 1999) to generate context specific matrices (Grabe 2002) and (4) JASPAR, another open-acess database of annotated matrix-based TFBS profiles for eukaryotes (Sandelin *et al*. 2004).

Another useful tool for locating new regulatory elements is phylogenetic analysis. Phylogenetic footprinting refers to the method of comparing evolutionary-related sequences, such as cross-species comparison of the promoter regions of orthologous genes (Tagle *et al.* 1988). Paradoxical to the findings by Guo and Jamison (2005; section 2.4.3), this method relies on the theory that essential regulatory regions tend to undergo less mutational events and because of their significance, are highly conserved through evolution (Tompa 2001; Boffelli *et al*. 2003; Nobrega and Pennacchio 2004; Van Hellemont *et al*. 2005). Of importance when using this method is the choice of species to be compared. It is essential for the efficiency of phylogenetic footprinting that the species compared are not too closely related to allow for enough evolutionary time for the accumulation of mutations at base positions. Conversely, if species are too distantly related, the detection of conserved regulatory elements may be impossible because they will have diverged too much to preserve any significant similarity (Duret and Buchert 1997). When the upstream regions of these orthologous genes are aligned, areas of sequence conservation may be detected (Fessele *et al.* 2002; Wray *et al.* 2003; Grad *et al.* 2004). These conserved sequences are likely to have biological importance such as that of TFBSs (Gumucio *et al.* 1996; Duret *et al.* 1993). A suitable program to align long, multiple genomic sequences over any evolutionary distance, is Multi-LAGAN (Brudno *et al.* 2003). Ultimately, TFBSs predicted by *in silico* methods, need to be confirmed by functional laboratory experiments.

2.4.4.2 *In vitro* assays

i) Electromobility shift assay (EMSA)

The EMSA is a sensitive technique for studying the effect of protein-DNA interactions *in vitro* (Fried and Crothers 1981). It is typically used to determine if variants in regulatory regions alter the binding of TFs to potential regulatory elements. Two radiolabelled oligonucleotides, approximately 20 bp in length, that match the variant in the two allelic forms, are synthesized. These are bound to a nuclear DNA extract to serve as a template to which purified proteins may bind. The products of the reactions are separated through gel-electrophoresis, and autoradiographed to reveal the greater physical retardation of protein-DNA complexes through the gel matrix. The limitation of the EMSA technique is that the DNA-binding region of interest must be known (Woo *et al.* 2002).

ii) DNase I footprinting assay

This protein-DNA interaction technique identifies at which position on a DNA sequence a specific protein-binding site is situated. When the protein is bound to the DNA, it protects the phosphodiester bonds from cleavage by DNase I. These regions can be detected by electrophoresis in the appropriately designed gel matrix (Galas and Schmitz 1978; Angel *et al*. 1987; Mueller and Wold 1989).

iii) Real-time PCR

The real-time PCR method uses reporter molecules that intercalate with double-stranded DNA to monitor the production of amplification products during each cycle of the PCR reaction. This method requires firstly the conversion of mRNA to cDNA by reverse transcription. During amplification, the fluorescent signal reaches a threshold level that correlates with the amount of original target sequence. Comparison of the initial quantity of mRNA from a wild type gene promoter versus one with a sequence variant, enables the determination of the effect of the variant on gene expression. Housekeeping genes such as glyceraldehyde-3-phosphate, that are not expected to change under experimental conditions, can serve as internal controls (Valasek and Repa 2005).

iv) Reporter gene assays

Expression vectors with reporter genes are used to test for the activity of a particular promoter or promoter element in a specific cell (Alam and Cook 1990). In this instance the reporter gene is placed under the control of the target promoter and the activity of the reporter gene is quantitatively measured. The results are normally reported relative to the activity of a “consensus” promoter known to induce strong gene expression.

v) Chromatin immunoprecipitation (ChIP) assay

This type of study is used to confirm protein-DNA interactions in particular cells under specific conditions and to determine the precise DNA targets for a specific protein (Orlando *et al.* 1997). HaploChIP is a *in vivo* modification to this *in vitro* approach that allows the allele-specific expression of a gene to be assayed by using RNA polymerase II as a marker. It has the advantage to analyse expression in the chromosomal environment of naturally occuring haplotypes in the presence of all the regulatory processes that operate in the cells investigated. Customarily the relative amount of phosphorylated RNA polymerase II bound to different alleles are measured with PCR and primer extention, and quantified by MALDI-TOF (Knight *et al.* 2003).

2.4.4.3 *In vivo* assays

i) Model organism

Although the evaluation of gene expression in cell culture conditions reveals important information regarding gene regulation, the activity of a specific gene at cellular level does not reveal satisfactory information about all the regulatory factors involved in the context of a complex animal. Cell cultures cannot simulate the intricate physiological interactions functioning in tissues and organ systems, which make experiments in transgenic animal models the ultimate functional studies to assess gene regulation. Although most transgenic animals used by scientists are small, simple animals such as *Caenorhabditis elegans, Drosophila melanogaster* and *Xenopus laevis*, examples are available of studies that used mouse or rat models to determine which region within the promoter is responsible for tissue-specific expression (Swoap 1998; reviewed in: Janson and During 2001; Ike *et al*. 2004).

CHAPTER 3:

MATERIALS AND METHODS

3. MATERIALS AND METHODS

MATERIALS

3.1 STUDY COHORT

Genomic DNA (gDNA) of 19 South African VP patients (Caucasian and Mixed Ancestry individuals) from a previous study (Steyn 2002) was available for the current study. All patients and controls in this study gave written informed consent for the use of their gDNA. In cases where DNA stocks were depleted, blood was collected from the patients and DNA extractions performed according to a modified salting out procedure as described by Miller *et al*. (1988). The patient group included 2 male and 17 female VP patients with ages between 14 and 70 years.

Urine and stool samples from the patient group were previously biochemically analysed by a private pathology laboratory to obtain porphyrin and porphyrin-precursor results. Based on these biochemical analyses, along with medical diagnoses by a dermatologist and neurologist and clinical questionnaires completed by the patients and/or their genetic counsellors, the patient group was divided into four phenotypic groups (Table 3.1; Steyn 2002; Warnich *et al.* 2002).

Table 3.1 Phenotypic classification of the VP patient group.

	Phenotypic classification				
Patient number	Asymptomatic	Skin symptoms	Acute attacks	Acute attacks + skin symptoms	Total
	5	5	5	4	19

All 19 patients were previously characterized as heterozygous for the R59W mutation in the *PPOX* gene (Warnich *et al.* 1996b). Additionally, this patient group has also formerly been screened for:

1) genetic variants in the exonic and intronic regions of the δ-aminolevulinate synthase-2 gene (*ALAS2*), the porphobilinogen deaminase gene (*PBGD*) and the *ALAS1* exonic- and splice-site regions (Steyn 2002),
2) three known variants in the 5'UTR region of the *PPOX* gene (Warnich *et al.* 2002), and
3) polymorphisms in the core promoter, exonic- and intronic regions of the heme oxygenase-1 gene (*HMOX1*) (Zaahl 2006, unpublished data).

The control population comprised of 90 unrelated individuals of Mixed Ancestry (also termed "coloured population", as described in Nurse *et al*. 1985), with no family history of VP. The study was approved by the Ethics Review Committee for Human Research, Stellenbosch University (ref no 95/126).

DETAILED EXPERIMENTAL PROCEDURES

3.2 NUCLEOTIDE NUMBERING

The sequence for the P1 artificial chromosome clone (accession number AC006252) containing the human *ALAS1* genomic reference sequence and flanking regions, was retrieved from the National Centre for Biotechnological Information (NCBI) Entrez Nucleotide Database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nucleotide) and used for all the references made to *ALAS1*. To maintain consistency in nomenclature and nucleotide numbering (given that *ALAS1* has multiple TSSs), the first base of the most 3' TSS was denoted as +1, with the nucleotide directly 5' to this, as -1 (Roberts and Elder 2001; Appendix 3.1).

3.3 MUTATION DETECTION

The 5' promoter region of the *ALAS1* gene was analysed for the presence of possible sequence variants. This area consisted of the -1286 bp to +6 bp *ALAS1* upstream region to encompass the core-, proximal- and a portion of the distal promoter, yielding a 1310 bp fragment [primer overhangs included (refer to section 3.3.1)]. Furthermore, two fragments in the upstream distal promoter region, each containing an ADRES element, were analysed (Podvinec *et al.* 2004). The most 5' ADRES element was termed ADRES1 (5'-GGGACTCCAGTGACCTCTCCTTGCACA-3') and contained in a 409 bp amplicon encompassing the region from position -21016 to -20607. The second ADRES element, termed ADRES2 (5'-GGGTGAGCTAAGTTCA-NN-TGTGCTGCCGTGACCT-3'), was contained in a 432 bp amplicon, encompassing the region from position -15686 to -16118.

3.3.1 Oligonucleotide primer design

In order to amplify these fragments, all primers (refer to Appendix 3.1 - 3.3) were designed according to the *ALAS1* reference sequence using Primer3 v0.2, a publicly available software program (Rozen and Skaletsky 2000; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi). Primers used for cloning were designed to include restriction endonuclease recognition sites and three base overhangs (indicated in Tables 3.2 & 3.3). As a result of this particular primer design, the 1310 bp promoter amplicons had *BglII* and

NheI restriction enzyme sites incorporated into their 5' and 3' ends respectively, while both the ADRES amplicons had *KpnI* and *NheI* restriction sites at their 5' and 3' ends respectively. These restriction sites were incorporated to facilitate the ligation of each ADRES amplicon to the 5' end of the *ALAS1* promoter fragment, and subsequent sub-cloning into an expression vector to perform *in vitro* promoter analysis in the event of possible variants in the ADRESs- and/or promoter fragments. The T_m for each specific primer pair was optimized and indicated in Tables 3.2 & 3.3. All primers were synthesized by Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa) unless stated otherwise.

3.3.2 DNA amplification

PCR reactions included 10 pmol of each primer, 0.5 U *Taq* polymerase enzyme (BIOTAQ™ DNA polymerase, Bioline Ltd., London, England), 0.2 mM of each dNTP (dATP, dCTP, dGTP, dTTP) (Bioline), 1 X *Taq* Buffer (Bioline), 50 ng gDNA and a specific concentration of $MgCl_2$ (Bioline) determined for each primer set (Tables 3.2 & 3.3). Each reaction was made up to a final volume of 25 µl with dH_2O. For every set of PCR reactions performed, a negative control (reaction without gDNA) was included to determine the likelihood of PCR contamination.

PCR amplifications were performed in a GeneAmp® PCR System 2700 thermocycler (Perkin Elmer Applied Biosystems, Warrington, WA, Great Britain) with individual annealing conditions for each primer-set as indicated in Tables 3.2 & 3.3. The general PCR cycling conditions implemented were: initial denaturation for 4 min at 94°C, 35 cycles of 30 s denaturation at 94°C, 45 s annealing at a temperature optimized for each primer set and an elongation step of 1 min at 72°C. This was followed by a final extension step of 6 min at 72°C and a cooling step at 4°C.

3.3.3 Agarose gel electrophoresis and PCR purification

Amplified PCR products were analysed on 1% (w/v) agarose gels (Appendix 1) with 1 X TBE buffer (Appendix 2.1) containing 0.05% (v/v) ethidium bromide (EtBr; Appendix 1). Loading Dye Solution (Fermentas Inc., Hanover, USA) (4 µl) was added to each PCR product and loaded into the wells to resolve and compare each specific fragment to the appropriate molecular size marker [100 bp (Generuler™; Fermentas) or 1 kb ladder (Generuler™; Fermentas or Hyperladder; Bioline)]. The products were subjected to electrophoresis at 120 V for 40 min after which the DNA fragments were visualized with an ultraviolet light (UV) transilluminator system and captured using the MultiGenius Image Capture System (Syngene, Cambridge, England). Although the PCR reactions were extensively optimized, so that following each PCR reaction, the nucleotides were

Table 3.2 Primers used for PCR amplification and DNA sequencing analysis of the *ALAS1* promoter.

Fragment name	Primer region	Primer (5' -> 3')	Annealing temperature (°C)	Fragment size (bp)	Analysis
ALAS1	-1286 to +6	∗Forward: AGTGCTAGCCAGGATCTCTTGGCTGAATG ∗Reverse: TGAAGATCTGGCACTCAAGTCGAGAAGTC	67 (PCR) or 55	**1310	PCR & Sequencing
ALAS1NPRF	-702 to +353	Forward: CCAAGTGCTCAGAGCTGTGG Reverse: AAGAAGGCAGTCCTCCGGCA	65 (PCR) or 55	1055	PCR & Sequencing
ALAS1R2	-600 to -580	Reverse: TGGCAGGATGCGGAAGAAGC	55	706	Sequencing
ALAS1NESTF	-702 to -682	Forward: CTCGCCTCGACCTTGCCATT	55	735	Sequencing
ALAS1RB	+333 to +353	Reverse: AAGAAGGCAGTCCTCCGGCA	55	1055	Sequencing

*Underlined sequences indicate the restriction endonuclease recognition sites of *Nhe1* (5'G↓CTAGC3') and *BglII* (5'A↓GATCT3').

** Fragment size incorporates primer overhangs.

Table 3.3 Primers used for PCR amplification and DNA sequencing analysis of *ALAS1* ADRES.

Fragment name	Primer region	Primer (5' -> 3')	Annealing temperature (°C)	Fragment size (bp)	Analysis
ADRES1	-21016 to -20997	*Forward: AGTGGTACCGCTGGAAGTCAGGCTTGTGT	64	409	PCR & Sequencing
	- 20626 to -20607	*Reverse: TGAGCTAGCCTCTAACCTTGAACTCCTGG	64		
ADRES2	-16118 to -16099	Forward: AGTGGTACCGCAGTCAGACAGGCCTAAGC	64	432	PCR & Sequencing
	-15707 to -15686	Reverse: TGAGCTAGCCACACAGGAGGACACAGAGG	64		
ADRES2BF	-16030 to -16011	Forward: GGTGGAGAAGACAGGCTTGG	55	354	Sequencing

*Underlined sequences indicate the restriction endonuclease recognition sites of *NheI* (5' G↓CTAGC 3') and *KpnI* (5' GGTAC↓C3 ').

exhausted and the minimal amount of primer dimer formation was generated, all amplicons were purified to avoid the interference of primer dimers during semi-automated DNA sequencing. Only PCR products that demonstrated sufficient amplification without non-specific background, were purified with a column-free nucleic acid recovery kit according to manufacturers' protocol (Quick Clean, Bioline) for subsequent semi-automated DNA sequencing analysis.

3.3.4 DNA quantification

The concentration of the purified amplicons was determined using a spectrophotometer (NanoDrop® ND-100, Nanodrop Technologies Inc., DE, USA) that measures the absorbance of DNA at 260 nm. The A260/A280 and A260/A230 ratios were used to estimate the purity of nucleic acids. Purified PCR products, and all other DNA preparations quantified in this project, with an OD_{260}/OD_{280} ratio in the range of 1.8-2.0 and OD_{260}/OD_{230} at 2.2, were regarded as pure and protein-free.

3.3.5 Semi-automated DNA sequencing analysis

Purified PCR products with concentrations equal to or more than 10 ng/µl were subjected to DNA sequence analysis. In order to obtain bidirectional results of the whole amplicon for each sample, the sequencing reactions were performed in overlapping fragments using the PCR primers designed, as well as nested primers created specifically for sequencing purposes (Tables 3.2 & 3.3). Cycle sequencing was executed as $1/8^{th}$ of the manufacturer's recommended reactions with 1 µl of the BigDye® Terminator v3.1 Cycle Sequencing mix (Applied Biosystems, Warringon WA, Great Britain), 3.3 pmol primer and 10 ng purified PCR product in a 7 µl total reaction volume. A PCR thermocycler was used to perform the cycling reactions with the program as follows: initial denaturation at 95°C for 2 min, followed by 25 cycles of denaturation at 95°C for 10 s, annealing at 55°C for 10 s and an extension step at 60°C for 4 min. The products were analysed by electrophoresis on an automated ABI 3130XL Genetic Analyser (Applied Biosystems).

The electropherograms were analysed using BioEdit Sequence Alignment Editor software version v6.0.7 (Hall 1999) and the corresponding sequences aligned with the published reference sequence (refer to section 3.2) using the ClustalW program (version 1.4) in the BioEdit program. Intense visual inspection was also employed to evaluate each sequence for differences to the reference.

3.3.6 Restriction Fragment Length Polymorphism (RFLP) analysis

The variants identified by DNA sequence analysis were subjected to RFLP analysis in order to identify the different alleles of each polymorphism in the control population. PCR amplifications were performed with primers designed to encompass the variant site (Tables 3.4 & 3.5). The originally designed forward primer (ALASMF1) was later substituted with another forward primer (ALASMF2), designed to create a shorter PCR fragment of 114 bp, to obtain improved gel resolution and more distinct genotyping results (Table 3.4). PCR products were digested overnight in a water bath set to the optimal temperature for each enzyme (Tables 3.4 & 3.5). The -853C/T variant was detected by digestion with the *AclI* restriction enzyme, while the -1253T/A variant was detected by digestion with the *BsaI* restriction enzyme. For each variant, a PCR sample of a patient, confirmed by direct sequencing as a heterozygote for the variant, was included as positive control in each digestion series. A negative PCR control sample (without DNA) was also digested and included as a negative control. Each reaction consisted of 6 U of the appropriate enzyme (New England Biolabs Inc., Beverly, USA), 1 x reaction buffer (New England Biolabs), 5 µl PCR product and dH_2O to a final volume of 20 µl.

Loading Dye Solution was added to each digested PCR product and fractionated on either 2% (w/v) agarose gels (as per section 3.3.3) or on 9 cm vertical 15% (w/v) non-denaturing polyacrylamide gels [40% (w/v) gel stock mixture (Appendix 2.2), 5 X TBE, 10% (w/v) APS (Appendix 1) and 0.1% (v/v) TEMED (Appendix 1)] in a Mighty Small system (Hoefer Pharmacia Biotech Inc., California, USA). Once polymerised, the gel was placed in the tank of the electrophoresis system and 15 µl of the digested PCR products were loaded into the wells. Electrophoresis was performed in 1.5 X TBE running buffer for 90 min at 200 V at room temperature. All gels were stained in an ethidium bromide solution (1.5 X TBE and 2 µg/ml EtBr), visualized under UV light and captured as described in section 3.3.3.

3.3.7 Statistical Analysis

The allele- and genotype frequencies of the detected variants were calculated in the control group and tested for departure from the Hardy-Weinberg equilibrium (HWE) by means of a chi-square-goodness-of-fit-test with Tools for Population Genetic Analysis (TPGA) version 1.3 (Miller 1997).

3.4 *IN SILICO* PROMOTER ANALYSIS

The 1.292 kb *ALAS1* promoter sequence was analysed for putative TF binding sites by searching the literature and using commercially available web-based promoter analysis software programs. The three different *ALAS1* promoter alleles were analysed with the following predictive programs: TRES (http://www.bic.nus.edu.sg.:888/tres) (Katti *et al.* 2000), Matinspector (http://www.genomatix.de/matinspector) (Quandt *et al.* 1995), Alibaba2 (http://www.gene-regulation.com/pub/programs/alibaba2) (Grabe 2002), Signal Scan (http://bimas.dcrt.nih.gov/molbio/signal) (Prestridge 1991; 1996) and Match (http://www.gene-regulation.com/pub/programs.html#match) (Kel *et al*. 2003).

Species comparison of the -21 kb *ALAS1* 5' region was performed using the Multi-LAGAN program (http://lagan.stanford.edu/lagan_web/index) (Brudno *et al.* 2003) to identify phylogenetically conserved motifs of possible evolutionary importance, with particular focus on the regions harbouring the variants identified. The *ALAS1* gene promoter sequences was retrieved from the genomic contigs of the human (AC_006252), chimp (NW_001232823), mouse (NT_039477), rat (NW_001084876), chick (AF_536192), cow (ENSBTAG00000004118), dog (ENSCAFG00000009881) and pufferfish (NEWSINFRUG00000127688) from the NCBI Entrez Nucleotide Database or Ensembl genome database (version 40 – Aug 2006).

Table 3.4 Primers used for *ALAS1* -853C->T polymorphism genotyping.

Primer name	Primer (5' -> 3')	Annealing temp (°C)	Amplicon (bp)	*Restriction enzyme	Digestion products (bp)
ALAS1MF1 or ALASMF2	Forward: GAGACTCACCTCACCTGCCT Forward: GCAAGGAAGCCAGACTGACA	63	342 (MF1&MR)	*AclI*[1]	C allele: 321 & 21; T allele: 342
ALAS1MR	Reverse: CCTCCATGCCAAAGGTCAAC	63	114 (MF2&MR)	*AclI*[1]	C allele: 93 & 21; T allele: 114

Table 3.5 Primers used for *ALAS1* –1253T>A polymorphism genotyping.

Primer name	Primer (5' -> 3')	Annealing temp (°C)	Amplicon (bp)	*Restriction enzyme	Digestion products (bp)
Sp1F	Forward: AGTGCTAGCCAGGATCTCTTGGCTGAATG	65	137	*BsaI*[2]	A allele: 90 & 47; T allele: 137
Sp1R	Reverse: AGGTGAGGTGAGTCTCCACCTTCAGAGGA	65	137	*BsaI*[2]	

***Restriction endonuclease recognition sites and digestion conditions:**

1. -853C>T, *AclI:* 5' AA↓CGTT 3', Digest with NEBuffer 4 and 0.01% BSA (New England Biolabs) at 37°C.

2. -1253A>T, *BsaI*: 5' GGTCTCN↓ 3', Digest with NEBuffer 3 (New England Biolabs) at 50°C.

3.5 CONSTRUCTION OF PROMOTER-LUCIFERASE REPORTER GENE CONSTRUCTS

3.5.1 Cloning into pGem®-T Easy Vector

3.5.1.1 Amplification and purification

The 1.292 kb *ALAS1* promoter fragment of two DNA samples, one containing the wild-type (wt) promoter and one containing the two identified variants (-853C/T; -1253T/A), located adjacent to the putative ERE- and Sp1-TFBS (see section 4.1.4), was amplified by PCR using high-fidelity *Taq* DNA polymerase with 3'-5' proofreading activity (ExTaq™, Takara Shuzo Co., Ltd., Kyoto, Japan). The PCR products were gel-purified using a PCR and Gel purification kit (Wizard®SV Gel and PCR Clean-Up System, Promega Corporation, Madison, WI, USA). In brief, from each sample, five PCR reactions were pooled (100 µl), mixed thoroughly with an equal volume of membrane binding solution (consisting of 4.5 M guanidine isothiocyanate, 0.5 M $C_2H_3KO_2$, pH 5.0) and incubated at room temperature for 1 min. A SV minicolumn was assembled into a collection tube. The prepared PCR product was placed onto the column assembly and incubated at room temperature. After 1 min incubation, the assembly was centrifuged at 16 000 x g for 1 min (Beckman GS-15 centrifuge, Beckman Instruments Inc., Palo Alto, California). The flow-through was discarded and 800 µl membrane wash solution [consisting of 10 mM $C_2H_3KO_2$, 16.7 µM EDTA, 80% EtOH (Appedix 1)] was added onto the column, and centrifuged for 1 min. A subsequent wash step with 500 µl membrane wash solution and 5 min centrifugation was performed. The flow-through was discarded after each wash step, and the column was positioned into a sterile 1.5 ml microcentrifuge tube. The DNA was eluted by incubation (5 min) with 15 µl dH_2O and a final 1 min centrifugation step (16 000 x g).

3.5.1.2 A-tailing

Since the proofreading enzyme ExTaq™ DNA polymerase creates blunt-ended fragments during PCR amplification, an A-tailing procedure was performed to add a single deoxyadenosine (dATP) to the 3´-ends of the amplified fragments, enabling ligation into the pGem®-T Easy Vector (Promega). As described by the pGem®-T Easy Vector (Promega) cloning kit, the 10 µl A-tailing reaction consisted of 5.9 µl purified PCR product, 1 x PCR reaction buffer (Mg-free) (Bioline), 50 mM $MgCl_2$ (Bioline), 1 mM dATP (Bioline) and 0.5 U *Taq* DNA polymerase enzyme (Bioline). Each reaction was incubated for 30 min at 70 °C.

3.5.1.3 Ligation to pGem®-T cloning vector

The promoter fragments were each ligated to the pGem®-T Easy cloning vector (Appendix 4.1) using the pGem®-T Easy Cloning kit (Promega). According to the manufacturer's protocol, to achieve an 1:1 insert:vector ratio, 3 µl of the A-tailing reaction was used in a 10 µl ligation reaction, in addition to 2 X ligation buffer (Promega) and 400 U of the enzyme T4 DNA Ligase (Promega). A positive control (insert provided with the kit) and negative ligation (dH_20) reactions were included and all ligation reactions incubated overnight at 4°C.

3.5.1.4 Preparation of competent cells

The calcium chloride protocol from Hanahan (1983) was adapted and used to prepare competent cells for transformation purposes. A single *E. coli* colony (strain DH5-α) was used to inoculate 5 ml LB medium (Appendix 2.3) and incubated overnight on a shaker rotating at 225 rpm at 37 °C. After 12 h incubation the overnight bacterial culture was used to inoculate 500 ml LB medium and cultured at 37°C whilst shaking. The optimal density (OD) was measured at 600 nM and when between 0.4-0.6, the culture was centrifuged in a swing-out rotor centrifuge (Unicen 20, Orto Alresa, Madrid) to pellet the cells (5000 x g, 10 min, 4°C). The cells were kept on ice and gently resuspended in 100 ml Buffer 1 (100 mM $MgCl_2$, 4°C). The resuspended cells were incubated on ice (30 min, 4°C) and centrifuged for 10 min (4000 x g, 4°C). The supernatant was discarded and cells resuspended in Buffer 2 (100 mM $CaCl_2$, 15% v/v glycerol). Competent cells were aliquoted and snap-frozen at -80 °C in pre-chilled microcentrifuge tubes (PLASTIBRAND®).

3.5.1.5 Transformation

Competent *E. coli* (DH5-α) cells (prepared as mentioned above) were thawed on ice and used for transformations. Transformation reactions were performed according to the specifications in the manual of the pGem®-T Easy cloning kit, with a few modifications. For each transformation, 5 µl of the ligation reaction was added to 100 µl competent cells. As negative controls, 1 µl unligated vector or 5 µl dH_2O was added to 100 µl competent cells. All reaction mixtures were gently flicked and incubated on ice for 30 min. The cells were heat-shocked for 45 s in a 50°C water bath and placed on ice for 2 min. Subsequently, 800 µl pre-warmed (37°C) LB-medium (Appendix 2.3) was added to the transformation reactions and incubated on a platform shaker (Innova 2100, New Brunswick Scientific Co., Inc., Edison, USA) at 225 rpm for 1 h at 37°C. From each transformation reaction, 200 µl was plated onto a pre-warmed LB agar plate (Appendix 2.4), supplemented with 50 µg/ml Amp, 40 µg/ml X-gal (Fermentas) and 0.2 mM IPTG (Fermentas) to function as selective pressure. The plates were incubated overnight at 37°C to facilitate colony formation.

3.5.1.6 Clone selection

Following incubation, blue and white colonies were visible on the plates. The white colonies, indicative of vectors containing inserts, were subjected to a colony screen. Briefly, a sterile needle was used to pick a colony, spread onto a pre-warmed masterplate (LB agar Amp/X-gal/IPTG) after which the needle was used to inoculate a PCR reaction mixture. The colony-PCR reaction consisted of the general PCR reagents (as per section 3.3.2) with 10 pM of the primer pair (M13F & M13R) flanking the multiple cloning site (MCS) of the pGem®-T Easy Vector (Table 3.7). Amplification products of the colony-PCR were assessed on an 1% (w/v) agarose gel. The colonies indicating successful amplification of the cloned fragment were picked with a sterile needle from the masterplate, and inoculated into 10 ml LB medium supplemented with 50 µg/ml Amp. The inoculated medium was incubated overnight at 37°C, on a platform shaker at 225 rpm.

3.5.1.7 Small-scale isolation of plasmid DNA (Miniprep)

After proper growth, plasmid DNA was extracted from the 10 ml overnight culture using a modified manual plasmid extraction procedure for high copy number plasmids adapted from Sambrook *et al*. (1989). In the first step, 2.5 ml of the overnight culture was collected, centrifuged at 5 000 x g and the supernatant discarded. The cells were resuspended in 400 µl lysis buffer [50 mM Glucose (Appendix 1), 25 mM Tris-HCL (Appendix 1), 10 mM EDTA (Appendix 1)] supplemented with 10 mg/ml Lysozyme (Roche Diagnostics GmbH, Mannheim, Germany) and 10 mg/ml RNAse A (Roche). NaOH/SDS [0.2 N NaOH (Appendix 1); 1% SDS (Appendix 1)] was freshly prepared, 400 µl was added to the mixture and incubated on ice for 5-10 min. To remove impurities, the lysate was treated with 300 µl of 7.5 M NH_4OAc (pH 7.6), mixed by inversion and incubated on ice for 5-10 min. Centrifugation at room temperature for 5 min at 14 000 x g were performed and the supernatant transferred to a new clean 1.5 ml microcentrifuge tube. Isopropanol (0.6 vol) was added to the supernatant and incubated for 10 min at room temperature after which a centrifugation step at 14 000 x g for 10 min followed. The supernatant was discarded and the pellet was resuspended in 100 µl 2 M NH_4OAc (pH 7.4). After a 5 min incubation period on ice, the suspension was centrifuged at 10 000 x g for 10 min, and the supernatant placed into a new microcentrifuge tube. Isopropanol (100 µl) was added to the supernatant, mixed by inversion and incubated at room temperature for 10 min. Following centrifugation at 10 000 x g for 10 min, the pellet was washed with 1 ml 70% ethanol and left to dry. The pellet was resuspended in 30 µl nuclease-free H_2O and the concentration and quality of the pGem®T-promoter plasmid DNA determined (as per section 3.3.4). The resultant constructs were designated, pGEM(ALAS-WT) and pGEM(ALAS-ERE/Sp1).

Table 3.6 Primers used for PCR-mediated mutagenesis.

Primer name	Primer (5' -> 3')	Fragment size (bp)
*ALASMut	Forward: AGTGCTAGCCAGGATCTCTTGGCTGAATGGCACAGAACTGGGTCTCG	**1310

* Primer synthesized by Integrated DNA Technologies, Inc (IDT). Underlined sequence indicate restriction endonucleases recognition site for *NheI* (5 'G↓CTAGC 3').

** Fragment size of PCR product when used with ALASR primer.

Table 3.7 Primers used for clone selection.

Primer name	Primer (5' -> 3')	Annealing temperature (°C)	Plasmid
T7	Forward: TAATACGACTCACTATAGGG	58	pGemT
Sp6	Reverse: TATTTAGGTGACACTATA G	58	pGemT
M13F	Forward: GTTGTAAAACGACGGCCAGT	58	pGemT
M13R	Reverse: CAGGAAACAGCTATGACC	58	pGemT
GL1	Forward: TGTATCTTATGGTACTGTAACTG	55	PGL2-Basic
GL2	Forward: CTTTATGTTTTTGGCGTCTTCCA	55	PGL2-Basic

3.5.2 Sub-cloning in pGL2-Basic luciferase reporter vector

3.5.2.1 Digestion and purification

The pGEM(ALAS-WT), pGEM(ALAS-ERE/Sp1) constructs and the luciferase-reporter vector, the pGL2-Basic plasmid (Promega) (Appendix 4.2), were each digested overnight in a 37˚C water-bath in double digestion reactions with *NheI* (New England Biolabs) and *BglII* restriction endonucleases (New England Biolabs). Each digestion reaction contained 10 U of each enzyme, 1 X buffer 2 (New England Biolabs), 0.01% BSA (New England Biolabs) and 2 µg of DNA, made up to a final volume of 20 µl with dH_2O.

The digested pGem-constructs were resolved on a 1% (w/v) agarose gel at 80 V for 60 min until the inserts (*ALAS1* promoter fragments) were visible under the UV light of the transilluminator system (Spectroline®, Spectronics Corporation). Each promoter fragment was excised from the gel with a sharp, sterile scalpel and purified with a PCR and Gel Purification kit (Promega Wizard®SV Gel and PCR Clean-Up System) as described in the Promega Technical Bulletin #TB308 and section 3.5.1.1. The concentrations of the purified fragments were determined using a spectrophotometer (as per section 3.3.4).

3.5.2.2 Dephosphorylation of pGL2-Basic vector

To prevent circularization of the digested pGL2-Basic luciferase reporter vector during ligation, the hydrolysis of the 5'-terminal phosphate residue (dephosphorylation) was performed using the enzyme shrimp alkaline phosphatase (SAP) in 1 x SAP dephosphorylation buffer (Roche) for 1 h at 37°C. Afterwards, the enzyme was heat-inactivated by incubating the sample for 10 min at 65°C. The SAP-treated vector was purified (Promega Wizard®SV Gel and PCR Clean-Up System, as per section 3.5.1.1) and the concentration determined using the spectrophotometer (as per section 3.3.4).

3.5.2.3 Ligation into pGL2-Basic vector

The two digested purified *ALAS1* promoter alleles (WT and ERE/Sp1) were ligated (3:1, insert:vector) into the pGL2-Basic vector, a promoter-less plasmid containing a MCS upstream of a luciferase gene (*Luc*). The pGL2-Basic vector was chosen as reporter plasmid to facilitate future comparisons to another study by our group, where it was used for the functional analysis of the *PPOX* gene promoter (Kimberg 2007, unpublished).

The protocol for the ligation reactions were performed as described by Ausubel (1997). In short, the 20 µl ligation reaction consisted of 70 ng insert DNA, 100 ng pGL2-Basic vector DNA, 1 x T4 DNA ligase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 1

mM ATP; 25 µg/ml BSA; New England Biolabs) and 400 U T4 Ligase (New England Biolabs). The reaction was incubated at 16˚C for 5 h.

3.5.2.4 Transformation

Competent DH5-α *E. coli* cells (prepared as per section 3.6) were transformed with the pGL2-Basic/*ALAS1*-promoter ligation products. The transformation procedure was performed as described in section 3.5.1.4 and the transformed cells were spread onto pre-warmed (37˚C) LB agar plates (Appendix 2.4) supplemented with 50 µg/ml Amp. Colony-PCRs, consisting of the general PCR reagents with a primer pair flanking the MCS of the pGL2-Basic vector (GL1 and GL2, Table 3.7), were used to discriminate between positive colonies and false positives due to vector re-ligation. Amplification products were assessed by electrophoresis on a 1% (w/v) agarose gel. The colonies indicating positive amplification were picked from the master plate and inoculated into 60 ml LB medium supplemented with 60 µg/ml Amp.

3.5.2.5 Large-scale endotoxin-free plasmid extraction (Maxiprep)

After overnight agitation at 225 rpm at 37˚C, 10 ml of the bacterial culture was used to prepare glycerol stocks: 1/10 volume 100% glycerol and 9/10 volumes of the bacterial culture were mixed and stored at -80°C. The remaining 50 ml of the bacterial culture was used for plasmid extractions with an endotoxin-free plasmid extraction kit without any deviations from the manufacturer's protocol (Promega, PureYieldTM Plasmid Midiprep System). The concentration of each extracted plasmid DNA sample was determined (as per section 3.3.4). The correct nucleotide sequence and orientation of the promoter fragments (inserts) were verified by DNA sequence analysis (as per section 3.3.5) with the primers GL1 and GL2 (Table 3.7). The resultant pGL2-Basic luciferase constructs were termed: pGL2(ALAS-WT) and pGL2(ALAS-ERE/Sp1).

3.5.3 Primer-mediated PCR mutagenesis

The pGemT(ALAS-ERE/Sp1) construct was used as a template to create an *ALAS1* promoter fragment containing only the sequence variant next to the ERE half-site (discussed in section 4.1.4). Since the detected variant adjacent to the putative Sp1-site was located within 14 bp (downstream) of the forward primer, a primer-mediated mutagenesis technique was used to replace this variant with the wt nucleotide. A 47-mer primer (ALASMut; Table 3.6), incorporating the original ALAS1F primer (Table 3.2) along with an extra 14 bp, to include the "Sp1-variant", was used for this technique. This primer consisted of the wild-type nucleotide at the Sp1-variation site, and was used in a PCR amplification reaction to generate the *ALAS*1 promoter fragment from the pGem-promoter construct (pGemERE/Sp1). The resultant PCR fragment, containing only the ERE mutation, was cloned into the pGemT-

Easy cloning vector (as per section 3.5.1) and sub-cloned into the pGL2-Basic vector (as per section 3.5.2). The resultant constuct was termed pGL2(ALAS-ERE).

3.6 CELL CULTURE, TRANSIENT TRANSFECTIONS AND GENE EXPRESSION ASSAYS

3.6.1 Cell culture

The HepG2 human hepatocarcinoma liver cell line from the American Type Culture Collection (ATCC, Rockville MD, USA) (obtained from Dr K Robson, WIMM, UK) was used to establish the effect of the various *ALAS1* promoter alleles on transcriptional activity. The HepG2 cells were cultured in 75 cm^2 sterile polystyrene culture flasks (Cellstar®, Greiner Bio-one, Frickenhausen, Germany) in 25 ml Dulbecco's Modified Eagles Medium with 4.5 g/L glucose and L-Glutamine (DMEM, Cambrex Bio Science, Walkersville, Inc., USA) fully supplemented with 10% fetal calf serum (v/v) (Sigma-Aldrich Co., St Louis, MO, USA) and 100 U/ml penicillin and streptomycin solution (Sigma). The serum provided cells of various supplements such as growth factors and hormones to assist in cell growth and attachment, while the antibiotics protected against infections. Flasks were kept in an incubator (Heraeus Cell 150, Kendro Laboratory Products, USA) set at 37˚C with an atmosphere of 5% CO_2. Growth medium was replaced with fresh, fully supplemented medium as needed (approximately every four days) in a sterile hood using sterile syringes.

Mycoplasma is a cell wall-less bacterium resistant to Penicillin and Streptomycin. Although *Mycoplasma* infection cannot be detected by visual inspection and may not noticeably affect cell culture growth rates, it may alter DNA, RNA and protein synthesis. To prevent the possibility of *Mycoplasma* contamination, the cells were routinely treated with the prophylactic bactericidal antibiotic Plasmocin™ (5 µg/ml), as this antibiotic does not to influence cell metabolism or transfection efficiency (http://www.plasmocin.com; Lieto *et al*. 2003). Cells were also frequently screened for *Mycoplasma* presence using a PCR-based detection kit (Mycoplasma Plus™ PCR Primer Set, Stratagene, La Jolla, CA, USA).

When cells were 80% confluent, they were passaged by removing the culture medium and rinsed with Hank's Balanced Salt solution (Cambrex Bio Science). To facilitate cell detachment, 1 ml trypsin-versene mix (Cambrex Bio Science) was added to the flask and replaced in the incubator (37˚C; 5% CO_2) for 5 min. The cells were resuspended in 10 ml fully supplemented DMEM, transferred to a 15 ml sterile tube (Cellstar®, Greiner Bio-One) and harvested by centrifugation at 700 x g for 5 min. The supernatant was discarded and the cells resuspended in 5 ml fully supplemented DMEM. Depending on the confluence and

growth-rate of the cells, the cell suspension was seeded to four flasks for a 1:4 split, and returned to the incubator.

It is recognized that continuous passaging prompt some cell lines to retain only their vital cellular functions. Subsequently, even though early passages of a HepG2 cell line contain ERα, the cells gradually lose their ability to express ERα (Knowles *et al*. 1980; Tam *et al*. 1985; Mao and Shapiro 2000). Consequently, in this project, HepG2 cells at a late passage stage, were selected for all transfection experiments, not only because these cells retain many properties of the hepatocytes from which they are derived from, but also because they express limited amounts of ERα, allowing them to be used as ERα-negative cells in co-transfection experiments.

3.6.2 Transfections, co-transfections and estrogen induction

A series of transfections were performed (discussed in section 4.3) in order to analyse the transcriptional response of the *ALAS1* promoter (Figure 3.1 and Tables 3.8 to 3.12). Transient transfections were performed using FuGENE® 6 Transfection Reagent (Roche) according to manufacturer's instructions. Twenty-four hours before transfection, confluent HepG2 cells were plated into 35 mm^2 culture dishes (Cellstar®). Briefly, the medium was removed from the 75 cm^2 culture flask, rinsed with Hank's Balanced Salt solution and 10 ml DMEM (fully supplemented) added. The cells were detached by gentle scraping with a glass pipette. The cell suspension was transferred to a 15 ml sterile tube, centrifuged at 700 x g for 5 min and resuspended in fully supplemented DMEM. The resuspended cells were transferred to 35 mm^2 culture dishes (3 ml/dish) and incubated overnight at 37˚C in 5% CO_2. The cells were plated in the 35 mm^2 culture dishes to ensure 50-80% confluent, freshly dividing cells on the day of the transfection.

3.6.2.1 Transfections

Five hours before transfection the culture medium was aspirated from the 35 mm^2 culture dishes and antibiotic-free DMEM, supplemented with 10% (v/v) fetal calf serum, added to the cells and returned to the CO_2 incubator (37˚C; 5% CO_2). All transfections were performed using FuGENE® 6 transfection reagent equilibrated to room temperature. Transfection reagent was mixed with DMEM in a 3:1 ratio (microlitre Fugene:microgram reporter construct) to the total amount of DNA. In short, 97 µl supplement-free DMEM at room temperature was placed in a sterile 1.5 ml microcentrifuge tube and 3 µl transfection reagent was added directly into the medium. Since chemical residues in plastic vials can significantly decrease the biological activity of the transfection reagent, FuGENE® 6 was carefully pipetted directly into the medium, without touching the sides of the tube. The mixture was incubated at room temperature for five min, after which the DNA was added

to include 1 µg of the *ALAS1* promoter reporter construct and 25 ng of the pSVβ-gal plasmid (provided by Dr MG Zaahl, Genetics Department, US). The pSVβ-Gal plasmid contains the bacterial β-galactosidase coding sequence under the control of the SV40 promoter and served as internal control for transfection efficiency.After a 30 min incubation period at room temperature, to allow the complexes to form, the 120 µl from each vial was added drop-wise to the cells in the appropriate culture dishes and incubated for 24 h at 37˚C in 5% CO_2 to allow transfection to occur.

3.6.2.2 Co-transfections and estrogen induction

Co-transfection experiments were performed as described by Menuet *et al*. (2004), with adaptions. In short, HepG2 cells were seeded into 35 mm^2 sterile culture dishes as described above. The fully supplemented medium was aspirated from the 35 mm^2 culture dishes 5 h prior to transfection and rinsed with 1 x PBS (Appendix 2.5). Because of the estrogen-like effects of phenol-red (Thompson *et al*. 1988), hormone-free medium (Phenol-Red-Free DMEM) (Sigma), supplemented with 5% Charcoal-treated stripped FCS (Highveld Biological, South Africa), was added to the cells. Charcoal-treated serum was used to exclude the likelihood of endogenous estrogen present in normal culture serum. The transfection procedure was performed as described in section 3.6.2.1, with some alterations: 1 µg of each promoter reporter construct were transfected together with 25 ng pSVβ-gal and 25 ng of pCDNA/ER, a human ERα expression vector (provided by Dr A Louw, Biochemistry Department, US). The 120 µl was added drop-wise to the cells in the appropriate culture dishes and incubated for 24 h at 37˚C in 5% CO_2 to allow transfection to occur.

Following the 24 h incubation period, cells were treated with the principal physiological form of estrogen, 17β-estradiol (E_2; Sigma), obtained in powder form and dissolved in 100% EtOH. Briefly, cells were washed once with 1 x PBS and treated with 3 x 10^{-6} nM E_2 or EtOH (vehicle) in phenol red-free medium without serum. The optimal concentration of estrogen stimuli in HepG2 cell culture was determined by a separate dose response experiment using an E_2 concentration series (described in section 4.3.5). According to these results, 3 x 10^{-6} were used as the optimal concentration for all E_2 inductions. The hormone treated cells were incubated for 24 h at 37˚C in 5% CO_2 before the Luciferase- and β-Galactosidase assays were performed.

For all quantification transfections and co-transfections, each experimental condition, i.e. construct type or incubation condition, triplicate transfections were performed, and each quantification assay (set) repeated separately at least three times to determine the significance and repeatability of the results.

Figure 3.1 Diagrammatic representation of the proposed transfection sets.

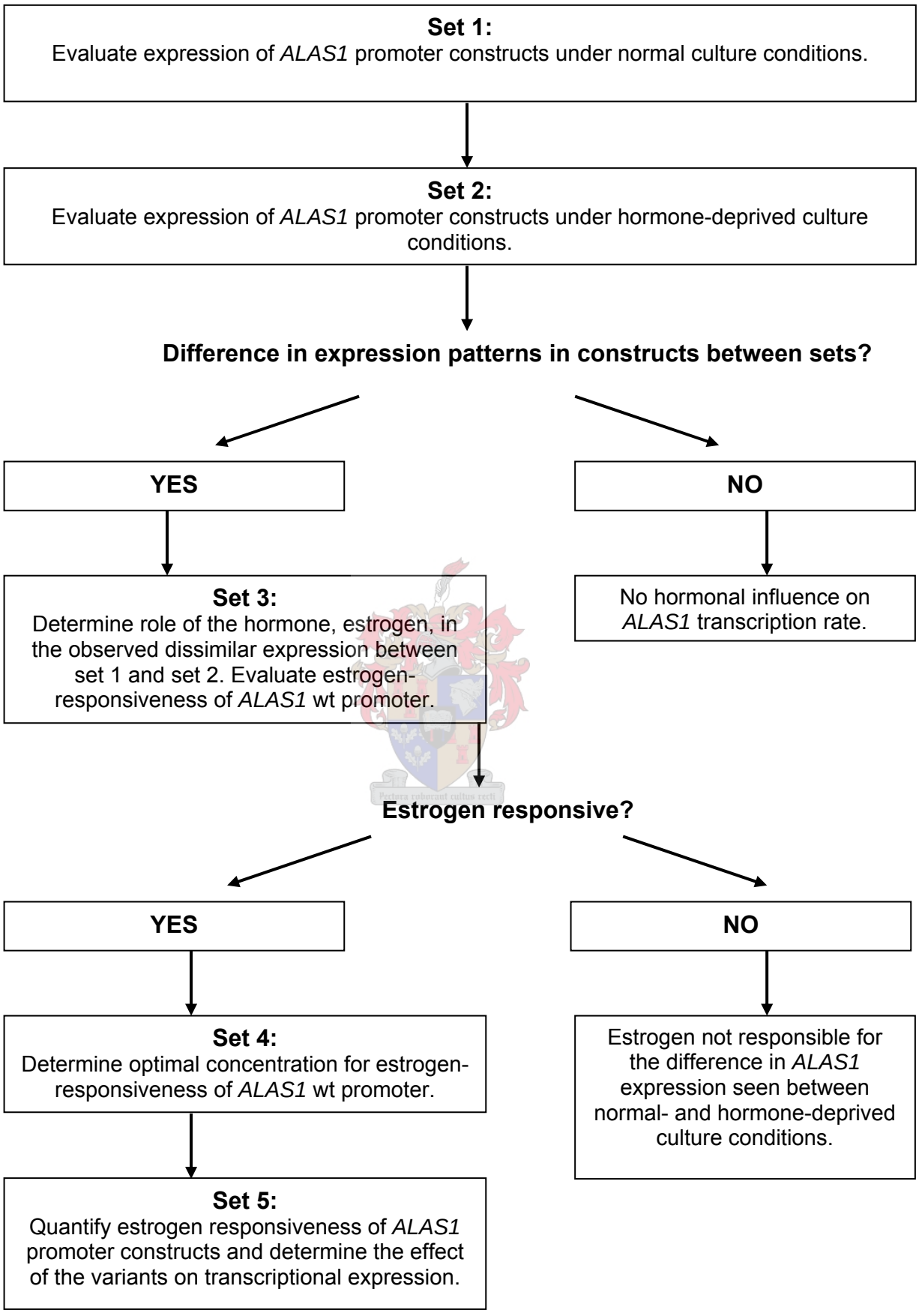


Table 3.8 Transfections of the *ALAS1*-promoter constructs in HepG2 cells under normal culture conditions.

Set 1: "normal" conditions	**1**	**2**	**3**	**4**	**5**	Repeat
Constructs	WT	ERE	ERE/Sp1	pGLvitERE	pGL2	X 3
	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	
Repeats	X 3	X 3	X 3	X 3	X 3	

Table 3.9 Transfections of the *ALAS1*-promoter constructs under hormone-free culture conditions.

Set 2: Hormone deprived medium	**1**	**2**	**3**	**4**	**5**	Repeat
Constructs	WT	ERE	ERE/Sp1	pGLvitERE	pGL2	X 3
	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	
Repeat	X 3	X 3	X 3	X 3	X 3	

Table 3.10 Transfection of the wild-type *ALAS1*-promoter construct in the presence or absence of estrogen-activated ERα.

Set 3: Hormone-free medium	**1**	**2**	**3**	**4**	**5**	**6**
Constructs	pGLvitERE	WT	pGL2	pGLvitERE	WT	WT
	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal
				pCDNA/ER	pCDNA/ER	pCDNA/ER
Stimuli	E_2 (10^{-8})	E_2 (10^{-8})	E_2 ($10^{-8)}$	E_2 (10^{-8})	E_2 (10^{-8})	EtOH
Repeat	X 3	X 3	X 3	X 3	X 3	X 3

Table 3.11 The dose-response of the wild-type *ALAS1*-promoter construct, co-transfected with ERα, in the presence of an estrogen concentration-series.

Set4: Hormone deprived medium	**1**	**2**	**3**	**4**	**5**	**6**
Constructs	WT	WT	WT	WT	WT	WT
	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal
	pCDNA-ER	pCDNA-ER	pCDNA-ER	pCDNA-ER	pCDNA-ER	-
Stimuli	E_2 ($3x10^{-6}$)	E_2 ($3x10^{-8}$)	E_2 ($3x10^{-10}$)	E_2 ($3x10^{-12}$)	EtOH	EtOH
Repeat	X 3	X 3	X 3	X 3	X 3	X 3

Table 3.12 Transfection of the *ALAS1*-promoter constructs, co-transfected with ERα, in the presence or absence of E_2.

Set5: Hormone deprived medium	**1**	**2**	**3**	**4**	**5**	**6**	**7**	Repeat
Constructs	WT	ERE	ERE/Sp1	pGL2	WT	ERE	ERE/Sp1	X 3
	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	pSVB-Gal	
	pCDNA-ER	pCDNA-ER	pCDNA-ER	pCDNA-ER	pCDNA-ER	pCDNA-ER	pCDNA-ER	
Stimuli	E_2 ($3x10^{-6}$)	E_2 ($3x10^{-6}$)	E_2 ($3x10^{-6}$)	E_2 ($3x10^{-6}$)	EtOH	EtOH	EtOH	
Repeat	X 3	X 3	X 3	X 3	X 3	X 3	X 3	

3.6.3 Luciferase- and β-Galactosidase Assay

Assays were performed 24 h after transfections or 24 h after hormone induction for co-transfections, in 96 well white-bottomed microtitre plates (Thermo Electro Co., Inc.). The detection of luciferase activity in cells, transfected with reporter vectors carrying the *Luc* gene from the North American firefly (*Photinus pyralis*), was performed by using the Luciferase Assay System (Promega) precisely according to the manufacturers' instructions (Promega Technical Bulletin no 281). The Luciferase assay is based on enzymatically catalyzed chemiluminescence where the firefly luciferase enzyme catalyses the oxidative decarboxylation of luciferin. This reaction is dependent on Mg^{2+} ions and ATP yielding photons as well as O_2, AMP and oxyluciferine. The firefly luciferase protein does not require posttranslational processing for enzymatic activity and can consequently function as a genetic reporter immediately upon translation. The amount of light emitted depends on the concentration of the luciferase enzyme, therefore, it allows the quantitative estimation of the expression level of the reporter gene.

Briefly, the medium was removed from the culture dishes and the cells washed once with 2 ml 1 x PBS. Sufficient 1 x reporter lysis buffer (Promega) was applied to each culture dish to completely cover the cells (250 µl). Dishes were incubated at -80˚C for 10 min until frozen. When thoroughly frozen, cells were thawed and harvested by gentle scraping with a sterile glass-pipette and transferred to a pre-cooled, sterile 1.5 ml microcentrifuge tube. All subsequent steps were performed on ice. After a brisk vortex, the contents were centrifuged at 12000 x g for 2 min at 4˚C, and the supernatant transferred to a sterile 2.2 ml microcentrifuge tube. The cell lysates were immediately used by adding 40 µl from each lysate into two separate wells of the 96 well plate for the luciferase- and β-galactosidase assays respectively.

Luciferase assay reagent (Promega) was prepared by adding the luciferase assay buffer (Promega) to the luciferase assay substrate (Promega) and equilibrated to room temperature. Luciferase activity was measured when the injector of the Veritas Microplate Luminometer (Turner Biosystems, Sunnyvale, California, United States), added 100 µl of the luciferase assay reagent (37˚C) to 40 µl of each cell lysate. Since the intensity of luminescence is constant for 20 s before it decreases, the relative light units (RLU) for the luciferase assays were counted by programming the luminometer to read the light emitted for a period of 10 s after a 2 s delay.

For the β-galactosidase assay, the Beta-Glo™ Assay System (Promega) was employed and performed exactly according to the manufacturers' instructions (Promega Technical Manual no 239). In short, the Beta-Glo™ reagent was prepared by adding the contents of the Beta-Glo™ assay buffer to the vial of Beta-Glo™ assay, gently mixed by inversion and equilibrated to

room temperature. After thorough mixing, 30 µl of the β-Glo™ reagent (37˚C) was added to 40 µl of each cell lysate. The mixture was incubated for 30 min at room temperature to allow the signal to stabilize after which the light emitted, was measured for 10 s.

Values obtained for luciferase activity directed by the reporter plasmids were normalized for variation in transfection efficiency by dividing it with the values obtained for β-galactosidase activity in the same culture dishes. The average value of the normalised results for each construct type in repeated experiments under specific conditions were determined to obtain the relative expression of each construct type for each variable (i.e. culture conditions).

3.6.4 Statistical analysis of reporter gene assays

The normalised values for reporter gene expression results were analyzed (SAS software, SAS Institute, Inc., Cary, NC) and only values within 20% of one another were used to determine arithmetic means, S.D. values, S.E. values, S.E. values of the means. The significance of differences among repeated measurements between multiple groups (sets) was determined by analysis of variance (ANOVA). This test examines the changes between groups to determine whether they are greater than would be expected by chance. In those cases in which significant differences ($p < 0.05$) were detected by ANOVA, determination of variance was followed by paired *t* tests, using the SAS software package and Microsoft Excel. Results were presented as means ± SEM or as a % of the reference (discussed in section 4.3).

CHAPTER 4:

RESULTS AND DISCUSSION

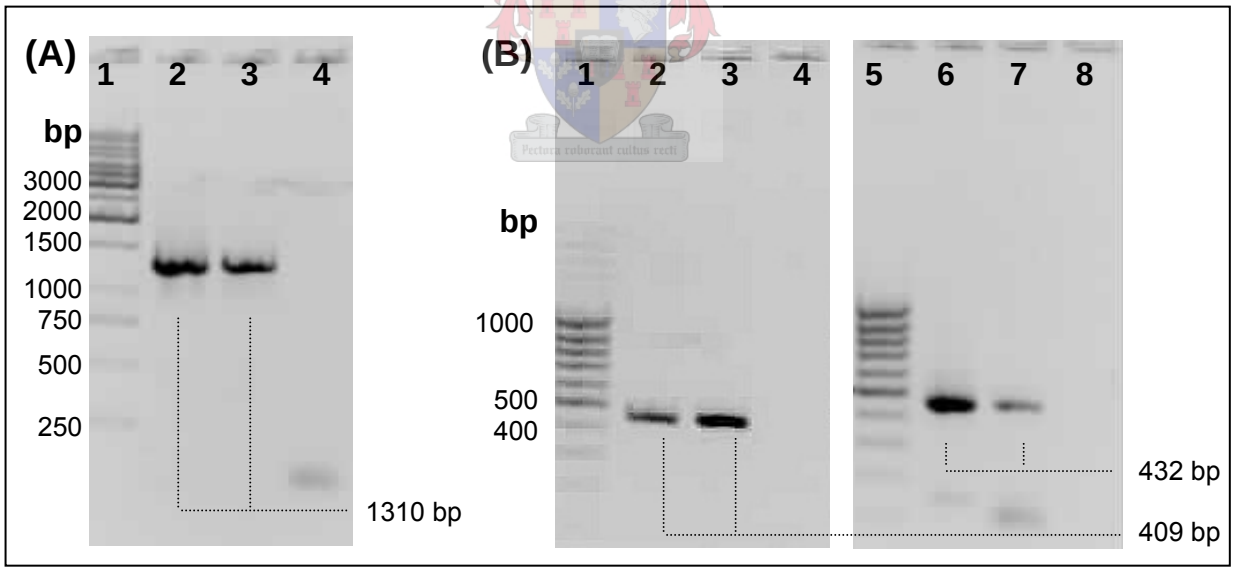
4. RESULTS AND DISCUSSION

4.1 *ALAS1* AS CANDIDATE MODIFIER GENE

4.1.1 PCR amplification

To assess the nature and extent of nucleotide variation within the proximal promoter and the distal ADRES elements in the 5' flanking region of the human *ALAS1* gene, PCR analyses were performed with primers designed based on the published sequence (Tables 3.1 & 3.3). Amplification of the -1286 bp to +6 bp region generated a 1310 bp fragment (including primer overhangs), while ADRES1 and ADRES2 PCRs produced fragments of 409 bp and 432 bp respectively. The resultant amplicons of 19 VP patient DNA samples, were verified and compared to a DNA molecular weight marker (1 kb or 100 bp) by agarose gel electrophoresis (Fig. 4.1). Results indicated that the amplified fragments obtained were of the expected size, good quality and sufficient quantity to perform sequence variation analysis. For every set of PCR reactions performed, no amplification product was observed in the negative control (without template DNA).

Figure 4.1 PCR amplicons of the *ALAS1* promoter- and the two ADRES fragments.



(A) An 1% agarose gel stained with EtBr (0.01%; v/v), indicating the *ALAS1* promoter amplicons in lanes 2 & 3 (1310 bp), compared to a 1 kb marker in lane 1 (Fermentas). A negative control was loaded in lane 4.
(B) An 1% agarose gel of the ADRES1 amplicon (lanes 2 & 3; 409 bp) and the ADRES2 amplicon (lanes 6 & 7; 432 bp). The fragments were compared to a 100 bp marker (Fermentas). The negative controls were loaded in lanes 4 and 8.

4.1.2 DNA Sequencing

The *ALAS1* promoter-, ADRES1- and ADRES2 amplicons of the 19 patients were screened by direct DNA sequencing using the primers indicated in section 3.3.5. Alignments of the sequencing results with the reference sequence clearly indicated amplification of the desired fragments. Chromatograms and sequences were subjected to intensive analysis (as described in section 3.3.5).

4.1.2.1 Variants detected

Comparisons of sequenced fragments with the reference sequence revealed two novel nucleotide substitutions: a C-to-T (C>T) substitution at position -853 and a T-to-A (T>A) substitution at position -1253. No sequence variants were observed in either of the two ADRES elements. Since sequence conservation in promoter regions is usually interpreted to be an indicator of functional significance (Crawford et al. 1999), the identification of no variants within the ADRES regions, suggests a strong preservation of these areas and the functional importance of the *ALAS1* regulatory regions. This confirms the expected sequence conservation of the drug-responsive regions by vital selective pressure to illustrate its important role in the regulation of haeme synthesis. An interesting article by Hoogendoorn *et al*. (2003) analysed 170 promoters representing 73 995 bases of proximal promoter sequence and revealed a total of 120 sequence variants (1 every 616 bases). Their findings however demonstrated that only around a third of promoter variants may alter gene expression to a functionally relevant extent. The identification of two variants in the patient group within the screened region (1292 bp), correlates to the frequencies as found by Hoogendoorn *et al*. (2003).

Chromatogram results indicated that three VP patients were heterozygous for both the C>T transition and the A>T transversion. Heterozygosity was clearly discernible on the chromatograms as double peaks. All three patients were female and members of the same extended family of Mixed Ancestry.

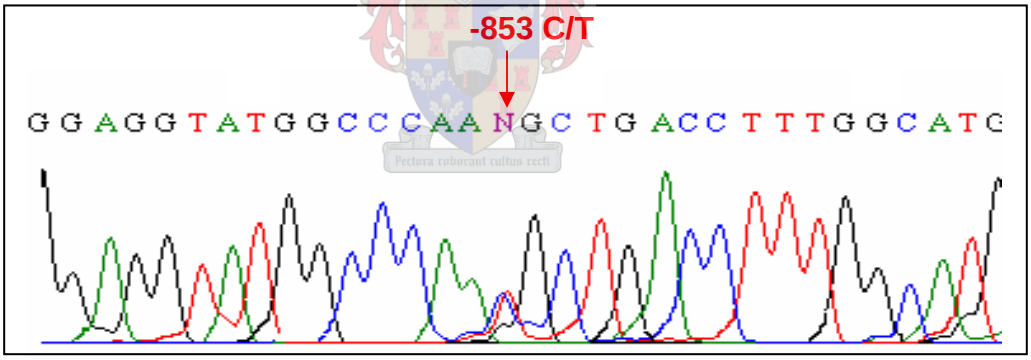
Genotyping analysis of the two identified variants was performed on members of the extended family of the heterozygous VP patients, and a population-matched control group to determine the prevalence of these genotypes in the Mixed Ancestry population in general and to establish whether the -853C>T and -1253T>A variants are situated on the same chromosome. A random subset of 90 healthy blood donors of Mixed Ancestry was used as a control population, and 31 family members of the heterozygous VP patients were available for screening.

i) -853 C>T Genotyping

PCR primers were initially designed to amplify a 342 bp region of the *ALAS1* 5' region to encompass the -853C>T transition (Table 3.4; Figure 4.2) with the reverse primer designed to create an *AclI* restriction site in the presence of the C (wt) allele. The correct sizes of PCR fragments were confirmed by comparison to a 100 bp molecular marker on 1% agarose gels. PCR products were digested with *AclI*.

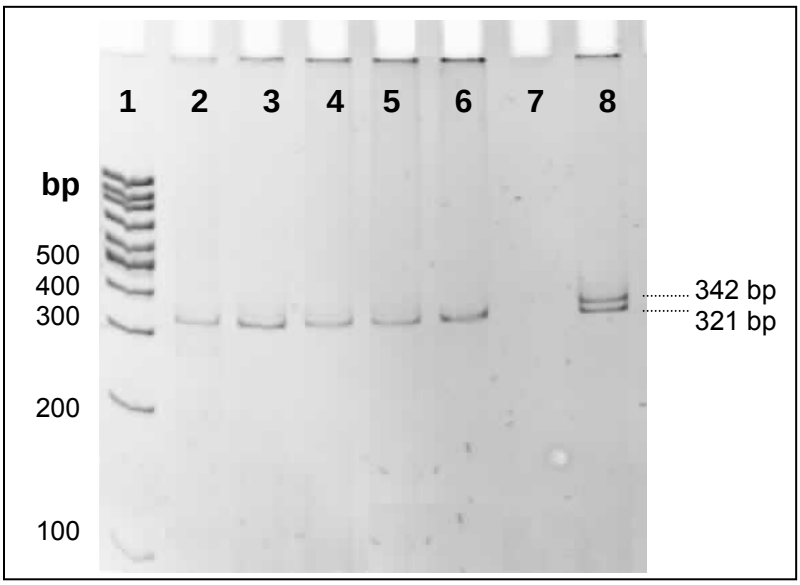
The digested 342 bp amplicons were subjected to polyacrylamide gel electrophoresis. As expected, digestion of the 342 bp PCR products yielded two fragments, 321 bp and 21 bp for individuals homozygous for the wt -853C allele, whereas individuals homozygous for the -853T/T genotype yielded one undigested band (342 bp) and heterozygous individuals (-853C/T) yielded all three bands (Figure 4.3A). The 114 bp digests produced two fragments (93 bp and 21 bp) for homozygotes with the wt -853C/C allele, one undigested band (114 bp) for -853T/T homozygotes, while heterozygotes produced all three bands (Figure 4.3B). The VP patient group, the population matched control group (frequencies indicated in Table 4.1) and the family members of affected individuals were genotyped.

Figure 4.2 DNA sequence chromatogram of a patient heterozygous for the *ALAS1* -853C>T variant.



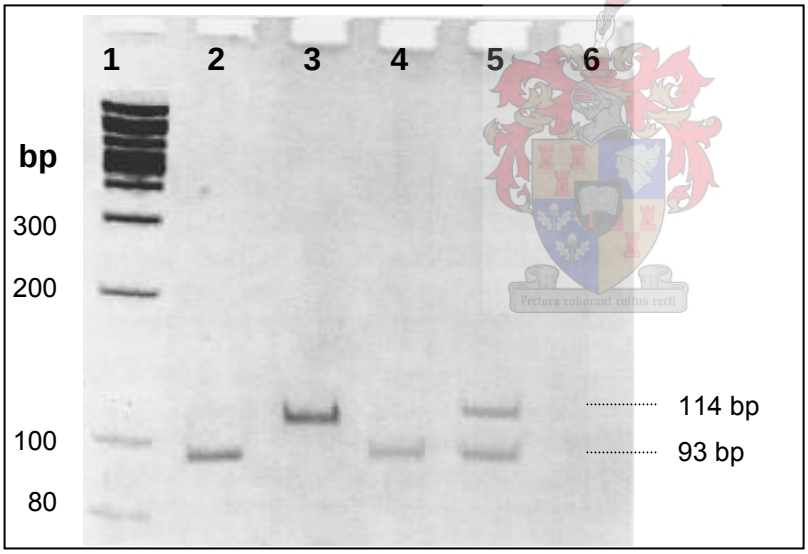
A chromatogram (5'-3' direction) indicating the double peak in DNA of an individual heterozygous for the -853C>T variant. Peaks specify nucleotides as: red, thymine (T); blue, cytosine (C); black, guanine (G); green, adenine (A). Arrowpoint indicates point of variation.

Figure 4.3 (A) Genotyping of the *ALAS1* -853C>T variant using a 342 bp amplicon.



AclI digest of the 342 bp fragment, assessed on a 15% polyacrylamide gel stained with EtBr.
Lane 1: 100 bp marker (Fermentas);
Lanes 2-6: 321 bp bands of individuals homozygous for -853C/C;
Lane 7: negative control;
Lane 8: 342 bp and 321 bp bands of an individual heterozygous for -853C/T.
Given its small size and increased migration rate, the 21 bp band was not visible on the gel.

Figure 4.3 (B) Genotyping of the *ALAS1* –853C>T variant using a 114 bp amplicon.



AclI digest of the 114 bp fragment, assessed on a 15% polyacrylamide gel stained with EtBr.
Lane 1: 100 bp marker (Fermentas);
Lanes 2 & 4: 93 bp bands of individuals homozygous for -853C/C;
Lane 3: 114 bp band of an individual homozygous for -853T/T;
Lane 5: 114 bp and 93 bp bands of an individual heterozygous for -853C/T;
Lane 6: negative control.
The 21 bp band was not visible on the gel.

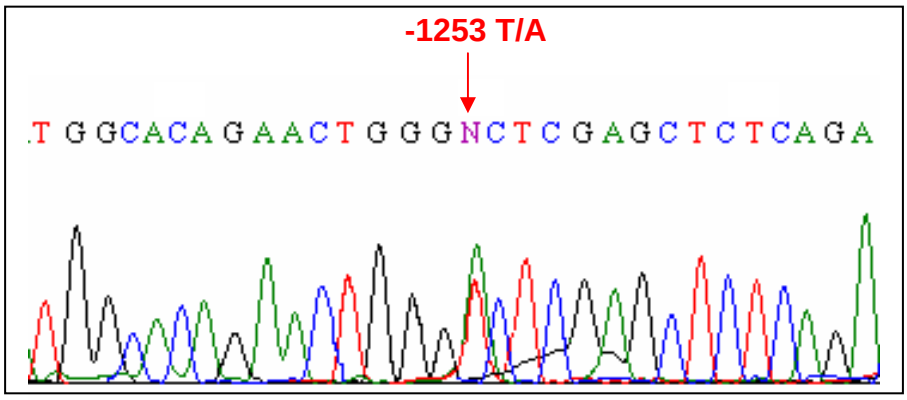
Table 4.1 Genotype- and allele frequencies of the *ALAS1* -853C>T variant in the control population.

Group	Genotype		Number	Total	Genotype Frequencies	Allele Frequencies
Control population (Mixed Ancestry)	-853	CC	84	90	0.933	C = 0.96
		CT	4		0.044	T = 0.04
		TT	2		0.022	

ii) -1253 T>A Genotyping

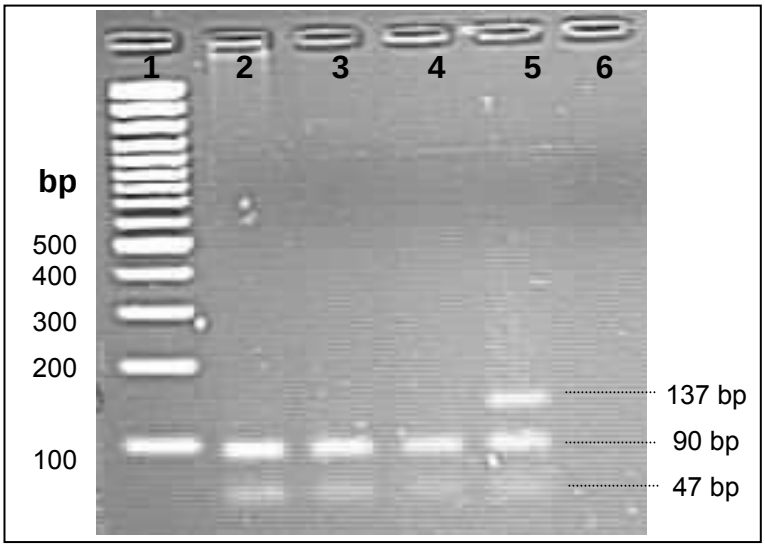
Primers were designed to amplify a 137 bp region of the *ALAS1* 5' region to encompass the -1253T>A transversion (Table 3.5; Figure 4.4). The PCR fragments were evaluated by comparison to a 100 bp molecular weight marker on a 1% (w/v) agarose gel. The amplicons were digested overnight with *BsaI* and genotyped on 2% (w/v) agarose gels. Individuals homozygous for the wt -1253T/T allele yielded two digested fragments of 90 bp and 47 bp, whereas individuals homozygous for the -1253A/A mutant allele produced an undigested fragment of 137 bp and heterozygous individuals (-1253T/A) yielded all three fragments (Fig 4.5). The VP patient group, the population matched control group (frequencies indicated in Tables 4.2) and the family members of affected individuals were genotyped for this variant.

Figure 4.4 DNA sequence chromatogram of a patient heterozygous for the *ALAS1* -1253T>A variant.



A chromatogram (5'-3' direction) indicating the double peak in the DNA of a heterozygous individual at -1253T>A. Peaks specify nucleotides as: red, thymine (T); blue, cytosine (C); black, guanine (G); green, adenine (A). Arrowhead indicates point of variation.

Figure 4.5 Genotyping of the -1253T>A variant.



A *BsaI* digest of the 137 bp fragment, assessed on a 2% agarose gel stained with EtBr, indicating two alleles of the -1253T>A variant.
Lane 1: 100 bp marker (Fermentas);
Lanes 2 – 4: 94 bp and 47 bp bands of individuals homozygous for -1253T/T;
Lane 5: 137 bp, 90 bp and 47 bp bands of an individual heterozygous for -1253T/A.
Lane 6: Negative control.

Table 4.2 Genotype frequencies of the *ALAS1* -1253T>A variant in the control population.

Group	Genotype		Number	Total	Genotype Frequencies	Allele Frequencies
Control population (Mixed Ancestry)	-1253	TT	83	90	0.922	T = 0.96
		TA	6		0.067	A = 0.04
		AA	1		0.011	

4.1.3 Statistical analysis

In the VP patient group, the -853C/T and -1253T/A variants were both identified in the same three female individuals of Mixed Ancestry, each classified in a different symptomatic class, i.e. asymptomatic, acute attacks or skin symptoms. The likelihood of the involvement of the detected variants in VP symptomatic expression was thus unlikely.

In this study, accurate statistical association analysis using DNA variants and clinical symptomatic classes were not possible, since the power of this study was compromised by the small VP patient sample size. Consequently, only variants that have definitive effects on the development of VP symptoms might have been identified, and thus, minor effects mediated by variants in this candidate gene, could not be excluded. Another confounding factor in identifying variants that produce minor phenotypic modifying effects in VP is the heterogeneous nature of this disorder, implicating that different modifier genes or different allelic variants of these genes, may play a part in mediating the symptomatic status in different subsets of VP patients. Also, inadequate information regarding exposure to precipitating factors was available for the patient group. However, an interesting observation from the accessible patient data, was that the

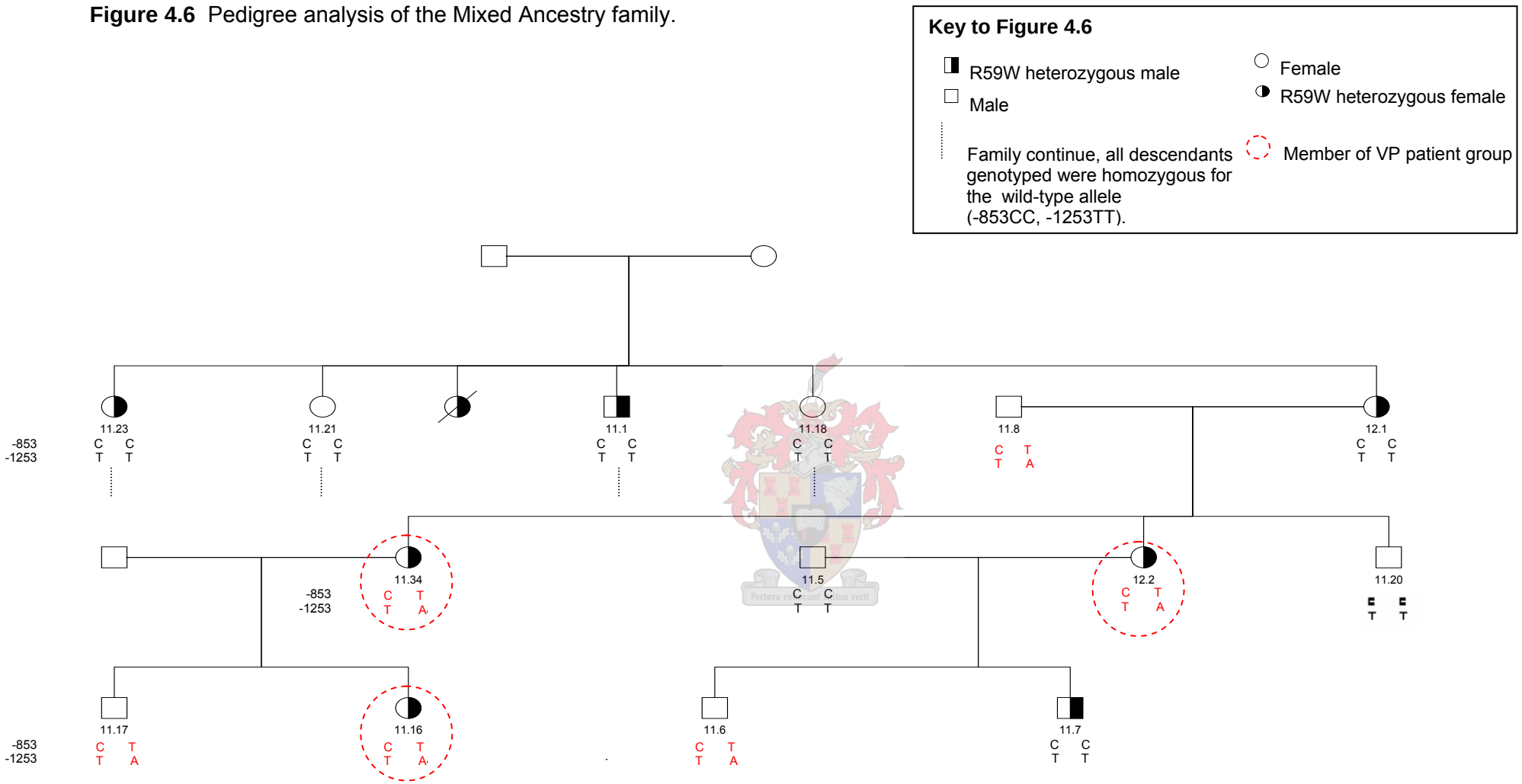
asymptomatic patient heterozygous for -853C/T and -1253T/A, had not yet reached puberty, with no previous history of hormonal contraceptive use. The second patient heterozygous for the two variants, classified as having only skin symptoms, was at post-menopausal age and also had no previous history of contraceptive use. However, the third heterozygous patient that presented with acute attacks had a record of contraceptive use. These observations raised questions about the possible involvement of hormonal effects on their symptomatic expression.

Since the variants were detected only in VP patients of the same Mixed Ancestry family, 31 available family members were screened for the presence of the identified variants. The variants were exclusively found together in the additional family members genotyped. These mutant alleles were found in three additional members (9.7%) of the Mixed Ancestry family. Pedigree analyses for the family using four generations revealed that, in this family, the two variant alleles at -853 and -1253 are co-inherited (Figure 4.6).

To establish whether these variants are also present in the Mixed Ancestry population, genotype analysis of a population-matched control group, comprised of 90 individuals, indicated that seven individuals had either one or both of the two identified variants. Of these seven individuals, three (42%) were heterozygous for both variants. Furthermore, the detection of two control individuals heterozygous for the -1253 variant but homozygous for the -853 mutant variant, one control individual homozygous for the -1253 mutant variant but heterozygous for the -853 variant, and one control individual heterozygous for the -1253 variant but homozygous for the -853 wt variant, suggests that no strong linkage disequilibrium exists. The Mixed Ancestry population is still a relatively young population that diverged from Caucasians, Africans and Asians (Nurse *et al.* 1985). Given this fact, the probability that two variants are found exclusively in this population, is highly unlikely, and merits the screening of other population groups to assess the frequency of the variants. According to the screening results obtained from the control group, the genotype frequencies in the Mixed Ancestry population are CC:CT:TT= 0.933:0.044:0.022 for the -853 variant, and TT:TA:AA= 0.922:0.067:0.011 for the -1253 variant. Although the genotype distribution for the -1253T/A variant in the control group was found to be distributed in the confines of the Hardy-Weinberg equilibrium (p=0.12), the genotype distribution for the variant at -853C/T was not (p=0.0). A possible reason for this deviation may be attributed to the small sample size and the fact that some members of the control population were residents of a relatively isolated area. Consequently, non-random mating or selection may have occurred.

As no evidence regarding the effect of these two variants could be obtained from the VP patients, bioinformatic- and functional analyses were performed to establish whether these variants could influence *ALAS1* transcriptional expression.

Figure 4.6 Pedigree analysis of the Mixed Ancestry family.



Three R59W individuals (12.2, 11.16 and 11.34), part of the VP patient group, and their available family members were genotyped for the -853 C/> and -1253 T>A variants. The family tree indicates that the two variants are co-inherited, with the mutant allele indicated in red.

4.1.4 Bioinformatic analysis of *ALAS1* promoter

In order to examine the presence of regulatory elements within the -1286 bp 5' flanking region of *ALAS1*, the area was screened for the presence of putative TFBSs by means of a bioinformatic approach. A multitude of putative regulatory motifs were predicted by the programs used (described in section 3.4), but subsequent experimental analysis of each predicted site was evidently unrealistic. The observed differences in TFBS prediction (Tables 4.3 to 4.5) between the various programs utilized, are probably an indication of the variation in sensitivity among the different programs. Some software programs only specify TFBSs that were experimentally proven, while other programs indicate all possible sites for a particular sequence region.

Analysis of the core promoter region revealed the consensus sequences of the TATA- and Inr elements and confirmed results from Roberts and Elder (2001), which indicated putative binding site sequences with homology to Sp1, nuclear factor κB (NFκB) and nuclear respiratory factor 1 (NRF1) TFBSs (Figure 2.3). Current information on TFBS sequences indicates that genetic variants may be actively involved in influencing gene expression in numerous ways, such as affecting TF binding or causing conformational changes in DNA molecules which may have an effect on transcription. Consequently, attention was primarily focused on the regions involving the novel variants detected in order to determine the possibility of these nucleotide substitutions affecting promoter activity.

The majority of programs utilized in this study predicted the presence of an optimal consensus half-palindromic (5'-TGACCT-3') estrogen response element (ERE), the binding site for the estrogen receptor, stretching from position -850 to -845 (Figure 4.7). This ERE half-site is located three bp 3' to the -853C/T bp variant. The program outputs of Matinspector, Alibaba2, TRES, Match and Signal Scan software were compared to identify similarities in ERE motif prediction (Table 4.3). The predicted gain or loss of a putative TFBS due to a given nucleotide change was ascertained using the default settings and parameters of the various programs used. As seen in Table 4.3, some programs specified that the -853T variant eliminate this ERE half-site (and *vice versa*), despite the fact that this variant is not located in the half-site itself.

Another predicted TFBS overlapping this particular site warrants mentioning: the chicken ovalbumin upstream promoter transcription factor (COUP-TF) is expressed in a wide variety of human tissues such as the liver, nervous system, uterus and mammary gland, where they regulate vital biological functions (Pereira *et al*. 1995; Pereira *et al.* 2000; Tripodi *et al*. 2004). COUP-TFs bind to a consensus imperfect direct repeat (3'-GTGTCAnAGGTCA-5') (Wang *et al*. 1989). One inverted half of this repeat is localized at this site (-850 to -845 bp) and is predicted to be a putative COUP-TF binding site by two of the software programs used. Based on their sequence, COUP-TFs can be seen as members of estrogen receptor TF subfamily, but COUP-

TFs are classified as orphan NRs for which no ligands have been identified (Tran *et al*. 1992). COUP-TFs are known to bind to consensus EREs or ERE half-sites in a number of gene promoters where it may either promote or antagonize ERα binding (Liu *et al*. 1992, Liu *et al*. 1993, Burbach *et al.* 1994; Klinge *et al*. 1997; Klinge 1999; Métivier *et al*. 2002).

These predictions directed the focus of the bioinformatic analyses to the involvement of putative ERE site/s in the regulation of *ALAS1*. After further bioinformatic analysis it was observed that some of these programs predicted two additional imperfect ERE half-sites (at -1169 to -1158 bp and at -820 to -814 bp) and one consensus ERE half-site (at -1101 to 1096 bp) in the -1292 bp of the *ALAS1* promoter examined (Table 4.4, Figure 4.7). Subsequent visual analysis of the published sequence confirmed that the imperfect ERE half-site predicted in the region between -1169 to -1158 bp, is in fact two putative overlapping imperfect ERE half-sites, while the region between -820 to -814 bp contains the recognition sequence of an imperfect ERE half-site element. Each of the predicted imperfect half-sites differs from the consensus half-site by one nucleotide. Visual examination also confirmed that the region between -1101 to -1096 bp contains one consensus ERE half-site (Figure 4.7).

Bioinformatic analysis of the region harbouring the -1253 T>A variant predicted only one regulatory site in this vicinity, a potential Sp1 site, located three bp 5' (-1262 bp to -1252 bp) of the detected variant. This TFBS was predicted by only one of the programs utilized (Table 4.5). Although Sp1 binding site sequences are reported to be highly variable (Huang and Chuang 1998; Skak and Michelsen 1999), this sequence has not been described as a recognition sequence for Sp1 in humans.

Table 4.3 Results of the software programs used that predicted an ERE half-site directly 3' to the -853C>T variant in the *ALAS1* promoter.

Region: –850 to –845 bp		
Program	**-853C *ALAS1* (Wild-type)**	**-853T *ALAS1* (Mutant)**
Matinspector	No ERE predicted	Predict ERE half-site
Alibaba	Predict ERE half-site	Predict ERE half-site
Signal Scan	Predict ERE half-site	Predict ERE half-site
Match	No ERE predicted	No ERE predicted
TRES	Predict ERE half-site	No ERE predicted

Table 4.4 A summary of additional ERE half-sites predicted in the *ALAS1* promoter fragment.

Regions:			
Program	**-1169 to -1158 bp** (Two overlapping imperfect half-sites)	**-1101 to -1096 bp** (Perfect half-site)	**-820 to -814 bp** (Imperfect ERE)
Matinspector	No ERE predicted	No ERE predicted	No ERE predicted
Alibaba	Predict imperfect ERE half-site	Predict perfect ERE half-site	No ERE predicted
Signal Scan	No ERE predicted	No ERE predicted	Predict imperfect ERE half-site
Match	No ERE predicted	No ERE predicted	No ERE predicted
TRES	No ERE predicted	Predict perfect ERE half-site	No ERE predicted

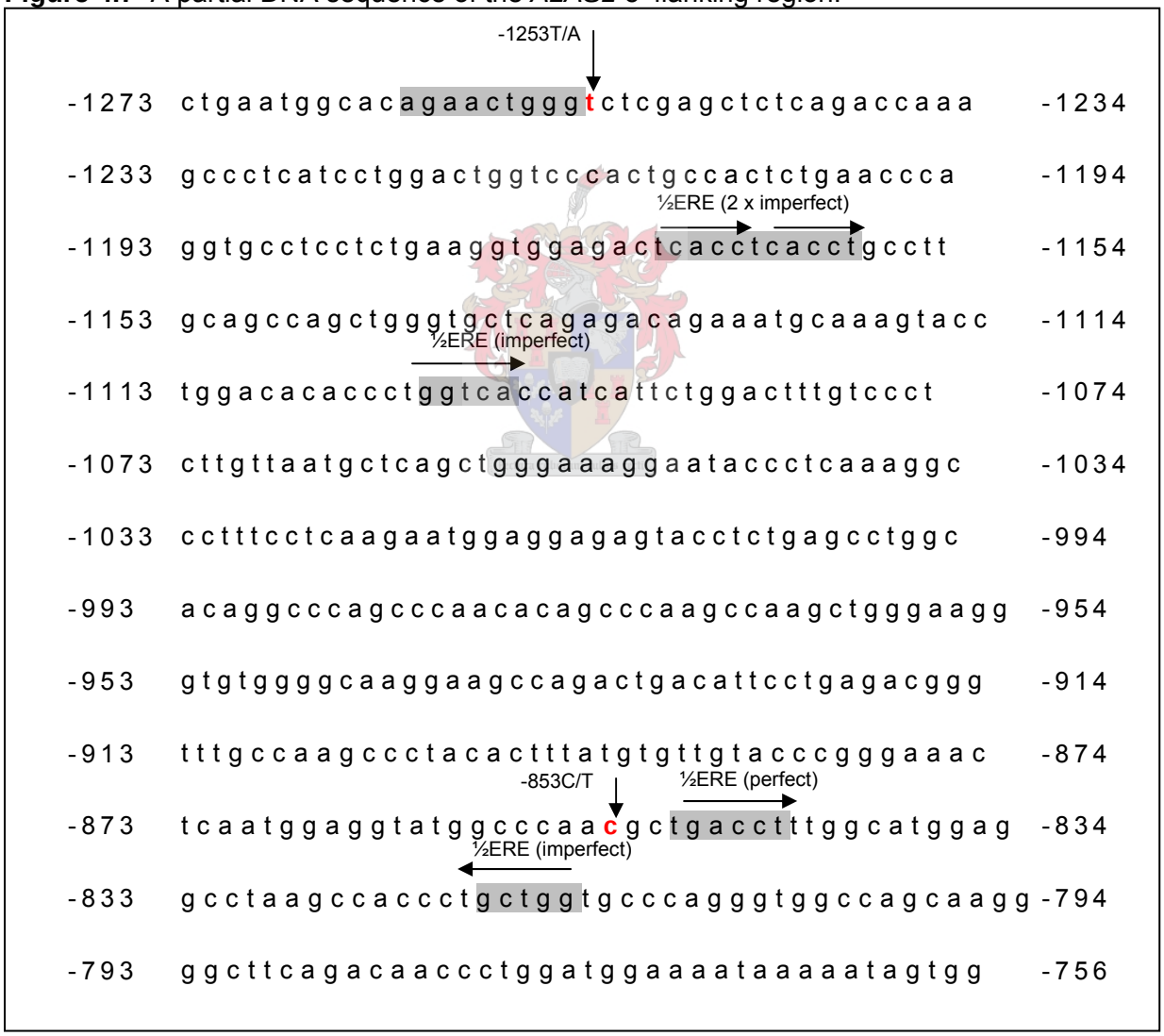
Table 4.5 A summary of the TFBSs predicted in the vicinity of the *ALAS1* -1253T>A variant.

-1253 T>A		
Program	**-1253T-ALAS (Wild-type)**	**-1253A-ALAS (Mutant)**
Matinspector	No site predicted in this area	No site predicted in this area
Alibaba	Predict Sp1 at –1262 bp	No Sp1 predicted at –1262 bp
Signal Scan	No site predicted in this area	No site predicted in this area
Match	No site predicted in this area	No site predicted in this area
TRES	No site predicted in this area	No site predicted in this area

Phylogenetic footprinting analysis was performed of the sequence between the TSS and the most 5' ADRES element, to cover a 21 kb upstream region of *ALAS1*. The Multi-Lagan program was implemented to align the 21 kb upstream region of the *ALAS1* gene in eight species, with the TSS of *ALAS1* of each species as anchor point, with the objective to identify evolutionally conserved regions with possible regulatory importance (as described in section 2.4.4.1). Special attention was given to the degree of conservation between species in the regions harbouring the variants identified (at -853 bp and -1253 bp) and the regions harbouring the putative ERE half-sites in the human promoter region. The alignment results compared to the human sequence are shown in Figure 4.8. As anticipated, the results revealed that the human and chimpanzee 21 kb upstream regions are almost completely homologous. In the area of the -853C>T variant, the mouse sequence indicated a homologous peak from -800 to -900 bp that show approximately 70% homology to the human sequence. In both the cow and the dog, this area displays a 75-80% homology to the human sequence. In the area of the -1253T>A variant, the chimpanzee, mouse, cow and dog sequences each share a degree of similarity. The chimpanzee and human sequences are almost completely homologous in this area, while the

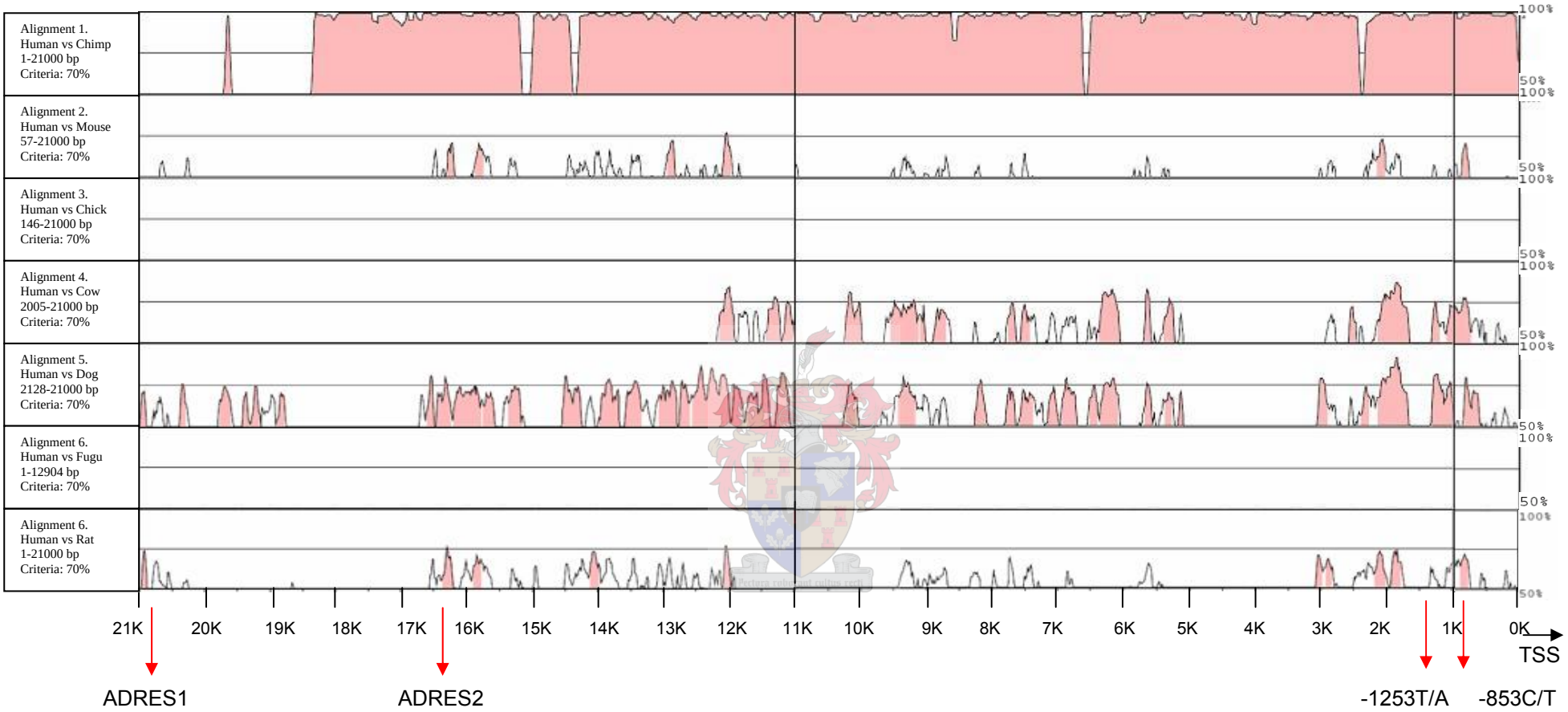
dog and cow sequences indicate an approximate 75% homology to the human sequence. When the positions of the two ADRES elements in the human *ALAS1* distal promoter, at -16 and -21 kb, were compared to the corresponding positions in the other species, the dog and rat appear to demonstrate approximately ~65% of conservation in these regions. However, the mouse and chicken *ALAS1* ADRES elements are found at -14 and -17.5 kb respectively (Fraser *et al*. 2002; 2003). Taken as a whole, in these species the -21 kb *ALAS1* sequence that are the most homologous to the human *ALAS1* 5' region is, in declining order, the chimpanzee, dog, cow, rat and mouse. Collectively, these bioinformatic observations indicated a distinct level of conservation of the analysed regions in some species, and prompted the functional analysis of the -853C/T and -1253T/A variants in the *ALAS1* promoter to determine whether these variants have a significant effect on transcriptional regulation of this gene.

Figure 4.7 A partial DNA sequence of the *ALAS1* 5' flanking region.



A section of the reference nucleotide sequence of the *ALAS1* 5' flanking region stretching from position -756 to -1273. The sequence was retrieved from the NCBI Entrez Nucleotide Database (accession number AC006252), with the first base of the most 3' TSS denoted as +1, and the nucleotide directly 5' to this set as -1 (according to Roberts and Elder 2001). A complete optimal ERE site has a consensus sequence of 5'-AGGTCAnnnTGACCT-3'. The optimal perfect and imperfect ERE half-sites are highlighted and overlined with their directions indicated with an arrowhead; the predicted Sp1-like site at -1262 to -1252 bp is highlighted. The wild-type nucleotides of the detected variants (-853C/T and –1253T/A) are identified in red.

Figure 4.8 A Multi-LAGAN multiple alignment comparison of the 21 kb region upstream of the *ALAS1* TSS.

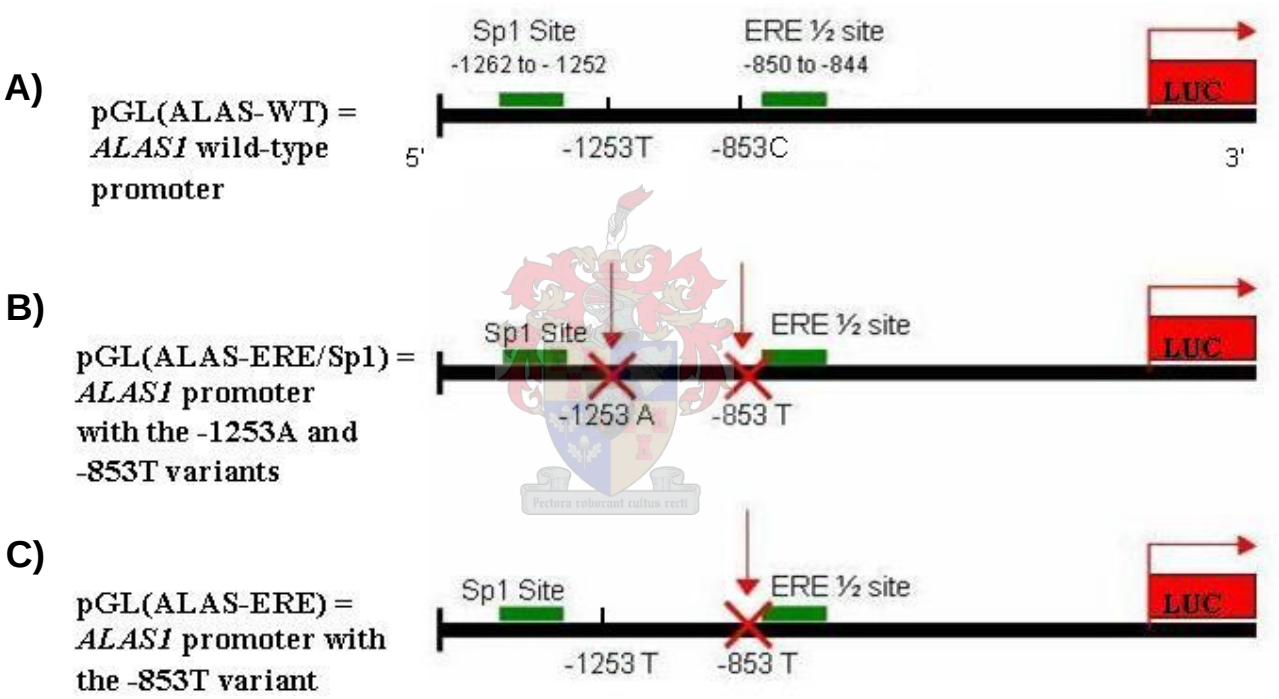


The Multi-LAGAN software program output of the alignments between the human *ALAS1* 21 kb upstream region with the corresponding *ALAS1* regions of the chimpanzee, mouse, chicken, cow, dog, pufferfish and rat. The location of the ADRES elements and the identified variants at -1253 and -853 are indicated by the arrows.

4.2 PREPARATION OF *ALAS1* PROMOTER CONSTRUCTS

Promoter constructs harbouring the -1286 bp to +6 bp region of the *ALAS1* promoter region were created. The DNA of two patients were used to generate two promoter constructs, one with the wt promoter [pGL2(ALAS-WT)] and one harbouring both detected variants which appear to be co-inherited in the VP family [pGL2(ALAS-ERE/SP1)]. The pGL2(ALAS-ERE/Sp1) construct was used to create the pGL2(ALAS-ERE) construct, harbouring only the -853T variant located adjacent to the putative ERE half-site. The genotypes of all constructs were verified by direct sequening. Fig 4.9 illustrates the three *ALAS1* promoter reporter constructs generated.

Figure 4.9 The pGL2-*ALAS1* promoter constructs.



A diagrammatic representation of the three different pGL2-*ALAS1* promoter constructs (not drawn to scale):
(A) pGL(ALAS-WT), contains the wt *ALAS1* promoter fragment;
(B) pGL2(ALAS-ERE/Sp1), contains the *ALAS1* promoter fragment with the -853T and -1253A variants;
(C) pGL2(ALAS-ERE), contains the *ALAS1* promoter with the -853T variant.
The putative Sp1- and ERE half-sites are indicated with green boxes and the TSS of the *Luc* gene indicated with red boxes.

4.3 EXPRESSION OF *ALAS1* IN HUMAN HEPATOCYTES

To verify the *in silico* predictions, and to assess if the -853C/T and -1253T/A variants influence the *in vitro* transcription rate of *ALAS1* in human liver hepatocytes, the *ALAS1* promoter fragments, fused to the promoter-less firefly luciferase gene of the pGL2-Basic plasmid, were used in transient transfection studies. Transfections were conducted in a HepG2 cell line at a passage stage accepted to express negligible amounts of ERα (as described in section 3.6.1). Transfection studies were designed to consist of four sets of initial analyses (set 1-3) and optimizations (set 4), followed by the final quantification in set 5 (sets indicated in Tables 3.8 to 3.12). In each set, the promoter-less pGL2-Basic control vector was used as the negative control construct and a β-galactosidase expression vector (pSVβ-Gal) was included in all transfections as an internal control to monitor transfection efficiency. The pGL-promoter plasmid with an incorporated consensus ERE (pGLvitERE; 5' GGTCAnnnTGACC 3'), found in the *Xenopus* vitellogenin A_2 gene promoter, was used as positive control in some optimization experiments to investigate estrogen-ERα mediated up-regulation *via* this consensus ERE. The normalized results obtained for each construct type were determined to establish the relative transcription rate for each construct over three experiments assayed in triplicate (as described in sections 3.6.3 & 3.6.4). The normalised values obtained in each set are indicated in Appendix 6.

4.3.1 Set 1: Expression analysis under normal culture conditions

The first set of transfections (Table 3.8) was performed to test the basal expression of the three pGL2-*ALAS1* promoter reporter constructs under "normal" (standard) culture conditions (phenol-red DMEM, 37°C; 5% CO_2), i.e. in the presence of trace elements of hormones such as estrogen (present in serum) and the estrogenic effect of phenol-red in culture medium. The constructs were each transfected, together with pSVβ-Gal, into the HepG2 cell line. For this set, transfections of each construct were performed in triplicate, and the set repeated independently on three occasions. The normalized values indicated consistent, repeatable expression during all repetitive transfections performed.

The normalized results obtained for each construct type were determined and the relative transcription rate for each construct was established. The expression value of the pGL(ALAS-WT) construct was used as reference and assigned a value of 1 and the expression of the other constructs expressed as a value relative to that of pGL(ALAS-WT). Comparison of the values obtained for pGL(ALAS-WT), pGL(ALAS-ERE) and pGL(ALAS-ERE/Sp1) revealed that the transcriptional expression between the three constructs varied under normal culture conditions (Figure 4.10). Expression of pGL(ALAS-ERE/Sp1) was 39% higher than that of pGL(ALAS-WT),

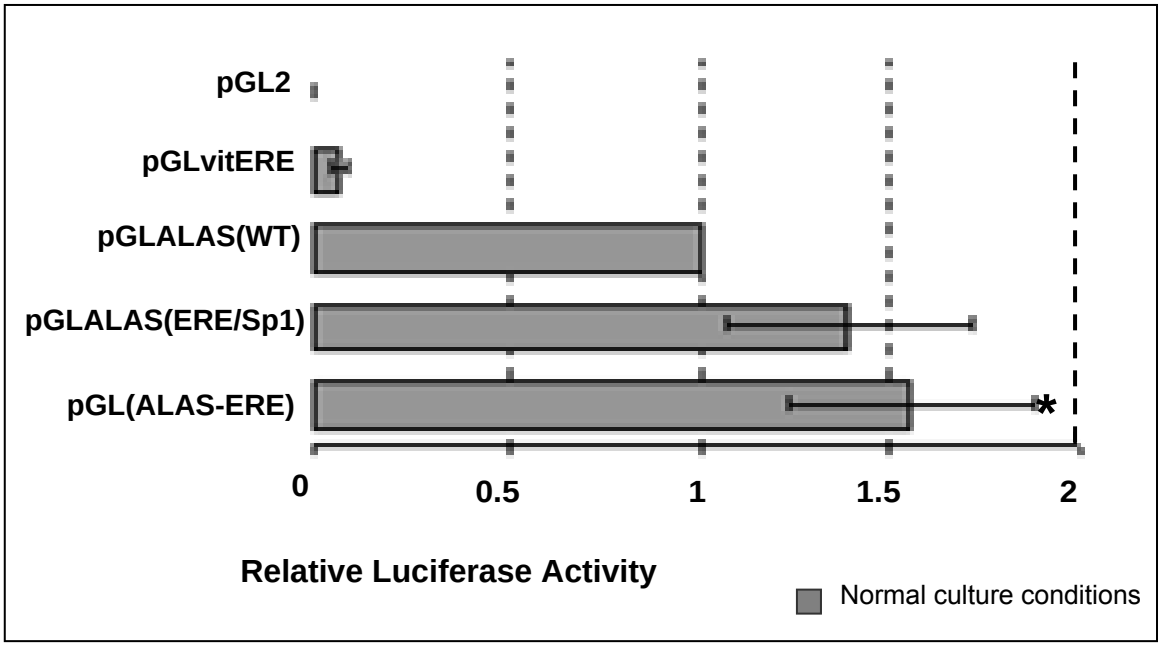
with a *p* value of 0.139. The pGL(ALAS-ERE) construct exhibited an 55% increase in transcription rate when compared to pGL(ALAS-WT), a statistical significant increase in expression with a *p* value of 0.046. The expression of pGL(ALAS-ERE) was 16% higher than pGL(ALAS-ERE/Sp1), and although this difference in expression was determined not to be statistically significant, a definite increase in expression of pGL(ALAS-ERE) was observed in all individual experiments. The pGLvitERE construct was included to monitor its expression under normal culture conditions, and normalized values indicated very low expression, not visible on the histogram (Figure 4.10).

Interpretation of these results illustrated that, under normal culture conditions, the -853T variant instigates a statistically significant increase in transcriptional response when compared to the wild-type promoter. The -853T variant in combination with the -1253A variant also caused a statistically significant higher expression rate than the wild-type promoter, albeit lower than that of the -853T on its own. Although the decreased transcriptional response of pGL(ALAS-ERE/Sp1) compared to pGL(ALAS-ERE) was not statistically significant, such a decrease may have substantial effects in an *in vivo* environment (Knight 2005).

Results indicate an increase in basal transcriptional activity (wild-type promoter expression) mediated by the -853T variant in hepatocytes under normal culture conditions, and an inhibiting effect of the -1253A variant. Since the -853C/T variant is closely situated to a putative ERE site, the difference in expression seen in the presence of this variant was contemplated to be linked to the possible functionality of this putative ERE site. It was hypothesized that these HepG2 cells may still express small amounts of ERα, which may, in combination with the estrogen present in the normal culture medium and serum, act positively on this possibly functional ERE half-site/s, with an added increase in the presence of the -853T variant. The ERα and estrogen levels, present in normal culture conditions, may however have been too low to have an effect on the positive control construct (pGLvitERE), or possibly, the high expression values obtained from the *ALAS1* promoter constructs compared to pGLvitERE, may have masked a slight up-regulation of pGLvitERE (Figure 4.10).

The inhibitory effect observed in the presence of the -1253A variant may be due to a myriad of mechanisms, possibly affecting the binding of Sp1 or another enhancer protein or by affecting DNA conformation or mediating the binding of a repressor etc.

Figure 4.10 Expression of *ALAS1* promoter constructs under normal culture conditions in HepG2 cells.



A histogram indicating the normalised mean ± SEM expression values of the three pGL2-*ALAS1* promoter reporter constructs in normal culture conditions (fully supplemented phenol-red DMEM, 37 °C; 5% CO_2) from three independent transfection experiments (assayed in triplicate). The promoter-less pGL2-Basic vector was included as negative control and the pGLvitERE was included as an estrogen responsive positive control vector. The expression value of pGL(ALAS-WT) was used as the reference and assigned (normalised to) a value of 1. Expression of the other constructs was conveyed as a value relative to the reference with asterisks' denoting significant differences from the reference (* $p < 0.05$). The error bars indicate the standard deviation (S.D).

4.3.2 Set 2: Expression analysis under hormone-free culture conditions

The results obtained in set 1, in particular the elevated expression observed under normal culture conditions of constructs harbouring the -853T variant, prompted further investigation into the possible role of estrogen in the detected expression patterns. Set 2 (Table 3.9) was performed to test the expression of the three *ALAS1* promoter reporter constructs under hormone-deprived conditions (antibiotic-free phenol-red-free DMEM; charcoal-treated FCS; 37°C; 5% CO_2) to test the hypothesis deduced from set 1, i.e. the possible functionality of the predicted ERE half-sites and that the variation in expression is due to the enhanced effect of trace amounts of endogenous ERα, activated by serum-estrogen, to the ERE/s in the presence of the -853T variant. Transfections of each construct were performed in triplicate, and the set repeated independently at three occasions. The normalized values for each construct were expressed relative to that of pGL(ALAS-WT), which was assigned a value of 1 (Figure 4.11).

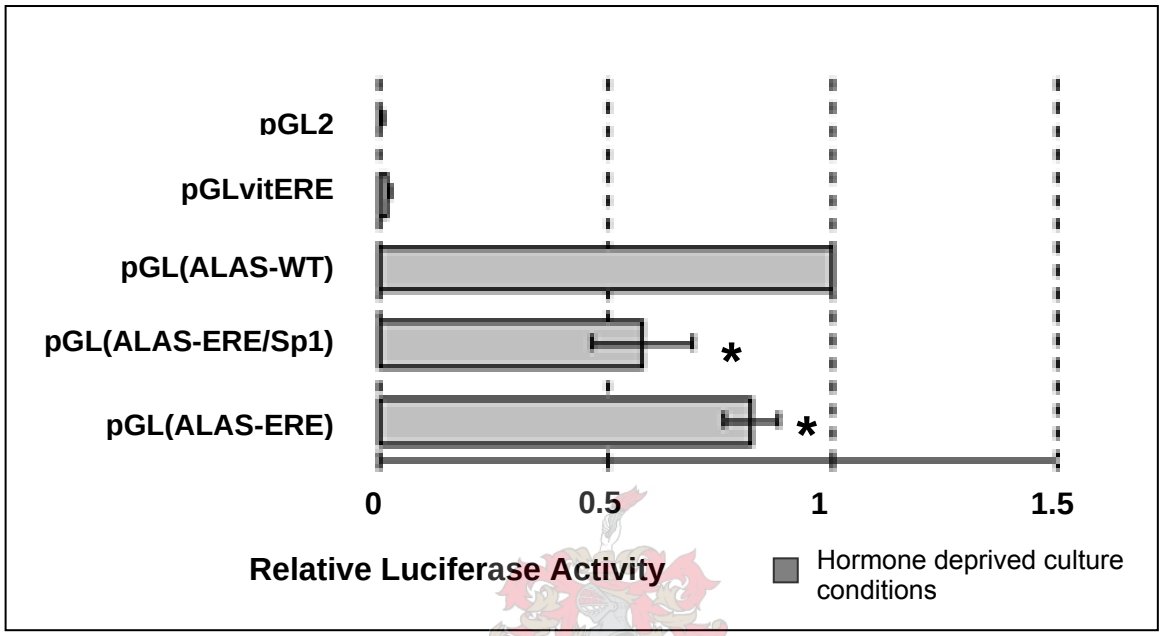
Results indicated that, in the absence of estrogen, pGL(ALAS-ERE) exhibited an 18% decrease in expression when compared to pGL(ALAS-WT), while the expression of pGL(ALAS-ERE/Sp1) was 42% lower than that of pGL(ALAS-WT). Statistical analysis indicated that the expression of pGL(ALAS-ERE) and pGL(ALAS-ERE/Sp1) were significantly lower than that of pGL(ALAS-WT) with *p* values of 0.05 and 0.0001 respectively. Expression of pGL(ALAS-ERE) compared to pGL(ALAS-ERE/Sp1) specified a statistical significant ($p = 0.013$) higher expression rate of 24%. The expression of the pGLvitERE construct indicated, similar to its expression under normal culture conditions, low transcriptional activity that was not visible on the histogram (Figure 4.11).

These results suggest that the positive influence of the -853T variant, as seen under normal culture conditions, requires the presence of hormones such as estrogen. According to these findings, the probability of estrogen-ERE-mediated regulation was formulated as follows: Estrogen-activated ERα may act as a positive regulator of *ALAS1 via* this, and possibly the other predicted ERE half-sites, but that the -853T variant amplifies the effect of estrogen-activated ERα, causing an increased transcriptional rate as seen in set 1. In the absence of estrogen (set 2), this increase in basal expression mediated by -853T did not occur, probably due to the absence of E_2-activated ERα.

The added decrease in the transcription rate of the *ALAS1* promoter harbouring both variants [pGL(ALAS-ERE/Sp1)], compared to the *ALAS1* promoter with only the -853T variant [pGL(ALAS-ERE)], was attributed to the additional -1253A variant. Results imply that the effect of the -1253A variant is independent of the presence or absence of estrogen, as the decrease in

transcription was observed under these hormone-free conditions and also under normal culture conditions (set 1). However, the effect was more substantial under hormone-free conditions.

Figure 4.11 Expression of *ALAS1* promoter constructs under hormone deprived culture conditions in HepG2 cells.



A histogram indicating the normalised mean ± SEM expression values of the three pGL2-*ALAS1* promoter reporter constructs in hormone-free culture conditions (antibiotic-free phenol-red-free DMEM; charcoal-treated FCS; 37 °C; 5% CO_2) from three independent transfection experiments (assayed in triplicate). The promoter-less pGL2-Basic vector was included as negative control and the pGLvitERE was included as an estrogen responsive positive control vector. The expression value of pGL(ALAS-WT) was used as the reference and assigned (normalised to) a value of 1. Expression of the other constructs was conveyed as a value relative to the reference with asterisks' denoting statistically significant differences from the reference (* $p < 0.05$). The error bars indicate the standard deviation (S.D).

4.3.3 Set 3: Estrogen responsiveness of *ALAS1* wild-type promoter

Set 3 (Table 3.10) was performed to establish whether the differences in *ALAS1* expression, seen between normal- (set 1) and hormone-free (set 2) culture conditions, involve estrogen-mediated regulation. The *ALAS1* wild-type promoter was employed to test if estrogen-activated ERα act as a positive regulator of the *ALAS1* promoter, and to determine the need for further estrogen regulated expression studies.

The effect of estrogen-activated ERα on the wild-type promoter was evaluated by performing co-transfections under hormone-free conditions. The transcription rate of the pGL(ALAS-WT) construct, in the presence of E_2-activated ERα could thus be measured and compared to its expression in the presence of unliganded (unactivated) ERα. Consequently, pGL(ALAS-WT) was co-transfected with an estrogen responsive vector expressing human ERα. Estrogen inductions were performed with E_2 dissolved in EtOH at a dose classified as the physiological concentration (10^{-8} M). For the transfections representing expression in the absence of activated ERα, control inductions were executed with the solvent/vehicle (EtOH) at a concentration (0.03%) equal to the amount used to dissolve E_2. The pGLvitERE vector was included as a positive control and the pGL2-Basic (promoter-less) vector served as negative control. Although transfections were performed in triplicate, it was not repeated since this set was not considered as a quantification experiment. The β-galactosidase activities were similar for vehicle- and E_2-treated cells, suggesting that E_2 did not affect cellular growth/toxicity and that the transfection efficiency were constant across all experiments performed. The normalized expression value of the pGL(ALAS-WT) construct, without ERα and induced with E_2, was used as reference and assigned a value of 1. Expression of the other constructs was indicated as a value relative to this construct (Figure 4.12).

When the co-transfection results of the positive control construct, pGLVitERE, were analysed, it was evident that the experiment was effective, as this construct displayed a 2604% increase in expression in the presence of ERα and E_2 stimulation, compared to its expression in the absence of E_2. In the presence of E_2, the pGL(ALAS-WT) construct, co-transfected with ERα, displayed a 81% increase in transcription when compared to the reference construct conditions (no ERα, induced with E_2).

However, in the absence of estrogen, pGL(ALAS-WT) co-transfected with ERα, also displayed an increase (52%) when compared to the reference conditions. We hypothesized that this increased expression was attributable to a possible inadequate cell-rinsing procedure or trace amounts of estrogen still present in the charcoal-treated serum. The presence of residual estrogen during transfection, would have, in the presence of ERα, acted positively on the

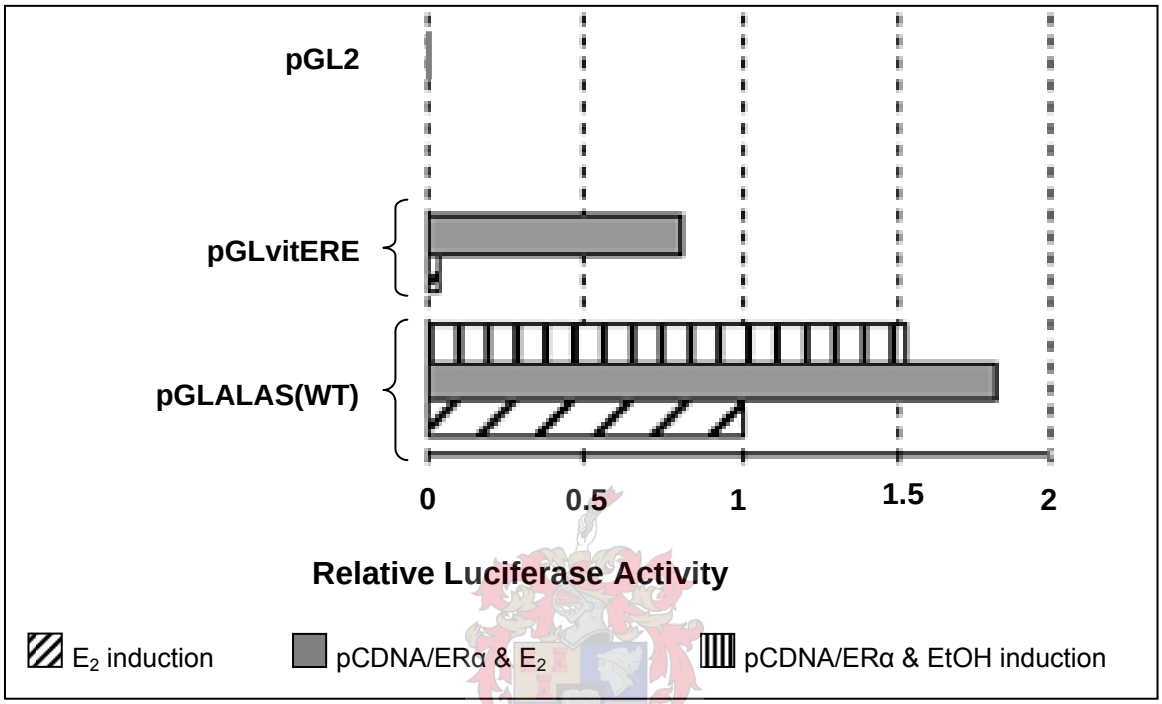
ALAS1 promoter, similar to the results observed for pGL(ALAS-WT) co-transfected with ERα in the presence of E_2. Alternatively, this may indicate that unliganded (unactivated) ERα bind to the *ALAS1* promoter, causing favourable conditions for transcriptional activity. Several quantitative experimental approaches have demonstrated ERα bound to promoters in the absence of hormone, with ERα binding increasing after estrogen addition (Kim *et al*. 2000). Numerous other studies have also demonstrated the binding of unliganded ERα to ERE-containing DNA *in vitro* (Brown and Sharp 1990; Zhuang *et al.* 1995), such as the use of chromatin immunoprecipitation data that demonstrated significant levels of ERα bound to the endogenous pS2- and Cathepsin D promoters in mammalian cells in the absence of hormone (Zheng *et al*. 2001). However, *in vivo* the situation has been less distinct (Klinge 2001). Cvoro *et al.* (2006) described the action of unliganded ERα as a co-activator in the transcription of the TNFα gene promoter.

Another explanation for this observation may be that *ALAS1* is transcriptionally up-regulated by EtOH. Various studies have reported on the direct transcriptional response of genes upon EtOH induction (Hassan *et al*. 2003; Uddin and Singh 2006a, 2006b). Wilke *et al*. (2000) reported the requirement of a consensus Sp1 site to mediate EtOH responsiveness. By insertion- and deletion analysis of the promoter region, they indicated the necessity of the proper context and spacing of the Sp1 site to confer EtOH transcriptional regulation. Their study expressed the capability of the Sp1 protein to interact with other co-factors that may also be responsible for the EtOH regulation. It might be possible that the *ALAS1* gene's increased expression during EtOH stimulation involve direct transcriptional responsive to EtOH, possibly *via* a Sp1 site. This hypothesis is further described in section 4.3.5 upon results obtained from set 5.

Results from set 3 confirmed that the difference in *ALAS1* expression seen between normal-(hormone) vs. hormone-free conditions, is (at least partly) due to the presence of estrogen. This gave credit to the hypothesis of the possible functionality of the putative ERE half-site/s (identified by *in silico* analyses) in the presence of estrogen-activated ERα. Nevertheless, it was recognized that this difference in expression seen upon estrogen induction, between the pGL(ALAS-WT) without ERα and pGL(ALAS-WT) with activated ERα, may also be a result of the incorporation of the additional expression vector with a strong viral promoter (pCDNA-ER), using transcriptional machinery that would otherwise have been available for luciferase gene expression of the promoter constructs. To eliminate this variable, and the increased expression seen during EtOH induction, it was decided to co-transfect all constructs in subsequent transfections with ERα to maintain accuracy regarding the influence of ERα and estrogen. Thereby, enabling the comparison of the expression rates among these constructs in the

presence of estrogen and evaluate it to the expression seen in the respective constructs, co-transfected with ERα, in the absence of estrogen (induced with EtOH).

Figure 4.12 Expression of the *ALAS1* WT-construct, co-transfected with ERα, in the presence and absence of estrogen in HepG2 cells.



A histogram indicating the normalised mean ± SEM expression values of pGL(ALAS1-WT) in hormone-free culture conditions, in the absence or presence of E_2-activated ERα. The promoter-less pGL2-Basic vector was included as negative control and the pGLvitERE was included as an estrogen responsive positive control vector. The expression value of pGL(ALAS-WT) without ERα, induced with E_2, was used as the reference and assigned a value of 1. Expression of the other constructs was conveyed as a value relative to the reference.

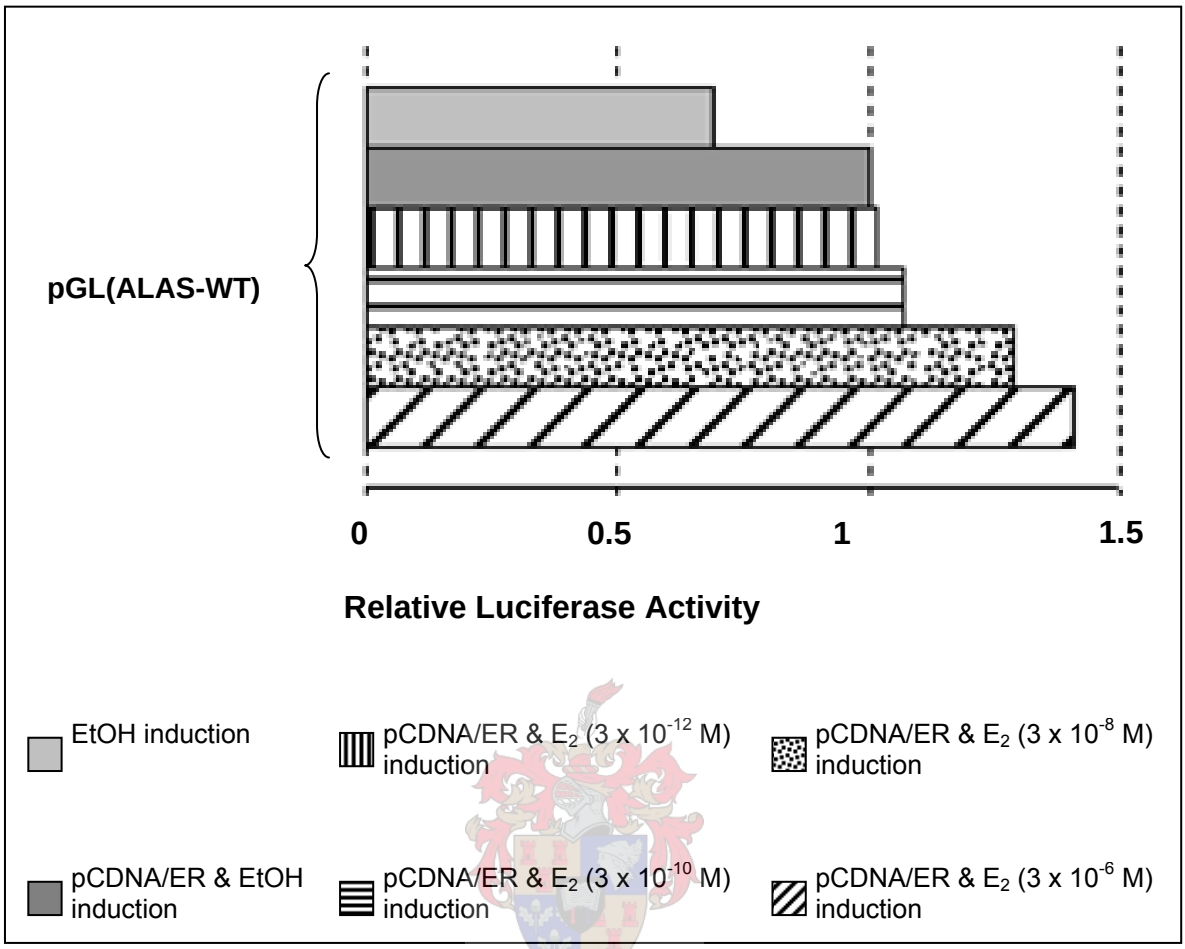
4.3.4 Set 4: Estrogen dose-responsiveness of *ALAS1* promoter

Prior to all estrogen inductions for the quantification set, a dose-response experiment (Table 3.11) was performed to test the activity of the *ALAS1* promoter on different E_2 concentrations, with the objective to obtain the optimal concentration of estrogen to be used in successive induction experiments.

Because of the difficulty to dissolve E_2, a working concentration of 3×10^{-3}M was prepared by using a sonicator to assure a homogenous solution. A dilution series was created from the working solution to ensure that the volume of EtOH-disolved-E_2, added during inductions, remains constant. The final E_2-concentrations of the treated cells were: 3×10^{-6}M, 3×10^{-8}M, 3×10^{-10}M or 3×10^{-12}M. The high concentration of the working solution prevented dose-response tests at E_2 concentrations higher than 3×10^{-6}M.

The normalised luciferase expression values indicated a linear increase in *ALAS1* promoter expression in response to augmented concentrations of estrogen (Figure 4.13). These results demonstrate that the *ALAS1* gene display a dose-dependent transcriptional up-regulation by E_2-activated ERα. Since estrogen at a concentration of 3×10^{-6}M induced the highest *ALAS1* promoter expression rate, this concentration was chosen as the most favourable to be used for subsequent induction experiments. Given that the liver is the major site of estrogen metabolism by cytochrome P450 enzymes, and the fact that haemoproteins (from the haeme synthesis pathway) are major constituents of the P450 enzymes, the high degree of “tolerance” by the HepG2 cells for E_2, and the increased expression of *ALAS1* to higher E_2 concentrations, were to some extent expected. Future dose-response experiments, using even higher levels of E_2 during inductions, will give an indication of the optimal level of estogen responsiveness of *ALAS1*.

Figure 4.13 Expression of the *ALAS1* WT-construct, co-transfected with ERα, induced with dilutions of estrogen, in HepG2 cells.



A histogram indicating the normalised mean ± SEM expression values of the pGL2-*ALAS1* wild-type promoter reporter construct in hormone-free culture conditions, co-transfected with ERα and induced with increasing concentrations of E_2. The expression value of pGL(ALAS-WT) co-transfected with ERα, induced with EtOH, was used as the reference and assigned (normalised to) a value of 1. Expression of pGL(ALAS-WT) in response to different E_2-concentrations, was conveyed as a value relative to the reference.

4.3.5 Set 5: Estrogen responsiveness of *ALAS1* promoter mutation-constructs

In order to substantiate the effect of co-transfected E_2-activated ERα, on the transcriptional expression of *ALAS1*, together with the influence of the two variants, human hepatocytes were grown under hormone deprived conditions and induced with either vehicle (0.03%) or E_2 at a concentration (3×10^{-6} M), as determined in set 4 (section 4.3.4). Co-transfections with each different construct were performed in triplicate (Table 3.12), and the set repeated on three independent occasions. The normalized expression value of the pGL(ALAS-WT) construct co-transfected with ERα and induced with EtOH, was used as reference and assigned a value of 1 (Figure 4.14)

Comparisons between the three constructs induced with EtOH, indicated that pGL(ALAS-WT) and pGL(ALAS-ERE) had approximately the same expression values, with the expression of pGL(ALAS-ERE/Sp1) slightly lower (11%) than these two. This reduced expression observed in pGL(ALAS-ERE/Sp1) correlates positively with the hypothesis deduced from set 3: the possible EtOH responsiveness of the *ALAS1* promoter *via* a Sp1 site. The reduced transcriptional expression observed in the context of the -1253A variant (located next to a putative Sp1-like TFBS), may therefore abolish the functionality of the putative Sp1 site (-1262 bp to 1252 bp), with a subsequent reduction in EtOH responsiveness, similar to the observations by Wilke *et al*. (2000). This reduction mediated by the -1253 variant is, however, also observed in the absence of EtOH (sets 1-2). Noteworthy, the observed EtOH responsiveness of *ALAS1* may also be related to the evidence that prolonged alcohol abuse induces a marked increase in ERα levels in livers of both male and female patients, especially in patients who had histological evidence of acute liver damage (alcoholic hepatitis) (Colantoni *et al*. 2002). Interestingly, as described in section 2.1.1.3, alcohol consumption is one of the major precipitating factors of VP symptomatic expression. Taken together, this observation may possibly provide credit to the hypothesis that ERα could interact with other regulatory proteins/regions, to cause an increase in *ALAS1* transcriptional response (Shang *et al*. 2000).

Comparisons between the different constructs in the presence of E_2, demonstrated a 23% difference in expression between pGL(ALAS-WT) and pGL(ALAS-ERE/Sp1). Although not statistically significant (p = 0.104), such a decrease may have significant influence *in vivo*. During E_2 stimulation, the expression of pGL(ALAS-ERE/Sp1) was 41% lower than that of pGL(ALAS-ERE), statistically significant with a p value of 0.0087. The increase in expression of pGL(ALAS-ERE) compared to pGL(ALAS-WT) was noticeable in E_2 induced conditions (17% increase), and even though this difference in expression was not statistically significant, in individual experiments the increase seen in pGL(ALAS-ERE) was much more prominent. (It is speculated that another repeat of the set will illustrate this significance in expression more

prominently). Nevertheless, a change in expression even of this modest degree, may have extensive consequences in an *in vivo* environment.

Transfections induced with E_2 resulted in a statistically significant increase in expression of all constructs, when compared to the respective constructs treated with the solvent, EtOH. The pGL(ALAS-WT) construct induced with E_2, demonstrated a 47% increase in expression compared to pGL(ALAS-WT) in the absence of E_2 (induced with EtOH), producing a *p* value of 0.0033. pGL(ALAS-ERE/Sp1) had a 35% increased expression rate in the presence of E_2 compared to its expression in EtOH ($p = 0.038$). The pGL(ALAS-ERE) construct displayed the highest expression: a 66% increase in the presence of E_2, compared to its expression in the absence of E_2 ($p = 0.0004$)

These results demonstrate that all three *ALAS1* promoter constructs are estrogen responsive. Initially it was speculated that the increase of all the *ALAS1* promoter constructs in response to estrogen, was due to the elevated demand for haeme to metabolize the high levels of estrogen. However, as described previously (refer to section 2.3.1.2), reports indicate that the transcriptional regulation of *ALAS1* transcription in response to haeme is highly unlikely (Srivastava *et al.* 1990; Hamilton *et al*. 1991) and that alternatively, haeme regulates *ALAS1* by decreasing its mRNA half-life (Hamilton *et al.* 1991; Munakata *et al*. 2004). Another mechanism of haeme-mediated regulation of *ALAS1* is *via* the haeme responsive regions, situated in the coding region of this gene, which is responsible for the intercellular translocation of the protein (Dailey *et al*. 2005). Since these haeme-mediated regulatory mechanisms all occur post-transcriptionally, the results observed in set 3, was solely accounted to the transcriptional effect of estrogen on the *ALAS1* promoter region. Consequently, it was hypothesized that the increase in luciferase reporter activity of the constructs may be consistent with the functionality of either one or more of the predicted ERE half-sites in the promoter region.

During estrogen induction, the -853T variant causes an evident increase in expression compared to the wt promoter with the C nucleotide at position -853. If the EREs are indeed functional, the elevated expression of the pGL(ALAS-ERE) construct may therefore demonstrate an enhanced influence of ERα on transcription, mediated by the -853T variant.

It has been long known that estrogen regulation can be mediated by imperfect half-sites and modulated by the sequences flanking ERE half-sites (Sanchez *et al*. 2002; Klinge 2001). A study by Anolik *et al.* (1993; 1995) showed that AT-rich sites adjacent to ERE sites increase the binding affinity of ERα-ERE binding and subsequent enhanced transcriptional activity. Surprisingly, they also reported that an ERE flanked by a GC-rich region, bound ERα in a

similar fashion to the AT-rich ERE, i.e. with enhanced affinity. They postulated that these sequences immediately flanking ERE sites, might share a common feature that stabilizes E_2-ERα DNA binding, conferring enhanced E_2-ERα-ERE binding and facilitate cooperative functionality of multiple ERE sites. In addition, they found common sequence features in the more distal flanking regions that allow binding of important co-factors. Driscoll *et al*. (1996) also described the enhanced binding affinity of E_2-ERα in AT-rich EREs. They postulated that the rationale behind this phenomenon is that, since AT-rich sequences are characterized by low melting temperatures and DNA bending, such a change in chromatin structure may bring DNA binding sites and their bound proteins in closer proximity to each other and the transcription initiation complex. In view of these findings, it is therefore possible that the -853T variant may influence the stability of ERα-ERE binding, resulting in an enhanced transcriptional effect. Since ERα-ERE function by creating a DNA bend, it can be hypothesized that the -853T variation, which has two A nucleotides at its 5' side, may create an area that is more easily bent, assisting in recruiting ERα to the ERE site/s and components of the initiation complex. It is therefore possible that the -853T variant may influence the stability ERα-ERE binding, resulting in an enhanced transcriptional effect.

In a study similar to this one, Harendza *et al*. (2003) identified a transition 1 bp from an ERE half-site in the Gelatinase A gene, which mediated a decrease in transcriptional activity in response to estrogen [as described in section 2.4.1.2(i)]. Although in the present study, the transcriptional effect of the transition (-853T) identified 3 bp from the ERE half-site, is opposite to the effect of the variant found by Harendza *et al.*, it reinforces the possibility of a closely situated variant influencing ERE action. It is important to bear in mind that although variations in ERE sites may elicit unpredictable transcriptional activity, ERα binding to ERE does not always result in a corresponding level of transcriptional activity and furthermore, the amount of transcriptional activity detected from the same ERE depends on cell-specific factors and surrounding promoter elements (Klinge 2001).

Evidence of half-sites acting synergistically has also been reported with some half-sites possessing greater responsive power due to their location with respect to the other half-sites. As described in the literature, these four predicted imperfect and perfect half-sites, might, if functional *in vivo*, function in a synergistic manner (Martinez and Wahli 1989; Kato *et al*. 1992; Cho *et al*. 1993). The number, location and spacing of these sites are known to be a determining factor in the significance of their functionality (Sathya *et al*. 1997). Although some of the predicted ERE half-sites in the *ALAS1* promoter are widely separated, they may be capable to function cooperatively, presumably because of their ability to interact by looping. It can be

speculated that the -853T variant may induce a DNA bend, placing the half-sites in a more favourable position with respect to each other to promote synergistic action.

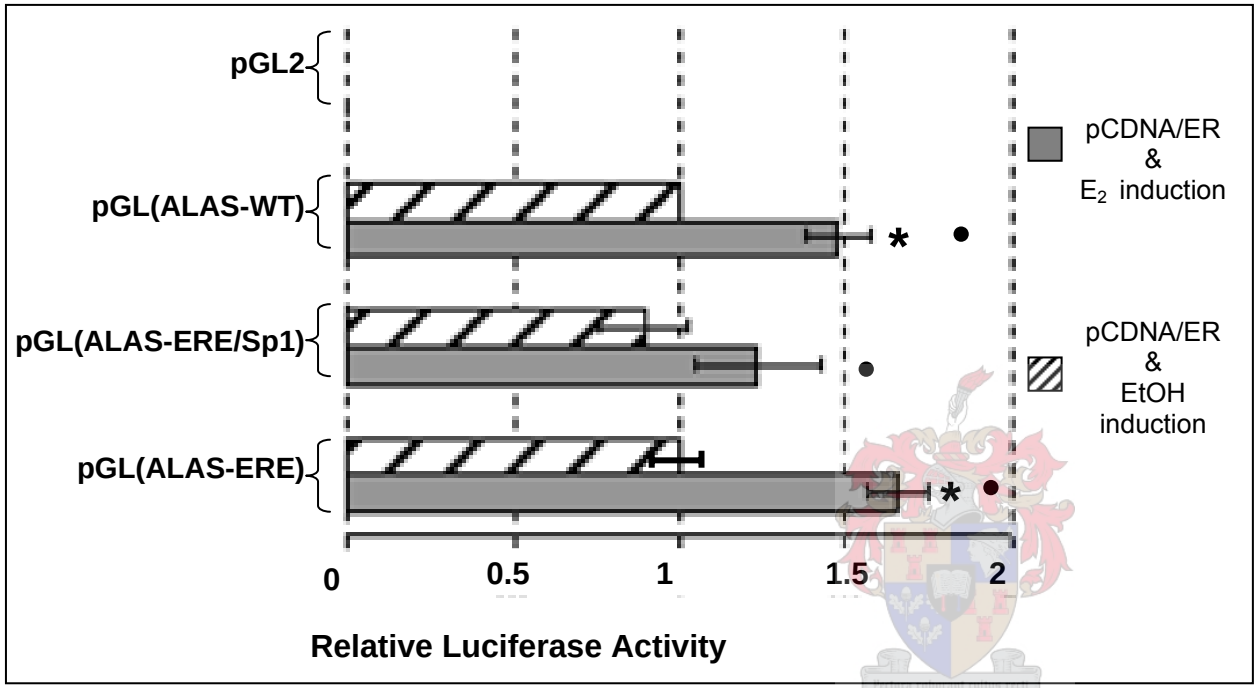
Many authors have observed that ERα works with multiple partners, including COUP-TF, to determine the transcriptional response of estrogen target genes to estrogens (Burbach *et al*. 1994; Lee *et al*. 1998; Krieg *et al*. 2001; Girault *et al*. 2006). Petit *et al*. (1999) demonstrated the synergistic action between an ERE half-site and an imperfect ERE site and the requirement for COUP-TF for ERα to achieve maximum estrogen-stimulation. Moreover, COUP-TFs have been shown to interact with ERα and bind to ERE half-sites (Klinge *et al*. 1997; Mètivier *et al*. 2002). A study by Klinge (1999) indicated that the direct interaction between ERα and a COUP-TF is influenced by ligand and that COUP-TF binding to an ERE half-site is increased by the addition of E_2-liganded ERα, mediating an enhanced transcription rate. Their study also provided evidence that the sequence of an ERE sequence and its immediate flanking regions, influence whether COUP-TF enhances, inhibits, or has no effect on ERα ligand-induced ERE reporter gene expression. Interestingly, a COUP-TF half-site, overlapping the -850 to -845 ERE half-site, has been predicted by bioinformatics analysis (refer to section 4.1.4). Consequently, the involvement of a COUP-TF in *ALAS1* estrogen-responsiveness should therefore be considered and the possibility that the -853T variant, adjacent to a putative ERE half-site/COUP-TF half-site, may mediate the enhancing effect of COUP-TF. Likewise, the following could also be true: Liu *et al*. (1993) reported on the transcriptional repression mediated by COUP-TFs. COUP-TFs have the ability to inhibit ERα action by both direct ERE-binding competition and through protein-protein interactions (Tsai and Tsai 1997). COUP-TF thus binds to consensus EREs and ERE half-sites or interacts directly with ERα, and subsequently inhibits E_2-induced gene expression. This suggests that the presence of the -853T variant might hinder the binding of the COUP-TF to the ERE half-site, making the site available for E_2-activated ERα-ERE binding. This latter hypothesis is, however, unlikely since the enhanced expression is also observed in the wt promoter.

Another consideration is that ERα may act indirectly on the *ALAS1* promoter region, with the -853T variant increasing its action. The indirect interaction of ERα to ERE half-sites have been extensively reported (Anderson 2000; Bahadur *et al*. 2005). In humans, indirect binding of ERα can indeed result in estrogen regulation, predominantly *via* Sp1 (Porter *et al*. 1997; Safe 2001). It has been shown that in a number of cases, Sp1/ERα interactions mediate estrogen inducibility of promoters, especially where half-sites are involved (Petz and Nardulli 2000). Typically, a DNA bend may bring a Sp1 binding site and an ERE site in close proximity, enabling a Sp1-ERα protein complex to bind cooperatively. In addition, estrogenic stimulation induces expression of several genes *via* ERα-Sp1 protein interactions with GC-rich promoter

elements in which Sp1, but not ERα binds to the DNA. A gene promoter may therefore lack a formal ERE site, with the regulatory effect of estrogen mediated through association of the activated ERα with relevant DNA binding transfactors (Dong *et al*. 2006). Wang *et al*. (1998) described another example where a GC-rich promoter sequence, which contains two overlapping Sp1 binding sites, was responsible for ERα-mediated transactivation and required formation of an ERα/Sp1 complex in which only the Sp1 protein bound DNA. In a later study they provided evidence that ERα enhances a Sp1-DNA binding, an action that is hormone independent (Wang *et al*. 2002). Krishnan *et al*. (1994) used the EMSA technique to identify a mechanism where both an ERE half-site and Sp1-like sequence are required for estrogen responsiveness of the Cathepsin D promoter. Of interest is that, several of the regulatory motifs predicted in the *ALAS1* promoter by bioinformatic analyses in this project, were putative Sp1 sites.

This enhancing effect mediated by E_2-activated ERα on the *ALAS1* wild-type promoter and the promoter with the -853T variant, is however severely reduced by the -1253A variant. The decreased expression of pGL(ALAS-ERE/Sp1) compared to pGL(ALAS-ERE) and pGL(ALAS-WT) under normal-, hormone-free- and estrogen-activated ERα conditions, confirms the negative functionality of the -1253A variant. As stated previously, this decrease caused by the -1253A variant, located next to a Sp1-like sequence, may be attributable to a myriad of factors, such as the interruption of the binding of a nearby activator, or creating a binding site for a repressor protein. As mentioned above, this non-consensus Sp1-like site may also work in cooperation with an ERα to contribute to the estrogen-responsiveness of *ALAS1*.

Figure 4.14 Expression of *ALAS1* promoter constructs in the absence and presence of estrogen in HepG2 cells.



A histogram indicating the normalised mean ± SEM expression values of the three pGL2-*ALAS1* promoter constructs in hormone-free culture conditions from three independent transfection experiments (assayed in triplicate). Each construct was co-transfected with ERα and induced with either EtOH or E_2. The promoter-less pGL2-Basic vector was included as negative control. The expression value of pGL(ALAS-WT) induced with EtOH, was used as the reference and assigned a value of 1. Expression of the other constructs was conveyed as a value relative to the reference with * denoting statistically significant differences from the reference, and • denoting significant differences between EtOH and E_2 inductions of a construct (* • $p< 0.05$). The error bars indicate the standard deviation (S.D).

4.4 SUMMARY OF RESULTS

In summary, we have described the nature and extent of nucleotide variation in the human *ALAS1* promoter region by identifying two novel, functional promoter variants and five putative functional ERE half-sites. Direct sequencing analysis revealed the two variants at -853C/T and -1253T/A in the region 5' to the TSS, with no variants in either two ADRES regions, indicating the high degree of sequence conservation of these regions. Although the *in silico* analyses indicated that neither of these novel variants are located within defined regulatory elements, examination of several promoter analysis programs indicated the -853C/T transition is located 5' to one half of an optimal consensus binding site for the estrogen receptor and that the -1253T/A transversion is located 3' to a putative non-consensus Sp1 site. In total, 5 ERE half-sites (perfect and imperfect) were predicted to be distributed unevenly across the 1.29 kb *ALAS1* 5' flanking region, with *in vitro* results of reporter gene assays indicating the estrogen responsiveness of the *ALAS1* promoter, suggesting the functionality of one or more of these ERE's. In the presence of E_2-activated ERα, the -853T variant mediated an additional increase in transcription, while the additional presence of the -1253A variant caused a decrease in transcription in the presence and absence of ERα and estrogen.

In disorders such as VP, attributes such as multiple etiologies, genetic heterogeneity, incomplete penetrance, gene-gene interactions, gene-environment interactions, genetic admixture and genetic drift, all act to obscure the genetic disease modifying component. In some instances, molecular genetic variation does not directly cause the disease, but creates a background of phenotypic susceptibility. It has been proposed that, by using a population where all all individuals have exactly the same disease-causing mutation, such our patient group consisting of individuals all heterozygous for the R59W disease-causing founder mutation, the power of the study to detect minor susceptibility effects, is increased (Warnich *et al*. 2002a). The small sample size and the fact that all patients were not exposed to the same exogenous factors such as estrogen usage, prevented statistical association analysis. Despite the effect of these variants on *ALAS1* transcription (illustrated by the functional studies), no significant trend was observed between these variants and VP symptomatic status.

Nevertheless, the estrogen-responsiveness of the *ALAS1* promoter in HepG2 cells is not unexpected given that ERα is the predominant estrogen receptor found in the liver, and the fact that the liver is the primary site for estrogen metabolism by the cytochrome P450 enzymes. Previously, all reports described the increased expression of *ALAS1* in response to estrogen, as the result of the negative feedback of haeme on *ALAS1*: an increased demand for haeme to produce the increased amount of cytochrome P450 enzymes to metabolize estrogen. Findings from this study, however, indicate a (additional) mechanism where estrogen, *via* ERα, acts on

the *ALAS1* promoter to increase its transcriptional expression. The precise mechanism in which the predicted perfect and imperfect half-sites function in the *ALAS1* promoter, appears to be complex and illustrates the need for further *in vitro* analyses. Also, the precise mechanism whereby the -853T and -1253A variants mediate their modifying effect, calls for further *in vitro* analyses.

CHAPTER 5:

CONCLUSIONS, IMPLICATIONS AND FUTURE PROSPECTS

5. CONCLUSIONS, IMPLICATIONS AND FUTURE PROSPECTS

Identification of transcriptional regulatory elements in promoter regions is of vital interest to biologists since these regions primarily govern the regulation of gene expression. The characterization of functional variants located in these regulatory regions of important candidate genes, is driven by the prospect of exposing phenotypic-modifying variants. In this study we identified two such variants with functional effects in the promoter of *ALAS1*, a candidate modifier gene of the porphyrias, including VP. Moreover, a direct transcriptional response of the *ALAS1* promoter to estrogen, in an *in vitro* environment, has been demonstrated.

Estrogen levels are not constant in an organism and vary greatly between sexes and at different stages of life. In females, the plasma levels of estrogens increase at puberty and also vary throughout the menstrual cycle, with estrogen levels being at its highest prior to ovulation (reviewed in: Gruber *et al*. 2002). At menopause, depletion of the ovarian follicles leads to a steady decline in ovarian estrogen production. In males, estrogens also play an important physiological role, although little is known about the regulation of estrogen production by extragonadal tissues (reviewed in Simpson *et al*. 2000). The two major physiological sources of estrogen are secretion by the ovaries and by biosynthesis in local tissues, catalyzed by enzymes such as aromatase (encoded by the *CYP19* gene). This enzyme is located in estrogen-producing cells in the adrenal glands, ovaries, placenta, testicles, adipose tissue and brain and catalyzes the conversion of testosterone to estradiol. Additionally, numerous dietary- and medicinal supplements may deliver additional estrogen to the body. Results obtained from this study imply that an individual will, during periods of augmented estrogen levels, experience elevated levels of *ALAS1* transcription, which will cause a subsequent increase in the rate of the entire haeme synthesis pathway. This estrogenic effect seems to be enhanced by the presence of the -853T variant. Acute porphyric attacks, such as observed in VP and other porphyrias, are primarily caused by the accumulation of haeme precursors in the presence of a defective enzyme in the haeme pathway. Consequently, an individual with the -853T *ALAS1* promoter variant and a defect in any one of the downstream enzymes in the haeme pathway may experience an added build-up of these haeme precursors during episodes of high estrogen levels, and subsequent enhanced porphyria phenotypic consequences. This theory is consistent with the occurrence of VP symptoms primarily in women and predominantly at the onset of puberty, with acute attacks often precipitated by contraceptive use and the frequency of the acute attacks decreasing with age (Thadani *et al*. 2000; http://www.porphyria-europe.com/01-for-patients/EN/for-patients.pdf). In other types of porphyrias, reports also exist of the association of acute attacks with the menstrual cycle (De Block *et al.* 1999).

The enhanced effect caused by estrogen is, however, eradicated by the presence of the -1253A variant. It is therefore speculated that a porphyria patient with the -1253A *ALAS1* variant may consequently be protected against severe acute symptoms, since the decrease in *ALAS1* transcription mediated by this variant, will slow down expression of this gene, with a possible reduction in the rate of haeme synthesis, and a resultant decrease in the amount of haeme precursors.

The clinical classification of the symptomatic status of a VP patient remains ambiguous and difficult to define, since a patient classified as asymptomatic, may develop symptoms when exposed to precipitating factors. With modifier genes, several variants may interact with each other to function cooperatively, which means that few single variants are universally beneficial (Nadeau 2005). In addition, several modifier genes may operate in unison, such that a particular combination of alleles may have a much stronger effect on the probability to develop VP symptoms, than each individual allele (Souery *et al*. 2001). Several possible mechanisms could account for variations in VP symptomatic expression, including variations in the promoter region of the *ALAS1* gene. In this regard, we have hypothesized that mutations in *cis*-acting elements of *ALAS1* could modulate the phenotype in VP patients and that clinical variability can be determined by the degree of *ALAS1* expression. Consequently, the comprehensive understanding of the transcriptional regulation of *ALAS1* expression is fundamental to understanding the variation in the amount of haeme-precursor accumulation in patients with VP. The two identified variants are considered not to be disease-causing owing to their occurrence in healthy individuals, but nevertheless, altered expression of the normal *ALAS1* allele in the presence of a defective downstream enzyme, is expected to significantly modify VP phenotype. It also remains possible that one of the described variants could exert a much greater or different effect on *ALAS1* gene expression *in vivo* than *in vitro*, since the sequence context of regulatory regions could be very different in chromatin in its natural state compared to in a reporter plasmid. It may therefore also be possible that, in the event of a disruption of an ERE (or other regulatory element), the cell may be adapted to employ an other upstream site, such as observed when the TATA-box of *ALAS1* is mutated (as discussed in section 2.3.1.2) It also remains interesting to see if VP patients, homozygote for the mutant variants, will experience more or less severe phenotypic consequences in response to estrogen stimulation.

Additional *ALAS1* reporter gene assays in other cell lines, such as HeLa cells, will indicate if the observed estrogen-responsiveness of this gene is liver-specific or operational in other tissues. Further studies, analyzing protein-DNA interactions, are now also required to validate findings from this project and confirm the involvement of ERα and the putative ERE half-sites in the estrogen-mediated up-regulation of *ALAS1*. Furthermore, the precise mechanisms involved in the enhancing and reducing effects of the -853T and -1253A variants respectively, also necessitate

protein-DNA binding analyses such as *in vitro* DNase1 footprinting assays and EMSA or *in vivo* analyses such as ChIP assays or a model organism (as described in section 2.4.4.2 and 2.2.4.3). It also remains certain that more regulatory elements are present in the *ALAS1* promoter, of which all may have either functional importance in VP symptomatic expression, when disrupted by sequence variants. Additionally, this study offered novel suggestions into possible future studies regarding *ALAS1* regulation: the importance of the 5' regulatory region of *ALAS1* justifies the screening of the *ALAS1* 3' non-coding region for possible functional modifying variants. The effect of ERα on *ALAS1* regulation also suggests the sequence of the coding and regulatory regions of the gene coding for the human ERα protein as a possible modifier gene of VP symptomatic expression.

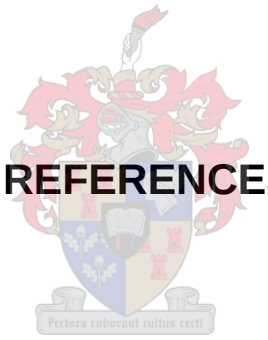
All the aims of this study, as stated on page 4 have been met:

1) The 1.292 kb region of the *ALAS1* gene promoter and the upstream ADRES elements of VP patients were successfully amplified by PCR amplification.
2) These amplified genomic regions were screened for the existence of sequence variants which revealed two novel single base substutions at -853 and -1253.
3) The allelic frequencies of the two sequence variants in the control population were calculated. No obvious trend between the variants and the clinical phenotypes of the VP patients, was observed.
4) The isolated genomic regions were analysed for the existence of putative regulatory elements by bioinformatic analyses and the level of conservation of these areas, determined by phylogenetic footprinting.
5) *In vitro* functional studies were performed:
    5.1) The transcriptional response of the *ALAS1* promoter upon estrogen stimulation, was assessed.
    5.2) The effect of the two identified variants on transcription were assessed under basal- (normal culture-) and drug-induced- (estrogen induction) conditions.

The most significant discoveries from this project are the transcriptional up-regulation of *ALAS1* in response to estrogen and the functionality of the two newly identified promoter variants. These findings describe a novel mechanism of haeme regulation *via ALAS1* in response to estrogen and contributed to unravelling the mode of regulation of *ALAS1*, providing the groundwork for further studies concerning haeme homeostasis. This study was the first of its kind to identify functional regulatory variants in the *ALAS1* promoter region, thereby contributing to the better understanding of how various alleles of a modifier gene may interact to give rise to a variable phenotype. Despite the limitations of this study, such as the small sample size and the lack of

extensive clinical data on the VP patients, our investigations made available important insights into the complexity of *ALAS1* expression, with possible influences on estrogen metabolism and the symptomatic expression of patients suffering from porphyrias, such as VP.

CHAPTER: 6



REFERENCES

6. REFERENCES

6.1 GENERAL REFERENCES

Adams TH and Timberlake WE (1990). Upstream elements repress premature expression of an Aspergillus developmental regulatory gene. Mol Cell Biol. **10**(9):4912-4919.

Agarwal S and Moorchung N (2005). Modifier genes in oligogenic disease. J Nippon Med Sch. **72**(6):326-334.

Ajioka RS, Phillips JD, Kushner JP (2006). Biosynthesis of heme in mammals. Biochim Biophys Acta. **1763**(7):723-36.

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

Ali S, Metzger D, Bornert JM, and Chambon P (1993). Modulation of transcriptional activation by ligand-dependent phosphorylation of the human estrogen receptor A/B region. EMBO J. **12**:1153-1160.

Anderson I and Gorski J (2000). Estrogen receptor α interaction with estrogen response element half-sites from the rat prolactin gene. Biochemistry. **39**:3842-3847.

Anderson KE, Bloomer JR, Bonkovsky HL, Kushner JP, Pierach CA, Pimstone NR, Desnick RJ. (2005). Recommendations for the diagnosis and treatment of the acute porphyrias. Ann Intern Med. **142**(6):439-450.

Sambrook J, Fritsch EF, Maniatis T (1989). Molecular cloning: a laboratory manual (2nd edition). Cold Spring Harbor Laboratory Press.

Angel P, Imagawa M, Chiu R *et al.* (1987) Phorbol ester-inducible genes contain a common cis-element recognized by a TPA-modulated trans-acting factor. Cell. **49**:729-739.

Anolik JH, Klinge CM, Bambara RA, Hilf R (1993). Differential impact of flanking sequences on estradiol- vs 4-hydroxytamoxifen-liganded estrogen receptor binding to estrogen responsive element DNA. J Steroid Biochem Mol Biol. **46**(6):713-730.

Anolik JH, Klinge CM, Brolly CL, Bambara RA, Hilf R (1996). Stability of the ligand-estrogen receptor interaction depends on estrogen response element flanking sequences and cellular factors. J Steroid Biochem Mol Biol. **59**(5-6):413-429.

Antonarakis SE, Irkin SH, Cheng TC, Scott AF, Sexton JP, Trusko SP, Charache S, *et al*. (1984). ß-Thalassemia in American blacks: novel mutations in the 'TATA' box and an acceptor splice site. Proc Natl Acad Sci USA. **81**:1154-1158.

Arbini AA, Pollak ES, Bayleran JK, High KA (1996). Functional characterization of the human factor VII 5' flanking region. J Biol Chem. **271**:1738-1747.

Arnold R, Burcin M, Kaiser B, Muller M, Renkawitz R (1996). DNA bending by the silencer protein NeP1 is modulated by TR and RXR. Nucleic Acids Res. **24**(14):2640-2647.

Ausubel F, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (ed.). (1997). Current protocols in molecular biology, vol. 1, chapter 9: Introduction of DNA into mammalian cells. John Wiley & Sons, Inc., NY.

Badminton MN and Elder GH (2005). Molecular mechanisms of dominant expression in porphyria. J Inherit Metab Dis. **28**:277-286.

Bahadur U, Ganjam GK, Vasudevan N, Kondaiah P (2005). Estrogen regulation of chicken riboflavin carrier protein gene is mediated by ERE half sites without direct binding of estrogen receptor. Mol Cell Endocrinol. **231**(1-2):1-11.

Barberis A and Petrascheck M (2003). Transcription activation in eukaryotic cells. Encyclopedia Of Life Sciences Macmillan Publishers Ltd, Nature Publishing Group. p1-7.

Barkhem T, Haldosen LA, Gustafsson J-A, Nilsson S (2002). pS2 Gene expression in HepG2 cells: complex regulation through crosstalk between the estrogen receptor alpha, an estrogen-responsive element, and the activator protein 1 response element. Mol Pharmacol. **61**(6):1273-1283.

Beato M (1989). Gene regulation by steroid hormones. Cell. **8**:335-344.

Beato M, Herrlich P, Schütz G (1995) Steroid hormone receptors: many actors in search of a plot. Cell. **83**:851-857.

Bell AC and Felsenfeld G (1999). Stopped at the border: boundaries and insulators. Curr Opin Genet Dev. **9**:191-198.

Belting HG, Shashikant CS, Ruddle FH. (1998). Modification of expression and cis-regulation of Hoxc8 in the evolution of diverged axial morphology. Proc Natl Acad Sci USA. **95**(5):2355-2360

Berry M, Nunez A-M, Chambon P (1989). Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. Proc Natl Acad Sci USA. **86**:1218-1222.

Bickers DR (1982). Environmental and drug factors in hepatic porphyria. Acta Derm Venereol Suppl. (Stockh). **100**:29-41.

Bishop DF (1990). Two different genes encode 3-aminolevulinate synthase in humans: nucleotide sequences of cDNAs for the housekeeping and erythroid genes. Nucleic Acids Res. **18**:7187-7188.

Blackwood EM and Kadonaga JT (1998). Going the distance: a current view of enhancer action. Science. **281**:60-63.

Bloomer JR, Logan GM, Weimer MK, Pierach CA (1990). Bile porphyrin and porphyrinogen levels in erythropoietic protoporphyria and variegate porphyria. Mol Aspects Med. **11**:54–55.

Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM (2003). Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science. **299**(5611):1391-1394.

Bonkovsky HL (2005). Neurovisceral Porphyrias: What a Hematologist Needs to Know. Hematology Am Soc Hematol Educ Program. 24-30.

Botstein D and Risch N (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. **33**:228-237.

Bourdeau V, Deschênes J, Métivier R, Nagai Y, Nguyen D, Bretschneider N, Gannon F, *et al*. (2004). Genome-wide identification of high-affinity estrogen response elements in human and mouse. Mol Endocrinol. **18**(6):1411-1427.

Braidotti G, Borthwick IA, May BK (1993). Identification of regulatory sequences in the gene for 5-aminolevulinate synthase from rat. J Biol Chem. **268**:1109-1117.

Brodie MJ, Moore MR, Goldberg A (1977). Enzyme abnormalities in the porphyrias. Lancet. **2**(8040):699-701.

Brown M and Sharp PA (1990). Human estrogen receptor forms multiple protein-DNA complexes. J Biol Chem. **265**:11238-11243.

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, NISC Comparative Sequencing Program, Green ED, *et al*. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. **13**(4):721-731.

Bucher P (1999). Regulatory elements and expression profiles. Curr Opin Struct Biol. **9**(3):400-407.

Buchner DA, Trudeau M, Meisler MH (2003). SCNM1, a putative RNA splicing factor that modifies disease severity in mice. Science. **301**(5635):967-969.

Buckland PR (2004). Allele-specific gene expression differences in humans. Hum Mol Genet. **13**(2):255-260.

Bullido MJ, Artiga MJ, Recuero M, Sastre I, Garcia MA, Aldudo J, Lendon C, *et al*. (1998). A polymorphism in the regulatory region of APOE associated with risk for Alzheimer's dementia. Nat Genet. **18**(1):69-71.

Burbach JP, Lopes da Silva S, Cox JJ, Adan RA, Cooney AJ, Tsai MJ, Tsai SY (1994). Repression of estrogen-dependent stimulation of the oxytocin gene by chicken ovalbumin upstream promoter transcription factor I. J Biol Chem. **269**(21):15046-15053.

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.

Cable EE, Miller TG, Isom HC (2000). Regulation of haem metabolism in rat hepatocytes and hepatocyte cell lines: delta-aminolevulinic acid synthase and haem oxygenase are regulated by different haem-dependent mechanisms. Arch Biochem Biophys. **384**(2):280-295.

Cáceres JFand Kornblihtt AR (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet. **18**:186-193.

Carroll SB (2000). Endless forms: The evolution of hene regulation and morphological diversity. Cell. **101**:577-580.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, *et al*. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. **21**(13):2933-2942.

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**(4):R32.

Chen ZG, Yu KL, Zheng HM, Dong KW (1999). Estrogen receptor-mediated repression of gonadotropin-releasing hormone (gnRH) promoter activity in transfected CHO-K1 cells. Mol Cell Endocrinol. **158**(1-2):131-142.

Cho H and Katzenellenbogen BS (1993). Synergystic activation of estrogen receptor-mediated transcription by estradiol and protein kinase activators. Mol Endocrinol. **7**:441-452.

Cho RJ, Huang M, Campbell MJ, Dong H, Steinmetz L, Sapinoso L, Hampton G, *et al*. (2001). Transcriptional regulation and function during the human cell cycle. Nat Genet. **27**:48-54.

Clark AR and Docherty K (1993). Negative regulation of transcription in eukaryotes. Biochem J. **296**:521-541.

Colantoni A, Emanuele MA, Kovacs EJ, Villa E, Van Thiel DH (2002). Hepatic estrogen receptors and alcohol intake. Mol Cell Endocrinol. **193**:101-104.

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

Corrigall AV, Hift RJ, Davids LM, Hancock V, Meissner D, Kirsch RE, Meissner PN (2000). Homozygous variegate porphyria in South Africa: genotypic analysis in two cases. Mol Genet Metab. **69**:323-330.

Cotter PD, Baumann M, Bishop DF (1992). Enzymatic defect in "X-linked" sideroblastic anemia: molecular evidence for erythroid delta-aminolevulinate synthase deficiency. Proc Natl Acad Sci USA. **89**:4028-4032.

Coulson RMR and Ouzounis CA (2003). The phylogenetic diversity of eukaryotic transcription. Nucleic Acids Res. **3**(2):653-660.

Couse JF, Lindsey J, Grandien K, Gustafsson J-A, Korach KS (1997). Tissue distribution and quantitative analysis of estrogen receptor-a (ERα) and estrogen receptor-b (ERβ) messenger ribonucleic acid in the wild-type and ER-α knockout mouse. Endocrinology. **138**:4613-4621.

Cowley SM, Hoare S, Mosselman S, Parker MG (1997). Estrogen receptors α and β form heterodimers on DNA. J Biol Chem. **272**:19858-19862.

Crawford DL, Segal JA, Barnett JL (1999). Evolutionary analysis of the TATA-less proximal promoter function. Mol Biol Evol*.* **6**:194-207.

Crossley M and Brownlee GG (1990). Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. Nature. **345**:444-446.

Cvoro A, Tzagarakis-Foster C, Tatomer D, Paruthiyil S, Fox MS, Leitman DC (2006). Distinct roles of unliganded and liganded estrogen receptors in transcriptional repression. Mol Cell. **21**(4):555-564.

Dailey TA, Woodruff JH, Dailey HA (2004). An examination of mitochondrial protein targeting of heme synthetic enzymes. In vivo identification of three functional heme-responsive motifs in 5-aminolevulinate synthase. Biochem J. **386**(Pt 2):381-386.

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh C, Minokawa T, *et al*. (2002). A genomic regulatory network for development. Science. **295**(5560):1669-1678.

Davies RC, Calvio C, Bratt E, Larsson SH, Lamond AI, Hastie ND (1998). WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. Genes & Dev. **12**:3217-3225.

Day RS (1986). Variegate porphyria. Semin Dermatol. **5**:138-154.

De Block CEM, De Leeuw IH, Van Gaal LF (1999). Premenstrual attacks of acute intermittent porphyria: hormonal and metabolic aspects – a case report. Eur J Endo. **141**:50-54.

De Bruin D, Zaman Z, Liberatore RA, Ptashne M (2001). Telomere looping permits gene activation by a downstream UAS in yeast. Nature. **409**:109-113.

De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, Gibbons RJ (2006). A Regulatory SNP Causes a Human Genetic Disease by Creating a New Transcriptional Promoter. Science **312**(5777):1215-1217.

De Gracia J, Mata F, Alvarez A, Casals T, Gatner S, Vendrell M, *et al.* (2005). Genotype-phenotype correlation for pulmonary function in cystic fibrosis. Thorax. 60(7):558-563.

De Matteis F (1972).The effects of drugs on the activities of 5-aminolaevulinate synthetase and other enzymes in the pathway of haem biosynthesis. Biochem J. **130**(2):52P-53P.

De Pontual L, Pelet A, Trochet D, Jaubert F, Espinosa-Parrilla Y, Munnich A, *et al*. (2006). Mutations of the RET gene in isolated and syndromic Hirschsprung's disease in human disclose major and modifier alleles at a single locus. J Med Genet. **43**(5):419-423.

De Villiers JNP, Kotze MJ, van Heerden CJ, Sadie A, Gardner HFJ, Liebenberg J, van Zyl R, *et al*. (2005). Overrepresentation of the founder PPOX gene mutation R59W in a South African patient with severe clinical manifestation of porphyria Exp Dermatol. **14**:50-55.

Dean G (1972). The Porphyrias. A Story of Inheritance and Environment. Philadelphia: JB Lippincott (pub) (2nd ed.).

Dermitzakis ET and Clark AG (2002). Evolution of TF binding sites in mammalian gene regulatory regions: conservation and turn-over. Mol Biol Evol. 19:1114

Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C, Deutsch S, Stevenson BJ, *et al*. (2002). Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature. **420**(6915):578-582.

Devuyst O, Persu A, and Vo-Cong M-T (2003). Autosomal dominant polycystic kidney disease: modifier genes and endothelial dysfunction. Nephrol Dial Transplant. **18**:2211-2215.

Deybach J-C, Puy H, Robreau AM, Lamoril J, Da Silva V, Grandchamp B, Nordmann Y (1996). Mutations in the protoporphyrinogen oxidase gene in patients with variegate porphyria. Hum Mol Genet. **5**:407-410.

Dipple KM and McCabe ERB (2000). Phenotypes of patients with “simple” mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. Am J Hum Genet. **66**:1729-1735.

Donaldson IJ and Gottgens B (2006). Evolution of candidate transcriptional regulatory motifs since the human-chimpanzee divergence. Genome Biol. **7**(6):R52.

Dork T, Wulbrand U, Richter T, Neumann T, Wolfes H, Wulf B, Maass G, Tummler B (1991). Cystic fibrosis with three mutations in the cystic fibrosis transmembrane conductance regulator gene. Hum Genet. **87**(4):441-446.

Drew PD, Ades IZ (1989). Regulation of the stability of chicken embryo liver N-aminolevulinate synthase mRNA by hemin. Biochem Biophys Res Commun. **162**:102-107.

Driscoll MD, Klinge CM, Hilf R, Bambara RA (1996). Footprint analysis of estrogen receptor binding to adjacent estrogen response elements. J Steroid Biochem Molec Biol. **58**(1):45-61.

Driscoll MD, Sathya G, Muyan M, Klinge CM, Hilf R, Bambara RA (1998). Sequence requirements for estrogen receptor binding to estrogen response elements. J Biol Chem. **273**(45):29321-29330.

Dröge P and Müller-Hill B (2001). High local protein concentrations at promoters: strategies in prokaryotic and eukaryotic cells. Bioessays. **23**:179-183.

Drumm ML (2001). Modifier genes and variation in cystic fibrosis. Respir Res. **2**:125-128.

Drumm ML, Konstan MK, Schluchter MD, Handler A, Pace RP, Zou F, Zariwala M, *et al*. (2005). Genetic modifiers of lung disease in cystic fibrosis. N Engl J Med. **353**(14):1443-1453.

Duarte J, Perrière G, Laudet V, Robinson-Rechavi M (2002). NUREBASE: database of nuclear hormone receptors. Nucl Acid Res. **30**(1):364-368.

Duncan R, Faggart MA, Roger AJ, Cornell NW (1999). Phylogenetic analysis of the 5-aminolevulinate synthase gene. Mol Biol Evol. **16**:383-396.

Duret L and Bucher P (1997). Searching for regulatory elements in human noncoding sequences. Curr Opin Struct Biol. **7**(3):399-406.

Duret L, Dorkeld F, Gautier C (1993). Strong conservation of noncoding sequences during vertebrates evolution-potential involvement in post-transcriptional regulation of gene expression. Nucleic Acids Res. **21**:2315-2322.

Edgar AJ, Losowsky MS, Noble JS, Wickramasinghe SN (1997). Identification of an arginine452 to histidine substitution in the erythroid 5-aminolaevulinate synthetase gene in a large pedigree with X-linked hereditary sideroblastic anaemia. Eur J Haematol. **58**:1-4.

Elder GH (1980). The porphyrias: clinical chemistry, diagnosis and methodology. Clin Hematol. **9**:371–398.

Elder GH (1998). Genetic Defects in the Porphyrias: Types and Significance. Clin Dermatol. **16**:225-233.

Elder GH, Hift RJ, Meissner PN (1997). The acute porphyrias. Lancet. **349**:1613-1617.

Escriva H, Delaunay F, Laudet V (2000). Ligand binding and nuclear receptor evolution. Bioessays. **22**:717-727.

Estivill X, Ortigosa L, Pérez-Frias J, Dapena J, Ferrer J, Peña J, *et al*. (1995) Clinical characteristics of 16 cystic fibrosis patients with the missense mutation R334W, a pancreatic insufficiency mutation with variable age of onset and interfamilial clinical differences. Hum Genet. **95**:331-336.

Evans RM (1988). The steroid and thyroid receptor superfamily. Science. **240**:889-895.

Fessele S, Maier H, Zischek C, Nelson PJ, Werner T (2002). Regulatory context is a crucial part of gene function. Trends Genet. **18**(2):60-63.

Fickett JW and Hatzigeorgiou AG (1997). Eukaryotic promoter recognition. Genome Res. **7**(9):861-878.

Fitze G, Cramer J, Ziegler A, Schierz M, Schreiber M, Kuhlisch E, Roesner D, Schackert HK (2002). Association between c135G/A genotype and RET proto-oncogene germline mutations and phenotype of Hirschsprung's disease. Lancet. **359**(9313):1200-1205.

Frank J, Jugert FK, Merk HF, Kalka K, Goerz G, Anderson K, Bickers DR, *et al*. (2001). A spectrum of novel mutations in the protoporphyrinogen oxidase gene in 13 families with variegate porphyria. J Invest Dermatol. **116**(5):821-823.

Frank J, McGrath J, Lam H, Graham RM, Hawk JL, Christiano AM (1998) Homozygous variegate porphyria: identification of mutations on both alleles of the protoporphyrinogen oxidase gene in a severely affected proband. J Invest Dermatol. **110**:452-455.

Fraser DJ, Podvinec M, Kaufmann MR, Meyer UA (2002). Drugs mediate the transcriptional activation of the 5-aminolevulinic acid synthase (*ALAS1*) gene via the chicken xenobiotic-sensing nuclear receptor (CXR). J Biol Chem. **277**(38):34717-34726.

Fraser DJ, Zumsteg A, Meyer UA (2003). Nuclear receptors constitutive androstane receptor and pregnane X receptor activate a drug-responsive enhancer of the murine 5-aminolevulinic acid synthase gene. J Biol Chem. **278**:39392-39401.

Fried C, Ahnert P, Stadler PF (2004). Distribution of SNPs in genes, phylogenetic footprints and non-coding regions. Poster presented at The Nineteenth International Course & Conference on the Interfaces among Mathematics, Chemistry & Computer Sciences. June 21-26, 2004, Dubrovnik, Croatia.

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

Fukada H, Casas A, Batlle A (2005). Aminolevulinic acid: from its unique biological function function to its star in photodynamic therapy. Int J Biochem Cell Biol. **37**:272-276.

Gaffney DJ and Keightley PD (2004). Unexpected conserved non-coding DNA blocks in mammals. Trends Genet. **20**(8):332-337.

Galas DJ and Schmitz A (1978). DNAase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5**:3157–3170.

Galbraith RA (1987). Hormonal regulation of the heme pathway. Ann N Y Acad Sci. **514**:256-260.

Ganley ARD, Hayashi K, Horiuchi T, Kobayashi T (2005). Identifying gene-independent noncoding functional elements in the yeast ribosomal DNA by phylogenetic footprinting. Proc Natl Acad Sci USA. **102**(33):11787-11792.

Garber K, Smith KT, Reines D, Warren ST (2006). Transcription, translation and fragile X syndrome. Curr Opin Genet Dev. **16**(3):270-275.

Garcia-Arencibia M, Davila N, Campion J, Carmen Carranza M, Calle C (2005). Identification of two functional estrogen response elements complexed with AP-1-like sites in the human insulin receptor gene promoter. J Steroid Biochem Mol Biol. **94**:1-14.

Garcia-Ramirez M, Rocchini C, Ausio J (1995). Modulation of chromatin folding by histone acetylation. J Biol Chem. **270**(30)17923-17928.

Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Grossfeld PD, Srivastava D (2005). Mutations in NOTCH1 cause aortic valve disease. Nature. **437**(7056):270-4.

Gaston K and Jayaraman PS (2003). Transcriptional repression in eukaryotes: repressors and repression mechanisms. Cell Mol Life Sci. **60**(4):721-741

Gerasimova TI and Corces VG (2001). Chromatin insulators and boundaries: effects on transcription and nuclear organization. Annu Rev Genet. **35**:193-208.

Germain P, Altucci L, Bourguet W, Rochette-Egly C, Gronemeyer H (2003). Nuclear receptor superfamily: principles of signaling. Pure Appl. Chem. **75**(11-12):1619-1664.

Giger U and Meyer UA (1981). Induction of d-aminolevulinate synthase and cytochrome P-450 hemoproteins in hepatocytes culture. Effect of glucose and hormones. J. Biol. Chem. **256**:11182-11190.

Giles RH, Peters DJ, Breuning MH (1998). Conjunction dysfunction: CBP/p300 in human disease. Trends Genet. **14**:178–183.

Giono LE, Varone CL, Ca´nepa ET (2001). 5-Aminolaevulinate synthase gene promoter contains two cAMP-response element (CRE)-like sites that confer positive and negative responsiveness to CRE-binding protein (CREB). Biochem J. **353**:307-316.

Girault I, Bieche I, Lidereau R (2006). Role of estrogen receptor alpha transcriptional coregulators in tamoxifen resistance in breast cancer. Maturitas. **54**(4):342-351.

Gonzalez FJ (1988). The molecular biology of cytochrome P450s. Pharmacol Rev. **40**:242-288.

Goodson M, Jonas BA, Privalsky MA (2005). Corepressors: custom tailoring and alterations while you wait. Nucl Recept Signal. **3**:e003

Gordon N (1999). The acute porphyrias. Brain Dev. **21**:373-377.

Gouya L, Puy H, Robreau A, Lyoumi S, Lamoril J, Da Silva V, Grandchamp B, Deybach J. (2004). Modulation of penetrance by the wild-type allele in dominantly inherited erythropoietic protoporphyria and acute hepatic porphyrias. Hum Genet **114**:256-262.

Grabe N (2002). AliBaba2: Context Specific Identification of Transcription Factor Binding Sites. In Silico Biol. **2**(1):S1-1.

Grad YH, Roth FP, Halfon MS, Church GM (2004). Prediction of similarly acting *cis*-regulatory modules by subsequence profiling and comparative genomics in *D.melanogaster* and *D.pseudoobscura.* Bioinformatics. **20**:2738-2750.

Granick S (1966). The induction in vitro of the synthesis of δ-aminolevulinate synthetase in chemical porphyria: a response to certain drugs, sex hormones and foreign chemicals. J Biol Chem. **241**:1359-1375.

Granick S and Kappas A (1967a). Steroid control of porphyrin and heme biosynthesis: a new biological function of steroid hormone metabolites. Proc Natl Acad Sci USA. **57**(5):1463-1467.

Granick S and Kappas A (1967b). Steroid induction of porphyrin synthesis in liver cell culture. I. Structural basis and possible physiological role in the control of heme formation. J Biol Chem. **242**:4587-4593.

Grasemann H, van's Gracesande KS, Buscher R, Nicola K, Silverman ES, Palmer LJ, *et al.* (2003). Endothelial nitric oxide synthase variants in cystic fibrosis lung disease. Am J Respir Crit Care Med. **167**:390-394.

Gray S and Levine M (1996).Transcriptional repression in development. Curr Opin Cell Biol. 8(3):358-364.

Green MR (2005). Eukaryotic transcription activation: right on target. Mol Cell. **18**(4):399-402.

Green S, Walter P, Kumar V, Krust A, Bornert JM, Argos P, Chambon, P (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb A. Nature. **320**:134-139.

Groenewald JZ, Liebenberg J, Groenewald IM, Warnich L (1998). Linkage disequilibrium analysis in a recently founded population: evaluation of the variegate porphyria founder in South African Afrikaners. Am J Hum Genet. **62**(5):1254-1258.

Gross U, Honcamp M, Daume E, Frank M, Dusterberg B, Doss MO (1995). Hormonal oral contraceptives, urinary porphyrin excretion and porphyrias.Horm Metab Res. **27**(8):379-83.

Gruber CJ, Tschugguel W, Schneeberger C, Huber JC (2002). Production and actions of estrogens. N Engl J Med. **346**(5):340-352.

Grunstein M (1997). Histone acetylation in chromatin structure and transcription. Nature **389**:349-352.

Guberman AS, Scassa ME, Giono LE, Varone CL, Canepa ET (2003). Inhibitory effect of AP-1 complex on 5-aminolevulinate synthase gene expression through sequestration of cAMP-response element protein (CRE)-binding protein (CBP) coactivator. J Biol Chem. **278**(4):2317-2326.

Guengerich FP (2003). Cytochromes P450, drugs, and diseases. Mol Interv. **3**(4):194-204

Gumucio DL, Shelton DA, Zhu W, Millinoff D, Gray T, Bock JH, Slightom JL, Goodman M (1996). Evolutionary strategies for the elucidation of cis and trens factors that regulate the developmental switching programs of the p-like globin genes. Mol Phylogenet Evol. **5**:18-32.

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

Hall JM and McDonnell DP (1999) The estrogen receptor β-isoform (ERβ) of the human estrogen receptor modulates ERα transcriptional activity and is a key regulator of the cellular response to estrogens and antiestrogens. Endocrinology. **140**:5566-5578.

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.

Hamilton JW, Bement WJ, Sinclair PR, Sinclair JF, Alcedo JA, Wetterhahn KE (1991). Heme regulates hepatic 5-aminolevulinate synthase mRNA expression by decreasing mRNA half-life and not by altering its rate of transcription. Arch Biochem Biophys. **289**:387-392.

Hanahan D (1983). Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol. **166**:557-580.

Handschin C, Blättler S, Roth A, Looser R, Oscarson M, Kaufmann MR, Podvinec M, Gnerre C, Meyer UA (2004). Nuclear receptors: the evolution of drug-activated nuclear receptors: one ancestral gene diverged into two xenosensor genes in mammals. Nuc Recept. **2**(1):7.

Handschin C, Lin J, Rhee J, Peyer AK, Chin S, Wu PH, Meyer UA, Spiegelman BM. (2005). Nutritional regulation of hepatic heme biosynthesis and porphyria through PGC-1 alpha. Cell. **122**(4):505-515.

Hardison RC (2000). Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet. **16**:369-372.

Harendza S, Lovett DH, Panzer U, Lukacs Z, Kuhnl P, Stahl RAK (2003). Linked common polymorphisms in the gelatinase a promoter are associated with diminished transcriptional response to estrogen and genetic fitness. J Biol Chem. **278**(23):20490-20499.

Hassan S, Duong B, Kim KS, Miles MF (2003). Pharmacogenomic analysis of mechanisms mediating ethanol regulation of dopamine beta-hydroxylase. J Biol Chem. **278**(40):38860-38869.

Hegele RA (1997). Small genetic effects in complex diseases: A review of regulatory sequence variants in dyslipoproteinemia and atherosclerosis. Clin Biochem. **30**(3):183-188.

Heinemeyer T, Chen X, Karas H, Kel AE, Kel OV, Liebich I, Meinhardt T, Reuter I, Schacherer F, Wingender E (1999). Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. Nucleic Acids Res. **27**:318-322.

Hift RJ (2005). Hepatology and porphyria. Student notes (Department of Medicine). Cape Town: University of Cape Town.

Hift RJ and Meissner PN (2005). An analysis of 112 acute porphyric attacks in Cape Town, South Africa: evidence that acute intermittent porphyria and variegate porphyria differ in susceptibility and severity. Medicine (Baltimore). **84**(1):48-60.

Hift RJ, Davidson BP, van der Hooft C, Meissner DM, Meissner PN (2004b). Plasma fluorescence scanning and fecal porphyrin analysis for the diagnosis of variegate porphyria: precise determination of sensitivity and specificity with detection of protoporphyrinogen oxidase mutations as a reference standard. Clin Chem. **50**(5):915-923.

Hift RJ, Meissner D, Meissner PN (2004a). A systematic study of the clinical and biochemical expression of variegate porphyria in a large South African family. Br J Dermatol. **151**:465-471.

Hift RJ, Meissner PN, Corrigall AV, Ziman MR, Petersen LA, Meissner DM, Davidson BP, *et al*. (1997). Variegate porphyria in South Africa, 1688–1996 - new developments in an old disease. S Afr Med J. **87**:722-730.

Hift RJ, Meissner PN, Todd G, Kirby P, Bilsland D, Collins P, *et al*. (1993). Homozygous variegate porphyria: an evolving clinical syndrome. Postgrad Med J. **69**(816):781-786. Erratum in: Postgrad Med J. 1994; **70**(829):855.

Hochschild A and Dove SL (1998). Protein-protein contacts that activate and repress prokaryotic transcription. Cell. **92**:597-600.

Honkakoski P and Negishi M (2000). Regulation of cytochrome P450 (CYP) genes by nuclear receptors. Biochem J. **347**:321-337.

Hoogendoorn B, Coleman SL, Guy CA, Smith K, Bowen T, Buckland PR, O'Donovan MC (2003). Functional analysis of human promoter polymorphisms. Hum Mol Genet. **12**(18):2249-2254.

Horwitz KB, Jackson TA, Bain DL, Richer JK, Takimoto GS and Tung L (1996) Nuclear receptor coactivators and corepressors. Mol Endocrinol. **10**:1167-1177.

Huang J, Li X, Maguire CA, Hilf R, Bambara RA, Muyan M (2005). Binding of estrogen receptor beta to estrogen response element in situ is independent of estradiol and impaired by its amino terminus. Mol Endocrinol. 19(11):2696-2712.

Huang YS and Chuang DT (1998). Mechanism for basal expression of rat mitochondrial branched-chain-2-oxo-acid dehydrogenase kinase [corrected]. Biochem J. **334**(Pt 3):713-22. Erratum in: Biochem J. (1999) **339**(Pt 3):775.

Hubert D, Bienvenu T, Desmazes-Dufeu N, Fajac I, Lacronique J, Matran R, Kaplan JC, Dusser DJ (1996). Genotype-phenotype relationships in a cohort of adult cystic fibrosis patients. Eur Respir J. **9**:2207-2214

Hull J, Thomson AH (1998). Contribution of genetic factors other than CFTR to disease severity in cystic fibrosis. Thorax. **53**:1018-1021

Ike A, Ohta H, Onishi M, Iguchi N, Nishimune Y, Nozaki M (2004). Transient expression analysis of the mouse ornithine decarboxylase antizyme haploid-specific promoter using in vivo electroporation. FEBS Lett. **559**(1-3):159-164.

Itzkovitz S, Tlusty T, Alon U (2006). Coding limits on the number of transcription factors. BMC Genomics. 19(7):239.

Jacolot S, Le Gac G, Scotet V, Quere I, Mura C, Ferec C. (2004). HAMP as a modifier gene that increases the phenotypic expression of the HFE pC282Y homozygous genotype. Blood. **103**(7):2835-2840.

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

Jakacka M, Ito M, Weiss J, Chien P-Y, Gehm BD, Jameson JL (2001). Estrogen receptor binding to DNA is not required for its activity through the nonclassical AP1 pathway. J Biol Chem. **276**:13615-13621.

James MFM and Hift RJ (2000). Porphyrias. Br J Anaesth. **85**(1):143-153.

Janson CG and During MJ (2001). Viral vectors as part of an integrated functional genomics program. Genomics. **78**(1-2):3-6.

Javahery R, Khachi A, Lo K, Zenzie-Gregory B, Smale ST (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol. **14**(1):116-127.

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

Johannsen W (1909). Elemente der exacten Erblichkeitslehre. Gustav Fischer, Jena.

Jones DR, Schmidt RJ, Pickard RT, Foxworthy PS, Eacho PI (2002). Estrogen receptor-mediated repression of human hepatic lipase gene transcription. J Lipid Res. **43**:383-391.

Jones N (1990). Structure and function of transcription factors. Semin Cancer Biol. **1**(1):5-17.

Jordan PM and Shemin D (1972). The Enzymes (Boyer PD, 3rd edn), Vol. 7, p348-356, Academic Press, New York.

Jover R, Hoffmann K, Meyer UA (1996). Induction of 5-Aminolevulinate Synthase by drugs is independent of increased apocytochrome P450 synthesis. Biochem Biophys Res Commun. **226**:152-157.

Jura J, Węgrzyn P, Jura J, Koj A (2006). Regulatory mechanisms of gene expression: complexity with elements of deterministic chaos. Acta Biochim Pol. **53**(1):1-9.

Kahn A, Cottreau D, Bernard JF, Boivin P (1975). Post translational modifications of glucose-6-phosphate dehydrogenase in human leukemias.
Biomedicine. 22(6):539-594.

Kamboh MI, Minster RL, Kenney M, Ozturk A, Desai PP, Kammerer CM, DeKosky ST (2006). Alpha-1-antichymotrypsin (ACT or SERPINA3) polymorphism may affect age-at-onset and disease duration of Alzheimer's disease. Neurobiol Aging. **27**(10):1435-1439.

Kamboh MI, Sanghera DK, Ferrell RE, DeKosky ST (1995). APOE*4-associated Alzheimer's disease risk is modified by alpha 1-antichymotrypsin polymorphism. Nat Genet. **10**(4):486-488. Erratum in: Nat Genet. **11**(1):104.

Kappas A and Granick S (1968). Steroid induction of porphyrin synthesis in liver cell culture. II. The effects of heme, uridine diphosphate glucuronic acid, and inhibitors of nucleic acid and protein synthesis on the induction process. J Biol Chem. **243**:346-351.

Karin M (1990). Too many transcription factors: positive and negative interactions. New Biol. **2**(2):126-31.

Kastner P, Mark M, Chambon P (1995). Nonsteroid nuclear receptors: what are genetic studies telling us about their role in real life? Cell. **83**:859-869.

Kato S, Tora L, Yamauchi J, Masushige S, Bellard M, Chambon P (1992). A far upstream estrogen response element of the ovalbumin gene contains several half-palindromic 5'-TGACC-3' motifs acting synergistically. Cell. **68**:731-742.

Katti MV, Sakharkar MK, Ranjekar PK, Gupta VS (2000). TRES: comparative promoter sequence analysis. Bioinformatics. **16**(8):739-740.

Kauppinen R (2005). Porphyrias. Lancet. **365**(9455):241-52.

Kauppinen R and Mustajoki P (1992). Prognosis of acute porphyria: occurrence of acute attacks, precipitating factors, and associated diseases. Medicine. **71**:1-13.

Kauppinen R, Timonen K, Laitinen E, Kuusisto K, Ahola H, Tenhunen R, Mustajoki P (1997) Molecular genetics and clinical characteristics of variegate porphyria. Acta Haematol. **98**(Suppl 1):96.

Kauppinen R, Timonen K, von und zu Fraunberg M, Laitinen E, Ahola H, Tenhunen R, Taketani S, Mustajoki P (2001). Homozygous variegate porphyria: 20 y follow-up and characterization of molecular defect. J Invest Dermatol. **116**(4):610-613.

Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res. **31**(13):3576-3579.

Kerem BS, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravati A (1989). Identification of the cystic fibrosis gene: genetic analysis. Science. **245**:1073-1080.

Kim J, Petz, LN, Ziegler YS, Wood JR, Potthoff SJ , Nardulli AM (2000). Regulation of the estrogen-responsive pS2 gene in MCF-7 human breast cancer cells. J Steroid Biochem Mol Biol. **74**:57-68.

Kirsch RE, Meissner PN, Hift RJ (1998). Variegate porphyria. Semin Liver Dis. **18**:33-41.

Klein-Hitpass L, Ryffel GU, Heitlinger E, Cato A.C (1988) A 13bp palindrome is a functional estrogen responsive element and interacts specifically with estrogen receptor. Nucleic Acid Res. **16**:647-663.

Klinge CM (1999). Role of estrogen receptor ligand and estrogen response element sequence on interaction with chicken ovalbumin upstream promoter transcription factor (COUP-TF). J Steroid Biochem Mol Biol. **71**:1-19.

Klinge CM (2000). Estrogen receptor interaction with co-activators and co-repressors. Steroids. **65**:227-251.

Klinge CM (2001). Estrogen receptor interaction with estrogen response elements. Nucleic Acid Res. **29**(14):2905-2919.

Klinge CM, Bodenner DL, Desai D, Niles RM, Traish AM (1997). Binding of type II nuclear receptors and estrogen receptor to full and half-site estrogen response elements in vitro. Nucleic Acids Res. **25**(10):1903-1912.

Knight JC (2003). Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. Clin Sci (Lon). **104**:493–501.

Knight JC (2005). Regulatory polymorphisms underlying complex disease traits J Mol Med **83**:97-109.

Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP (2003). *In vivo* characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nat Genet. **33**:469-475.

Knoepfler PS, and Kamps MP (1995). The pentapeptide motif of hox proteins is required for cooperative DNA binding with pbx1, physically contacts pbx1 and enhances DNA binding by pbx1. Mol Cell Biol. **15**:5811-5819.

Knowles BB, Howe CC, Aden DP (1980). Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. Science **209**:497-499.

Koivisto UM, Palvimo JJ, Janne OA, Kontula K (1994). A single-base substitution in the proximal Sp1 site of the human low density lipoprotein receptor promoter as a cause of heterozygous familial hypercholesterolemia. Proc Natl Acad Sci USA. **91**:10526-10530.

Kolluri S, Sadlon TJ, May BK and Bonkovsky, BM (2005). Heme Repression of the Housekeeping 5-Aminolevulinate Synthase Gene in the LMH Hepatoma Cell Line. Biochemical J. **392**(Pt1):173-180.

Kornblihtt AR, de La Mata M, Fededa JP, Muñoz MJ, Nogués G (2004). Multiple links between transcription and splicing. RNA. **10**:1489-1498.

Krieg SA, Krieg AJ, Shapiro DJ (2001). A unique downstream estrogen responsive unit mediates estrogen induction of proteinase inhibitor-9, a cellular inhibitor of IL-1beta- converting enzyme (caspase 1). Mol Endocrinol. **15**(11):1971-1982.

Krishnan V, Wang X, Safe S (1994). Estrogen receptor-Sp1 complexes mediate estrogen-induced cathepsin D gene expression in MCF-7 human breast cancer cells. J Biol Chem. **269**(22):15912-15917.

Kuiper GG, Shughrue PJ, Merchenthaler I, Gustafsson JA (1998). The estrogen receptor beta subtype: a novel mediator of estrogen action in neuroendocrine systems. Front Neuroendocrinol. **19**:253-286.

Kumar R and Thompson EB (1999). The structure of the nuclear hormone receptors. Steroids. **64**:310-319.

Kumar V and Chambon P (1988). The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. Cell. **55**(1):145-56.

Kushner PJ, Agard DA, Greene GL, Scanlan TS, Shiau AK, Uht RM, Webb P (2000). Estrogen receptor pathways and AP-1. J Steroid Biochem Mol Biol. **74**:311-317.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, *et al*. (2001). Initial sequencing and analysis of the human genome. Nature. **409**:860-921.

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

Latchman DS (1996a). Inhibitory Transcription Factors. Int J Biochem Cell Bid. **28**(9):965-914.

Latchman DS (1996b). Transcription-factor mutations and disease. N Engl J Med. **334**:28-33.

Le Corvoisier P, Park HY, Rockman HA (2003). Modifier genes and heart failure. Minerva Cardioangiol. **51**(2):107-120.

Le Goff P, Montano MM, Schodin DJ, and Katzenellenbogen BS (1994). Phosphorylation of the human estrogen receptor. Identification of hormone-regulated sites and examination of their influence on transcriptional activity. J Biol Chem. **269**:4458-4466.

Lecha M, Badenas C, Puig S, Orfila J, Mila M, To-Figueras J, *et al*. (2006). Genetic studies in variegate porphyria in Spain. Identification of gene mutations and family study for carrier detection. J Eur Acad Dermatol Venereol. 20(8):974-979.

Lee SK, Choi HS, Song MR, Lee MO, Lee JW (1998). Estrogen receptor, a common interaction partner for a subset of nuclear receptors. Mol Endocrinol. **12**(8):1184-1192.

Lee TI and Young RA (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet. **34**:77-137.

Lehmann AR (2001). The xeroderma pigmentosum group D (*XPD*) gene: one gene, two functions, three diseases. Genes Dev. **15**:15-23.

Lejeune F, Cavaloc Y, Stevenin J (2001). Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene: identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. J Biol Chem. **276**(11):7850-7858.

Lemon B and Tjian R (2000). Orchestrated response: a symphony of transcription factors for gene control. Genes & Dev. **14**:2551-2569.

Lettice LA, Horikoshi T, Heaney SJH, van Baren MJ, van der Linde HC, Breedveld GJ, Marijke (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci USA. **99**(11):7548-7553.

Levine M and Tjian R (2003). Transcription regulation and animal diversity. Nature. **424**:147-151.

Li C, Briggs MR, Ahlborn TE, Kraemer FB, Liu J (2001). Requirement of Sp1 and estrogen receptor alpha interaction in 17ß-estradiol-mediated transcriptional activation of the low density lipoprotein receptor gene expression. Endocrinology. **142**:1546-1553.

Li Q and Johnston SA (2001). Are all DNA binding and transcription regulation by an activator physiologically relevant? Mol Cell Biol. **21**(7):2467-2474.

Lieto LD, Borrego F, You CH, Coligan JE (2003). Human CD94 gene expression: dual promoters differing in responsiveness to IL-2 or IL-15. J Immunol. **171**(10):5277-5286.

Lindberg RL, Martini R, Baumgartner R, Erne B, Borg J, Zielasek J, Ricker K, *et al*. (1999). Motor neuropathy in porphobilinogen deaminase-deficient mice imitates the peripheral neuropathy of human acute porphyria. J Clin Invest. **103:**1127-1134.

Liu LG, Yan H, Zhang W, Yao P, Zhang XP, Sun XF, Nussler AK (2004). Increased haeme oxygenase-1 and decreased delta-aminolevulinate synthase expression in the liver of patients with acute liver failure. Int J Mol Med. **14**:1001-1005.

Liu Y and Teng CT (1992). Estrogen response module of the mouse lactoferrin gene contains overlapping chicken ovalbumin upstream promoter transcription factor and estrogen receptor-binding elements. Mol Endocrinol. **6**:355-364.

Liu Y, Yang N, Teng CT (1993). COUP-TF acts as a competitive repressor for estrogen receptor-mediated activation of the mouse lactoferrin gene. Mol Cell Biol. **13**:1836-1846.

Liu YP, Lien WC, Fang CC, Lai TI, Chen WJ, Wang HP (2005). ED presentation of acute porphyria. Am J Emerg Med. **23**(2):164-167.

Lopez AJ (1998). Alternative splicing of pre-mRNA: Developmental consequences and mechanisms of regulation. Annu Rev Genet. **32**:279-305.

Love JD, Gooch JT, Nagy L, Chatterjee VKK, Schwabe JWR (2000). Transcriptional repression by nuclear receptors: mechanisms and role in disease. Biochem Soc Trans. **28**(4):390-396.

Maguire DJ, Day AR, Borthwick IA, Srivastava G, Wigley PL. May BK, Elliott WH (1986). Nucleotide sequence of the chicken 5-aminolevulinate synthase gene. Nucleic Acids Res. **14**:1379-1391.

Mangelsdorf DJ, Thummel C. Beato M, Herrlich P, Schültz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, Evans RM (1995). The nuclear receptor superfamily: the second decade. Cell. **83**:835-839.

Mao C and Shapiro DJ (2000). A histone deacetylase inhibitor potentiates estrogen receptor activation of a stably integrated vitellogenin promoter in HepG2 Cells. Endocrinology. **141**(7):2361-2369.

Martin MP, Dean M, Smith MW, Winkler C, Gerrard B, Michael NL, Lee B, *et al*. (1998). Genetic acceleration of AIDS progression by a promoter variant of CCR5. Science. **282**(5395):1907-1911.

Martinez E and Wahli W (1989). Cooperative binding of estrogen receptor to imperfect estrogen-responsive DNA elements correlates with their synergistic hormone-dependent enhancer activity. EMBO J. **8**:3781-3791.

Thompson EW, Reich R, Shima TB, Albini A, Graf J, Martin GR, Dickson RB, Lippman ME (1988). Differential regulation of growth and invasiveness of MCF-7 breast cancer cells by antiestrogens. Cancer Res. **48**:6764-6768.

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

Matlin AJ, Clark F, Smith CW (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol. **6**(5):386-98.

May BK, Dogra SC, Sadlon TJ, Bhasker CR, Cox TC, Bottomley SS (1995). Molecular regulation of heme biosynthesis in higher vertebrates. Prog Nucleic Acid Res Mol Biol. 51:1-51.

Meissner PN, Dailey TA, Hift RJ, Ziman M, Corrigall AV, Roberts AG, Meissner D, *et al*. (1996) A R59W mutation in protoporphyrinogen oxidase results in decreased enzyme activity and is prevalent in South Africans with variegate porphyria. Nat Genet. **13**:95–97.

Menuet A, Le Page Y, Torres O, Kern L, Kah O, Pakdel F (2004). Analysis of the estrogen regulation of the zebrafish estrogen receptor (ER) reveals distinct effects of ERalpha, ERbeta1 and ERbeta2. J Mol Endocrinol. **32**(3):975-86.

Mercier B, Verlingue C, Lissens W, Silber SJ, Novelli G, Bonduelle M, Audrézet MP, Férec C (1995). Is congenital bilateral absence of vas deferens a primary form of cystic fibrosis? Analyses of the CFTR gene in 67 patients. Am J Hum Genet. **56**:272-277.

Mètivier R, Gay FA, Hubner MR, Flouriot G, Salbert G, Gannon F, *et al*. (2002). Formation of a hER alpha-COUP-TFI complex enhances hER alpha AF-1 through Ser118 phosphorylation by MAPK. EMBO J. **21**(13):3443-3453.

Meyer UA, Schuurmans MM, Lindberg RL (1998). A review of the pathogenesis of the neurological manifestations of the acute porphyrias. Semin Liver Dis. **18**:43-52.

Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. **16**(3):1215.

Miller M (1997). Tools for population genetic analysis (TPGA) 1.3: A windows programme for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.

Mitchell PJ and Tjian R (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science. **245**:371-378.

Mitchison NA (2001). Polymorphism in regulatory gene sequences. Genome Biol. **2**(1):Comment2001.

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.

Montano MM, Muller V, Trogaugh A, Katzenellenbogen BS (1995). The carboxyterminal F domain of the human estrogen receptor: role in transcriptional activity of the receptor and the effectiveness of antiestrogens as estrogen antagonists. Mol Endocrinol. **9**:814-825.

Moore JT, McKee DD, Slentz-Kesler K, Moore LB, Jones SA, Horne EL, Su JL, Kliewer SA, Lehmann JM, Willson TM (1998). Cloning and characterization of human estrogen receptor beta isoforms. Biochem Biophys Res Commun. **247**:75-78.

Moore MR (1998). The biochemistry of heme synthesis in porphyria and in the porphyrinurias. Clin Dermatol. **16**:203-223.

Mueller PR and Wold B (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science. **246**:780-786.

Munakata H, Sun JY, Yoshida K, Nakatani T, Honda E, Hayakawa S, Furuyama K, Hayashi N (2004). Role of the heme regulatory motif in the heme-mediated inhibition of mitochondrial import of 5-aminolevulinate synthase. J Biochem (Tokyo). **136**(2):233-238.

Mustajoki P and Nordmann Y (1993). Early administration of heme arginate for acute porphyric attacks. Arch Intern Med. **153**(17):2004-8.

Nadeau JH (2001). Modifier genes in mice and humans. Nat Rev Genet. **2**:165-174.

Nadeau JH (2003a). Modifying the message. Science. **301**:927-302.

Nadeau JH (2003b). Modifier genes and protective alleles in humans and mice. Curr Opin Genet Dev. **13**:290-295.

Nadeau JH (2005). Listening to genetic background noise. N Engl J Med. **352**(15):1598-1599.

Nardulli AM, Grobner C, Cotter D (1995). Estrogen receptor-induced DNA bending: orientation of the bend and replacement of an estrogen response element with an intrinsic DNA bending sequence. Mol Endocrinol. **9**(8):1064-1067.

Nardulli AM, Romine LE, Carpo C, Greene GL, Rainish B (1996). Estrogen receptor affinity and location of consensus and imperfect estrogen response elements influence transcription activation of simplified promoters. Mol Endocrinol. **10**:694-704.

Nastiti W, Norbert K, Stephan I (2004). Biology of heme in health and disease. Current Medicinal Chemistry. **11**(8):981-986.

Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, *et al*. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics. **6**(1):1-42.

Nikolova G and Vilain E (2006). Mechanisms of Disease: transcription factors in sex determination—relevance to human disorders of sex development. Nat Clin Pract Endocrinol Metab. **2**:231-238.

Nilsson S, Mäkelä S, Treuter E, Tujague M, Thomsen J, Andersson G, Enmark E, *et al*. (2001). Mechanisms of Estrogen Action. Physiol Rev. 81: 1535-1565.

Nobrega MA, Ovcharenko I, Afzal V, Rubin EM (2003). Scanning human gene deserts for long-range enhancers. Science. **302**(5644):413.

Nordmann Y and Puy H (2002). Human hereditary hepatic porphyrias. Clinica Chimica Acta **325**:17-37.

Nuclear Receptors Nomenclature Committee (1999). A unified nomenclature system for the nuclear receptor superfamily. Cell. **97**:161-163.

Nurse GT, Weiner JS, Jenkins T (1985). The growth of hybrid communities. In: The Peoples of Southern Africa and Their Affinities. Oxford, UK, Oxford Clarendon, p 218.

O'lone R, Frith MC, Karlsson EK, Hansen U (2004). Genomic targets of nuclear estrogen receptors. Mol Endocrinol. **18**(8):1859-1875.

Ogawa S, Inoue S, Watanabe T, Hiroi H, Orimo A, Hosoi T, Ouchi Y, Muramatsu M (1998). The complete primary structure of human estrogen receptor beta (hER beta) and its heterodimerization with ER alpha in vivo and in vitro. Biochem Biophys Res Commun **243**:122-126.

Ogbourne S and Antalis TM (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J. **331**:1-14.

Olefsky JM (2001). Nuclear receptor minireview series. J Biol Chem. **276**(40):36863-36864.

Oliver M (2004). From SNPs to function: the effect of sequence variation on gene expression. Focus on "A survey of genetic and epigenetic variation affecting human gene expression". Physiol Genomics. **16**:182-183.

Onate SA, Tsai SY, Tsai M-J, O'Malley BW (1995). Sequence and characterization of a co-activator for the steroid hormone receptor superfamily. Science. **270**:1354-1357.

Orlando V, Strutt H, Paro R (1997). Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods. **11**(2):205-214.

Orphanides G, Lagrange T, Reinberg D (1996). The general transcription factors of RNA polymerase II. Genes Dev. **10**:2657-2683.

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

Paranjape SM, Kamakaka RT, Kadonaga JT (1994). Role of chromatin structure in the regulation of transcription by RNA polymerase II. Annu Rev Biochem. **63**:265-297.

Pastinen T and Hudson TJ (2004). Cis-acting regulatory variation in the human genome. Science. **306**:647-650.

Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, *et al*. (2004). A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics. **16**:184-193.

Peale Jr FV, Ludwig LB, Zain S, Hilf R, Bambara RA 1988 Properties of a high-affinity DNA binding site for estrogen receptor. Proc Natl Acad Sci USA 85:1038–1042.

Pearson H (2006). Genetics: What is a gene? Nature. **441**:398-401.

Pedersen AG, Baldi P, Chauvin Y, Brunak S (1999). The biology of eukaryotic promoter prediction - a review. Comput Chem. **23**(3-4):191-207.

Pelletier G (2000). Localization of androgen and estrogen receptors in rat and primate tissues. Histol. Histopathol. **15**:1261-1270.

Peltonen L and McKusick VA (2001). Genomics and medicine. Dissecting human disease in the postgenomic era. Science. 291(5507):1224-1229.

Pereira FA, Qiu Y, Tsai MJ, Tsai SY (1995). Chicken ovalbumin upstream promoter transcription factor (COUP-TF): expression during mouse embryogenesis. J Steroid Biochem Mol Biol. 53(1-6):503-508.

Pereira FA, Tsai MJ and Tsai SY (2000). COUP-TF orphan nuclear receptors in development and differentiation. Cell Mol Life Sci. 57:1388-1398.

Persu A, Stoenoiu MS, Messiaen T, Davila S, Robino C, El-Khattabi O, Mourad M, *et al*. (2002). Modifier effect of ENOS in autosomal dominant polycystic kidney disease. Hum Mol Genet. **11**(3):229-241.

Petit FG, Métivier R, Valotaire Y, Pakdel F (1999). Synergism between a half-site and imperfect estrogen-responsive element, and cooperation with COUP-TF1 are required for estrogen receptor (ER) to achieve a maximal estrogen-stimulation of rainbow trout ER gene. Eur J Biochem. **259**:385-395.

Petz LN and Nardulli AM (2000). Sp1 binding sites and an estrogen response element half-site are involved in regulation of the human progesterone receptor α promoter. Mol Endocrinol. **14**(7): 972-985.

Pike AC, Brzozowski AM, Hubbard RE, Bonn T, Thorsell AG, Engstrom O, Ljunggren J, *et al*. (1999). Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. EMBO J. **18**:4608-4618.

Podvinec M, Handschin C, Looser R, Meyer UA (2004). Identification of the xenosensors regulating human 5-aminolevulinate synthase. Proc Natl Acad Sci USA. **101**(24):9127-9132.

Poh-Fitzpatick MB (1980). A plasma fluorescence marker for variegate porphyria. Arch Dermatol. **116**:543-547.

Poh-Fitzpatrick MB (1998). Clinical features of the porphyrias. Clin Dermatol. **16**(2):251-264.

Ponka P (1997). Tissue-specific regulation of iron metabolism and heme synthesis: distinct control mechanisms in erythroid cells. Blood. **89**:1–25.

Ponka P (1999). Cell biology of heme. Am J Med Sci. **318**:241-256.

Porter LE, Elm MS, Van Thiel DH, Dugas MC, Eagon PK (1983). Characterization and quantitation of human hepatic estrogen receptor. Gastroenterology. **84**(4):704-712.

Privalsky ML (2004). The Role of corepressors in transcriptional regulation by Nuclear Hormone Receptors. Annu Rev Physiol. **66**:315-60.

Porter W, Saville B, Hoivik D, Safe S (1997). Functional synergy between the transcription factor Sp1 and the estrogen receptor. Mol Endocrinol. **11**(11):1569-1580.

Prestridge DS (1991). SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. Comput Appl Biosci. **7**(2):203-206.

Prestridge DS (1996). SIGNAL SCAN 4.0: additional databases and sequence formats. Comput Appl Biosci.**12**:157-160.

Ptashne M and Gann A (2002). Genes and Signals. Cold Spring Harbor, NY., Cold Spring Harbor Laboratory Press.

Qin C, Singh P, Safe S (1999).Transcriptional activation of insulin-like growth factor-binding protein-4 by 17ß-estradiol in MCF-7 cells: role of estrogen receptor–Sp1 complexes. Endocrinology. **140**:2501-2508.

Quandt K, Frech K, Karas H, Wingender E, Werner T (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. **23**:4878-4884.

Rasmuson M (2002). Review: the genotype-phenotype link. Hereditas. **136**:1-6.

Ray P, Ghosh S, Zhang D, Ray R (1997). Repression of interleukin-6 gene expression by 17β-estradiol: inhibition of the DNA-binding activity of the transcription factors NF-IL6 and NF-κB by the estrogen receptor. FEBS Lett. **409**:79-85.

Roberts AG and Elder GH (2001). Alternative splicing and tissue-specific transcription of human and rodent ubiquitous 5- aminolevulinate synthase (*ALAS1*) genes. Biochim Biophys Acta. **1518**:95-105.

Roberts AG, Redding SJ, Llewellyn DH (2005). An alternatively-spliced exon in the 5'-UTR of human *ALAS1* mRNA inhibits translation and renders it resistant to haem-mediated decay. FEBS Lett. **579**:1061-1066.

Robinson-Rechavi M, Carpentier A-S, Duffraisse M, Laudet V (2001). How many nuclear hormone receptors are there in the human genome? Trends Genet. **17**(10):554-556

Robledo M, Gil L, Pollán M, Arancha Cebrián, Ruìz S, Azañedo M, *et al*. (2003). Polymorphisms G691S/S904S of *RET* as Genetic Modifiers of MEN 2A. Cancer Res. **63**(8):1814-1817.

Romeo G and McKusick VA (1994). Phenotypic diversity, allelic series and modifier genes. Nat Genet. **7**(4):451-453.

Rosenfeld MG and Glass CK (2001). Coregulator Codes of Transcriptional Regulation by Nuclear Receptors. J Biol Chem. **27**6(40)36865-36868.

Rozen S and Helen J. 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.

Ruau D, Duarte J, Ourjdal T, Perriere G, Laudet V, Robinson-Rechavi M (2004). Update of NUREBASE: nuclear hormone receptor functional genomics. Nucleic Acids Res. **32**(Database issue):D165-167.

Ryder K, Silver S, DeLucia AL, Fanning E, Tegtmeyer P (1986). An altered DNA conformation in origin region I is a determinant for the binding of SV40 large T antigen. Cell. **44**(5)719-725.

Sadlon TJ, Dell'Oso T, Surinya KH, May BK (1999). Regulation of erythroid 5-aminolevulinate synthase expression during erythropoiesis. Int J Biochem Cell Biol. **31**(10):1153-1167.

Safe S (2001). Transcriptional activation of genes by 17β-estradiol through estrogen receptor-Sp1 interactions. Vitam Horm. **62**:231-252.

Saltzman AG, Weinmann R (1989). Promoter specificity and modulation of RNA polymerase II transcription. FASEB J. **3**(6):1723-1733.

Salvatore F, Scudiero O, Castaldo G (2002). Genotype–phenotype correlation in cystic fibrosis: the role of modifier genes. Am J Med Genet. **111**:88-95.

Sanchez R, Nguyen D, Rocha W, White JH, Mader S (2002). Diversity in the mechanisms of gene regulation by estrogen receptors. Bioessays. **24**:244-254.Sanchez-Garcia I (1997). Consequences of chromosomal abnormalities in tumour development. Annu Rev Genet. **31**:429-453.

Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004).JASPAR: an open-access database for eukaryotic transcription factor binding profiles.Nucleic Acids Res. **32**(Database issue):D91-4.

Sassa S and Kappas A (2000). Molecular aspects of the inherited porphyrias. Journal of Internal Medicine **247**:169-178.

Sassa S and Nagai T (1996). The role of heme in gene expression. Int J Hematol. **63**:167-178.

Sathya G, Li W, Klinge CM, Anolik JH, Hilf R, Bambara RA (1997). Effects of multiple estrogen responsive elements, their spacing, and location on estrogen response of reporter genes. Mol Endocrinol. **11**(13):1994-2003.

Scassa ME, Guberman AS, Ceruti JM, Canepa ET (2004). Hepatic nuclear factor 3 and nuclear factor 1 regulate 5-aminolevulinate synthase gene expression and are involved in insulin repression. J Biol Chem. **279**:28082-28092.

Scassa ME, Guberman AS, Varone CL, Ca´nepa ET (2001). Phosphatidylinositol 3-kinase and ras/mitogen-activated protein kinase signaling pathways are required for the regulation of 5-aminolevulinate synthase gene expression by insulin. Exp Cell Res. **271**:201-213.

Scassa ME, Varone CL, Montero L, Canepa ET (1998). Insulin inhibits delta-aminolevulinate synthase gene expression in rat hepatocytes and human hepatoma cells. Exp Cell Res. **244**(2):460-469.

Schena M (1989). The evolutionary conservation of eukaryotic gene transcription. Experientia. **45**(10):972-983.

Schneider J, Sassa S, Kappas A (1983). Metabolism of estradiol in liver cell culture. Differential responses of C-2 and C-16 oxidations to drugs and other chemicals that induce selective species of cytochrome P450. J Clin Invest. **72**:1420-1426.

Schneider-Yin X and Minder EI (2006). Swiss patients with variegate porphyria have unique mutations. Swiss Med Wkly. **136**(31-32):515-519.

Scholz A, Truss M, Beato M (1998). Hormone-induced recruitment of Sp1 mediates estrogen activation of the rabbit uteroglobin gene in endometrial epithelium. J Biol Chem. **273**:4360-4366.

Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP (1998). Congenital heart disease caused by mutations in the transcription factor NKX2-5. Science. **281**:108-111.

Schultz JH (1874). Fall von Pemphigus leprosus compliciert durch Lepra visceralis. Thesis. University of Greifswald.

Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell. **103**:843-852.

Shanley BC, Neethling AC, Percy VA, Carstens M. (1975). Neurochemical aspects of porphyria. Studies on the possible neurotoxicity of delta-aminolaevulinic acid. S Afr Med J. **49**(14):576-80.

Sheffield VC, Stone EM, Carmi R (1998). Use of isolated inbred human populations for identification of disease genes. Trends Genet. **14**:391–396.

Shibata H, Spencer TE, Onate SA, Jenster G, Tsai SY, Tsai MJ, O'Malley BW (1997). Role of co-activators and co-repressors in the mechanism of steroid/thyroid receptor action. Recent Prog Horm Res. **52**:141-164.

Shou M, Korzekwa KR, Brooks EN, Krausz KW, Gonzalez FJ, Gelboin HV (1997). Role of human hepatic cytochrome P450 1A2 and 3A4 in the metabolic activation of estrone. Carcinogenesis. **18**(1):207-214.

Silverman LM and Mifflin TM (2005). Genotype–Phenotype Correlations (II): Assessing the Influence of Sequence Variants on the Clinical Phenotype in Multifactorial Disorders. Clin Chem. **51**(6):929-930.

Simeonia E, Winkelmannc BR, Hoffmannd MM, Fleurya S, Ruize J, Kappenbergera L, März W, Vassallia G (2004). Association of RANTES G-403A gene polymorphism with increased risk of coronary arteriosclerosis. Eur Heart J. **25**:1438-1446.

Simpson E, Rubin G, Clyne C, Robertson K, O'Donnell L, Jones M, Davis S (2000). The role of local estrogen biosynthesis in males and females. Trends Endocrinol Metab. **11**(5):184-188.

Skak K and Michelsen BK (1999). The TATA-less rat GAD65 promoter can be activated by Sp1 through non-consensus elements. Gene. **236**(2):231-41.

Slavotinek A and Biesecker LG (2003). Genetic modifiers in human development and malformation syndromes, including chaperone proteins. Hum Mol Genet. **12**(Spec 1):R45-50.

Slieker MG, Sanders EA, Rijkers GT, Ruven HJ, van der Ent CK (2005). Disease modifying genes in cystic fibrosis. J Cyst Fibros. Suppl **2**:7-13.

Smale ST (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim Biophys Acta. **1351**:73-88.

Smale ST (2001). Core promoters: active contributors to combinatorial gene regulation. Genes & Dev. **15**(19):2503-2508.

Smale ST and Baltimore D (1989). The "initiator" as a transcription control element. Cell. **57**(1):103-113.

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

Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000). dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res. **28**:352-355.

Souery D, Rivelli SK, Mendlewicz J (2001). Molecular genetic and family studies in affective disorders: state of the art. J Affect Disord. **62**(1-2):45-55.

Spana C and Corces VG (1990). DNA bending is a determinant of binding specificity for a *Drosophila* zinc finger protein. Genes Dev. **4:** 1505-1515.

Spiegelman BM and Heinrich R (2004). Biological Control through Regulated Transcriptional Coactivators. Cell. **119**:157-167.

Srivastava G, Hansen AJ, Bawden MJ, May BL (1990). Hemin administration to rats reduces levels of hepatic mRNAs for phenobarbitone-inducible enzymes. Mol Pharmacol. **38**:486-493.

Stegmaier P, Kel AE, Wingender E (2004). Systematic DNA-binding domain classification of transcription factors. Genome Inform. **15**(2):276-286.

Steyn I (2002). Investigation of the possible influences of candidate modifier genes on the clinical expression of variegate porphyria (VP). Stellenbosch: University of Stellenbosch, Thesis presented for M.Sc.

Stormo GD (2000). DNA binding sites: representation and discovery. Bioinformatics. **16**:16-23.

Swoap SJ (1998). In vivo analysis of the myosin heavy chain IIB promoter region. Am J Physiol Cell Physiol. **274**:C681-C687.

Szutorisz H, Dillon N, Tora L (2005). The role of enhancers as centres general transcription factor recruitment. Trends Biochem Sci. **30**(11):593-599.

Taft RJ and Mattick JS (2003). Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. Genome Biol. **5**(1):P1.1-P1.22.

Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT (1988). Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J Mol Bio. **203**:439-455.

Taketani S (2005). Acquisition, mobilization and Utilization of cellular iron and heme: Endless findings and growing evidence of tight regulation. Tohoku J Exp Med. **205**:297-318.

Tam SP, Archer TK, Deeley RG (1985). Effects of estrogen on apolipoprotein secretion by the human hepatocarcinoma cell lines, HepG2. J Biol Chem **260**:1670–1675.

Taylor AH and Al-Azzawi F (2000). Immunolocalization of oestrogen receptor beta in human tissues. J Mol Endocrinol. **24**:145-155.

Tenhunen R and Mustajoki P (1998). Acute porphyria: treatment with heme. Semin Liver Dis **18**:53-55.

Thadani H, Deacon A, Peters T (2000). Diagnosis and management of porphyria. BMJ. **320**:1647-1651.

The International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. Nature. **431**:931-945.

Thunnel K, Floderus Y, Henrichson A, Moore MR, Meissner P, Sinclair J (1992). Alcoholic beverages in acute porphyria. J Stud Alcohol. **53**:272-276.

Tjian R (1995). Molecular machines that control genes. Sci Am. **272**(2):54-61.

Tompa M (2001). Identifying functional elements by comparative DNA sequence analysis. Genome Res. **11**(7):1143-1144.

Tran P, Zhang XK, Salbert G, Hermann T, Lehmann JM, Pfahl M (1992). COUP orphan receptors are negative regulators of retinoic acid response pathways. Mol Cell Biol. **12**(10):4666-4676.

Tripodi M, Filosa A, Armentano M, Studer M (2004). The COUP-TF nuclear receptors regulate cell migration in the mammalian basal forebrain. Development. **131**(24):6119-6129.

Tsai SY and Tsai MJ (1997). Chick Ovalbumin Upstream Promoter-Transcription Factors (COUP-TFs): Coming of Age. Endocr Rev. **18**(2):229-240.

Tschudy D, Waxman A, Collins A (1967). Oscillations of hepatic δ-aminolevulinic acid synthetas produced by estrogen: a possible role of “rebound induction” in biological clock mechanisms. Proc Natl Acad Sci USA. **58**:1944-1948.

Tsiftsoglou AS, Tsamadou AI, Papadopoulou LC (2006). Heme as key regulator of major mammalian cellular functions: Molecular, cellular, and pharmacological aspects. Pharmacol Ther. **111**(2):327-345.

Tsuchiya Y, Nakajima M, Yokoi T (2005). Cytochrome P450-mediated metabolism of estrogens and its regulation in human. Cancer Lett. **227**:115-124.

Tupler R, Perini G, Green MR (2001). Expressing the human genome. Nature. **409**:832-833.

Uddin RK and Singh (2006b). cis-Regulatory sequences of the genes involved in apoptosis, cell growth, and proliferation may provide a target for some of the effects of acute ethanol exposure. Brain Res. **1088**(1):31-44.

Uddin RK and Singh SM (2006a). Ethanol-responsive genes: identification of transcription factors and their role in metabolomics. Pharmacogenomics J. 2006 May 2; [Epub ahead of print].

Valasek MA and Repa JJ (2005). The power of real-time PCR. Adv Physiol Educ. **29**:151-159.

Van Hellemont R, Monsieurs P, Thijs G, de Moor B, Van de Peer Y, Marchal K (2005). A novel approach to identifying regulatory motifs in distantly related genomes.Genome Biol. **6**(13):R113.

Van’t Hooft FM, von Bahr SJF, Silveira A, Iliadou A, Eriksson P, Hamsten A (1999). Two common, functional polymorphisms in the promoter region of the β-Fibrinogen gene contribute to regulation of plasma fibrinogen concentration. Arterioscler Thromb Vasc Biol. **19**:3063-3070.

Villard J (2004). Transcription regulation and human diseases. Swiss Med Wkly. **134**:571–579.

Virbasius JV and Scarpulla RC (1994). Activation of the human mitochondrial transcription factor A gene by nuclear respiratory factors: a potential regulatory link between nuclear and mitochondrial gene expression in organelle biogenesis. Proc Natl Acad Sci USA. **91**(4):1309-1313.

Von Und Zu Fraunberg M, Timonen K, Mustajoki P, Kauppinen R (2002). Clinical and biochemical characteristics and genotype – phenotype correlation in Finnish variegate porphyria patients. Eur J Hum Genet. **10**:649-657.

Vyhlidal C, Samudio I, Kladde MP, Safe S (2000). Transcriptional activation of transforming growth factor alpha by estradiol: requirement for both a GC-rich site and an estrogen response element half-site. J Mol Endocrinol. **24**:329-338.

Wang F, Hoivik D, Pollenz R, Safe S (1998). Functional and physical interactions between the estrogen receptor Sp1 and nuclear aryl hydrocarbon receptor complexes. Nucleic Acids Res. **26**(12):3044-3052.

Wang F, Samudio I, Safe S (2002). Transcriptional activation of rat creatine kinase B by 17β-estradiol in MCF-7 cells involves an estrogen responsive element and GC-rich sites. J Cell Biochem **84**:156-172.

Wang JC and Giaever GN (1988). Action at a distance along a DNA. Science. **240**(4850):300-304.

Wang LH, Tsai SY, Cook RG, Beattie WG, Tsai M-J, O'Malley BW (1989). COUP transcription factor is a member of the steroid receptor superfamily. Nature. **340**:163-166.

Wang Z, Xiao X, Van Nostrand E, Burge CB (2006). General and Specific Functions of Exonic Splicing Silencers in Splicing Control. Mol Cell. **23**(1):61-70.

Wanscher JH (1975). An analysis of Wilhelm Johannsen's genetical term "genotype" 1909-26. Hereditas. **79**(1):1-4.

Warnich L, Kimberg M, Kotze MJ, Ohashi T, Taketani S, Louw BJ (2002a). Haplotype analysis excludes the functional protoporphyrinogen oxidase promoter polymorphism -1081G>A as a modifying factor in the clinical expression of variegate porphyria. Cell Mol Biol. **48**:57-60.

Warnich L, Kotze MJ, Groenewald IM, Groenewald JZ, van Brakel MG, van Heerden CJ, de Villiers JN, *et al*. (1996b). Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South African families with variegate porphyria. Hum Mol Genet. **3**:981-984.

Warnich L, Meissner PN, Hift RJ, Louw JH, van Heerden CJ, Retief AE (1996a). Mapping of the variegate porphyria (VP) gene: contradictory evidence for linkage between VP and microsatellite markers at chromosome 14q32. Hum Genet. **97**:690-692.

Warnich L, Waso HF, Groenewald IM, Bester AE, de Villiers JN, Kotze MJ, Lynch AG, Louw JH (2002b). Single nucleotide polymorphisms of the protoporphyrinogen oxidase gene: inter-population heterogeneity of allelic variation. Mol Cell Probes. **15**:217-221.

Warren AJ (2002). Eukaryotic transcription factors. Curr Opin Struct Biol. **12**(1):107-14.

Watson CJ, Pierach CA, Bossenmaier I, Cardinal R (1977). Postulated deficiency of hepatic heme and repair by hematin infusions in the "inducible" hepatic porphyrias. Proc Natl Acad Sci USA. **74**:2118-2120.

Weatherall DJ (2001). Phenotype–genotype relationships in monogenic disease: lessons from the thalassaemias. Nature Rev Genet. **2**:245-255.

Whatley SD, Puy H, Morgan RR, Robreau AM, Roberts AG, Nordmann, *et al*. (1999). Variegate porphyria in Western Europe: identification of PPOX gene mutations in 104 families, extent of allelic heterogeneity, and absence of correlation between phenotype and type of mutation. Am J Hum Genet. **65**(4):984-994.

Wilke N, Sganga MW, Gayer GG, Hsieh KP, Miles MF (2000). Characterization of promoter elements mediating ethanol regulation of hsc70 gene transcription. J Pharmacol Exp Ther. **292**(1):173-180.

Wiman Å, Harper P, Floderus Y (2003). Nine novel mutations in the protoporphyrinogen oxidase gene in Swedish families with variegate porphyria. Clin Genet. **64**:122-130.

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

Winter WE (2003). Newly defined genetic diabetes syndromes: maturity onset diabetes of the young. Rev Endocr Metab Disord. **4**:43-51.

With TK (1980). A short history of porphyrins and the porphyrias. Int J Biochem. **11**:189-200.

Wittkopp PJ (2005). Genomic sources of regulatory variation in cis and in trans. Cell Mol Life Sci. **62**(16):1779-83.

Wolfson S, Bartczak A, Bloomer JR (1979). Effect of endogenous heme generation on delta-aminolevulinic acid synthase activity in rat liver mitochondria J Biol Chem. **254**:3543-3546.

Woo AJ, Dods JS, Susanto E, Ulgiati D, Abraham LJ (2002). A proteomics approach for the identification of DNA binding activities observed in the electrophoretic mobility shift assay. Mol Cell Proteomics. **1**:472-478.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003). The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol. **20**(9):1377-1419.

Wright S (1993). Regulation of Eukaryotic Gene Expression by Transcriptional Attenuation Molecular Biology of the Cell **4**:661-668.

Wurtz J-M, Bourguet W, Renaud J-P, Vivat V, Chambon P, Moras D, Gronemeyer H (1996). A canonical structure for the ligand-binding domain of nuclear receptors. Nat Struct Biol. **3**:87-94.

Xing W, Archer TK (1998) Upstream stimulatory factors mediate estrogen receptor activation of the cathepsin D promoter. Mol Endocrinol. **12**:1310-1321.

Yamagata K, Furuta H, Oda N, Kaisaka PJ, Menzel S, Cos NJ, *et al*. (1996). Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). Nature. **384**:458-460.

Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW (2002). Allelic variation in human gene expression. Science. **297**(5584):1143.

Yan Q, Bykhovskaya Y, Li R, Mengesha E, Shohat M, Estivill X, Fischel-Ghodsian N, Guan MX (2006). Human TRMU encoding the mitochondrial 5-methylaminomethyl-2-thiouridylate- methyltransferase is a putative nuclear modifier gene for the phenotypic expression of the deafness-associated 12S rRNA mutations. Biochem Biophys Res Commun. **342**(4):1130-1136.

Yomogida K, Yamamoto M, Yamagami T, Fujita H, Hayashi N (1993). Structure and expression of the gene encoding rat nonspecific form of delta-aminolevulinate synthase. J Biochem. **113**:364-371.

Zhang L, Vincent GM, Baralle M, Baralle FE, Anson BD, Benson DW, Whiting B, *et al*. (2004). An intronic mutation causes long QT syndrome. J Am Coll Cardiol. **44**(6):1283-91.

Zhao Z and Zhang F (2006). Sequence context analysis of 8.2 million single nucleotide polymorphisms in the human genome. Gene. **366**(2):316-24.

Zheng L, Annab LA, Afshari CA, Lee WH, Boyer TG (2001). *BRCA1* mediates ligand-independent transcriptional repression of the estrogen receptor. Proc Natl Acad Sci USA. **98**(17):9587-9592.

Zhuang Y, Katzenellenbogen BS, Shapiro DJ (1995). Estrogen receptor mutants which do not bind 17 beta-estradiol dimerize and bind to the estrogen response element in vivo. Mol Endocrinol. **9**(4):457-466.

6.2 WEBSITE REFERENCES

Alibaba2: http://www.gene-regulation.com/pub/programs/alibaba2

American Porphyria Foundation: http://www.porphyriafoundation.com

Cape Town Porphyria service: http://web.uct.ac.za/depts/porphyria

Ensembl: http://www.ensembl.org;

Human Gene Mutation Database (December 2006): http://www.hgmd.org

International HapMap project [online]: http://www.hapmap.org

Match: http://www.gene-regulation.com/pub/programs.html#match

Matinspector: http://www.genomatix.de/matinspector

National Center for Biotechnology Information (NCBI): http://www.ncbi.nlm.nih.gov

Nuclear Receptor Database: http://www.ens-lyon.fr/LBMC/laudet/nurebase/nurebase.html

Plasmocin: http://www.plasmocin.com/activity.htm

Primer 3: http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi

Signal Scan http://bimas.dcrt.nih.gov/molbio/signal

The European Porphyria Initiative (EPI): http://www.porphyria-europe.com

Transcription Regulatory Element Search (TRES): http://www.bic.nus.edu.sg.:888/tres

APPENDIX 1: REAGENTS

List of suppliers used for general reagents.

Reagent	Supplier
Agarose	Biobasic Inc.
Agar A	Bio Basic Inc.
Ethidium Bromide	Sigma Chemical Company
APS	Promega Corporation
Bis Acrylamide	Promega Corporation
Boric acid	Promega Corporation
EDTA	Sigma Chemical Company
EtOH	Merck Chemicals
Glucose	Merck Chemicals
NaCl	Laboratory Solutions
NaOH	Laboratory Solutions
SDS	BDH laboratories
TEMED	Promega Corporation
Tris-HCl	Promega Corporation
Bacto-Tryptone	Merck Chemicals
Yeast extract	Becton Dickenson

APPENDIX 2: SOLUTIONS, BUFFERS AND MIXES

Buffers and solutions were prepared with autoclaved, distilled H_2O at room temperature unless mentioned otherwise.

2.1 TBE buffer (10 x stock)

89 mM Tris-HCl (pH 8)

2 mM EDTA (pH 8)

89 mM Boric acid

Made up to 1 litre with dH_2O. pH set to 8.0. Diluted 1 in 10 for 1 x TBE buffer.

2.2 Acrylamide gel mix (40% PAA stock; 5% crosslinking)

95 g AA

5 g BAA

Made up to 100 ml with dH_2O.

2.3 LB medium (Luria-Bertani medium)

10 g Bactro-tryptone

5 g Yeast extract

10 g NaCl

Filled up to 1 L dH_2O and sterilized by autoclaving.

2.4 LB Agar plates

10 g Bacto-Tryptone

5 g Bacto-yeast extract

10 g NaCl

15 g Agar

Filled up to 1 litre with dH_2O, sterilized by autoclaving and cooled to 55 °C before necessary antibiotic was added. LB-Agar was poured into sterile petri dishes (~20 ml/plate).

2.5 PBS buffer (10 x stock)

2 g KCl

80 g NaCl

17.8 g Na_2HPO_4

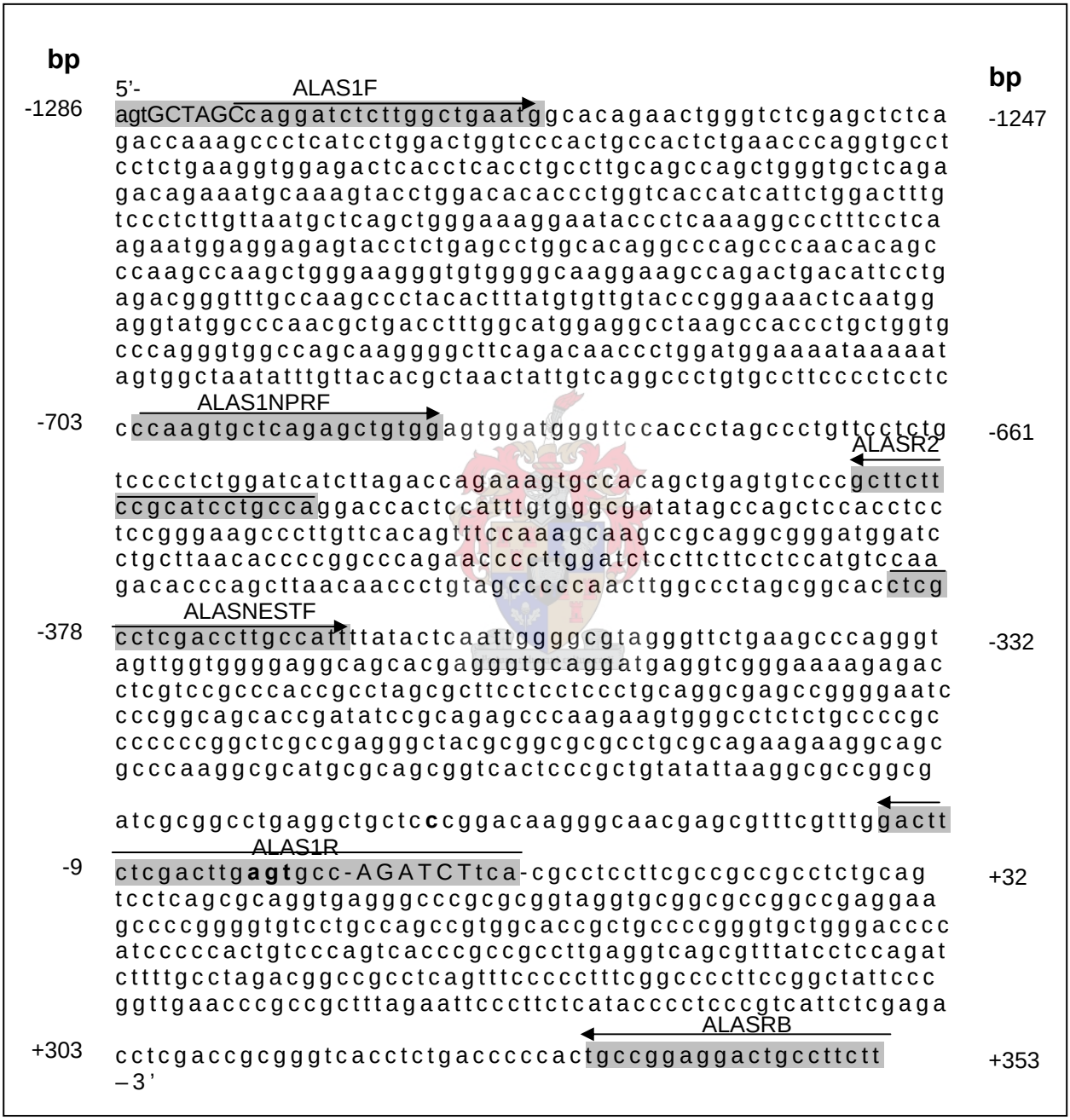
2.4 g KH_2PO_4

Filled up to 1 litre with dH_2O. Autoclaved and pH set to 7.4.

APPENDIX 3: FRAGMENT SEQUENCES

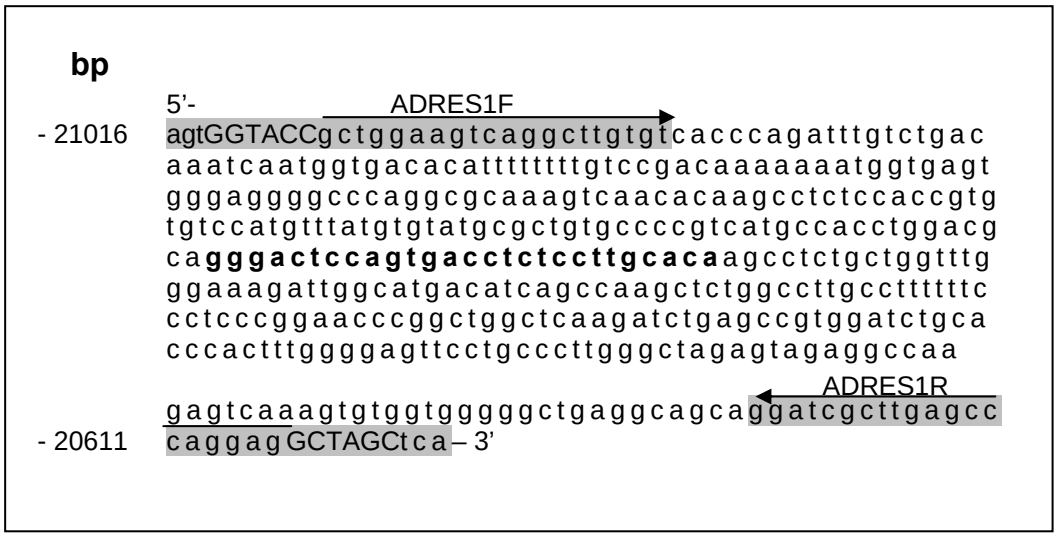
3.1 A DNA sequence of the ALAS1 5' fragment.

Shaded areas indicate primers used for PCR amplification and DNA sequencing, with the arrowheads signifying the primer orientation. Uppercase letters indicate restriction endonuclease sites.



3.2 A DNA sequence of the ADRES1 fragment.

Shaded areas indicate the primers used for PCR amplification and DNA sequencing analysis with the arrowhead signifying the primer orientation. The ADRES1 elements are indicated in bold font.

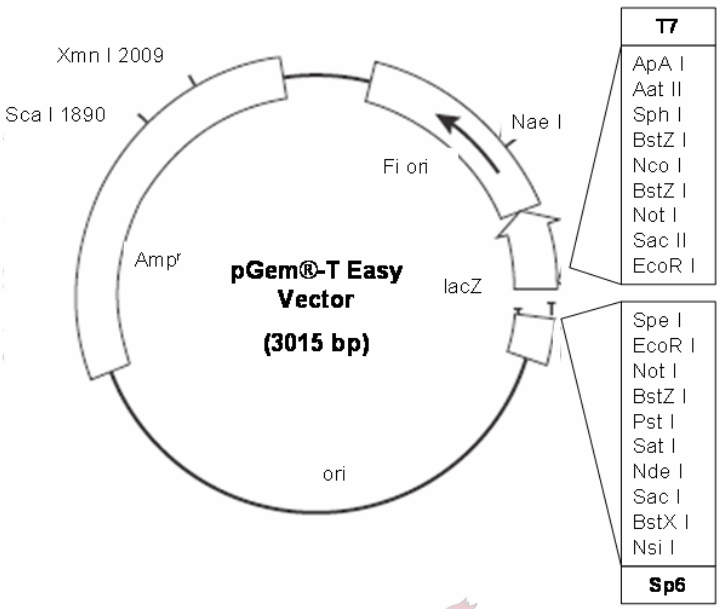


3.3 A DNA sequence of the ADRES2 fragment.

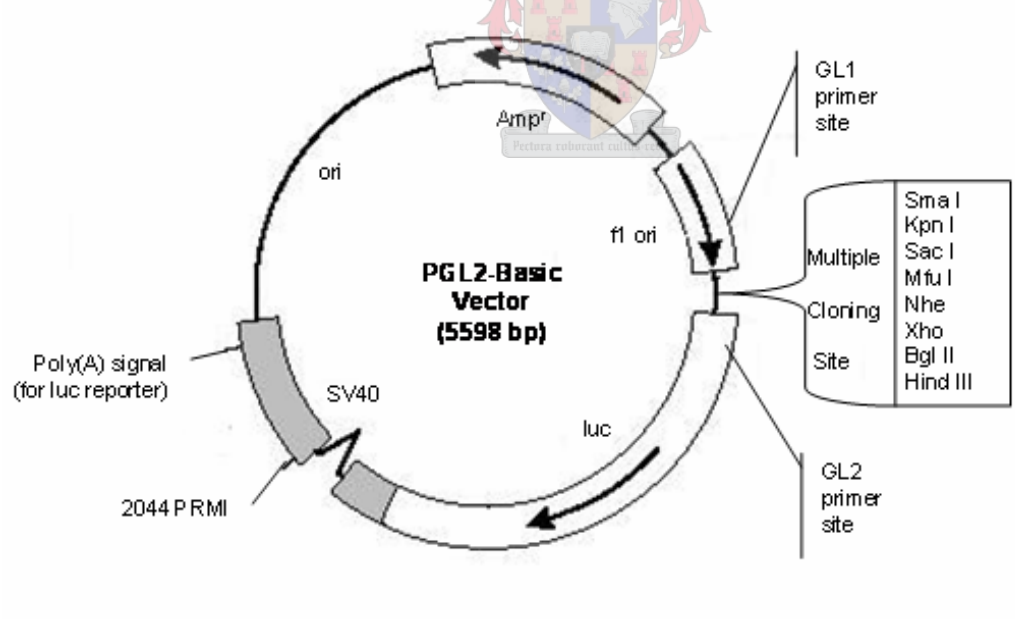
Shaded areas indicate the primers used for PCR amplification and DNA sequencing analysis with the arrowhead signifying the primer orientation. The ADRES2 elements are indicated in bold font.

APPENDIX 4: PLASMID VECTORS

4.1 Structure of the pGEMT-Easy cloning vector map and reference points.



4.2 Structure of the pGL2-Basic reporter vector map and reference points.



APPENDIX 5: TRANSCRIPTION FACTOR (TF) SUPERCLASSES AND THEIR SUBCLASSES.

Superclass	Class	Representative transcription factors	Major functions
Basic domains:	Leucine zipper (bZip)	AP-1-& C/EBP-like, CREB, bZIP/PAR	Liver differentiation; fat cell specification
	Helix-loop-helix (bHLH)	MyoD, Hairy, INO, HLH, Twist	Muscle and nerve specification; *Drosophila* sex determination
	NF-1	NF-1	Transcriptional activators; interact with TFIIB
	RF-X	RF-X	Regulator of ciliated sensory neuron differentiation
	Helix-span-helix (bHSH)	AP-2	Regulator of genes involved in the morphogenesis of the peripheral nervous system, face, limbs and skin
Zinc-coordinating domains:	Nuclear Receptors (NR)	Steroid-, Thyroid receptors	Secondary sex determination; reproduction; craniofacial & limb development;metabolism of steroids & xenobiotics
	Cys2His2	Krüppel, GL1-like	Cell-cycle & developmental regulators
	Diverse	GATA, Trithorax	Maintenance of hematopoietic cells
Helix-turn-helix domains:	Paired Homeodomain	LIM, Pax	Head development; neural specification; eye development
	POU	Pit-1, Oct-1, Oct-2	Pituitary development; neural fate
	Forkhead/ Winged	GABP	Developmental regulators; cell-cycle controlling factors
	Heatshock	HSF	Transcriptional response to heat stress
	Standard	Tryptophan	Repressor protein
	TEA domain	TEF-1	Transcriptional activators
Beta-scaffold domains with minor groove contacts:	RHR	Rel/ankyrin	Transcription regulators for genes in cellular defence and differentiation mechanisms
	HMG	Sox, TCF-1	Mammalian primary sex determination; ectoderm differentiation
	STAT	Stat	Signal transducers and activators of transcription
	P53	P53	DNA repair pathway and apoptosis transcription factor
	MADS	MEF-2, Homeotic	Regulators of differentiation; responders to external signals
	TATA-binding	TBP	Eukaryotic TATA-binding
	CCAAT factors		
	Cold-Shock	csd	Transcriptional response to lower temperatures
	Grainyhead	Grainyhead	Development

APPENDIX 6: UNPROCESSED NORMALISED DATA OF LUCIFERASE- AND β-GALACTOSIDASE ASSAYS

Shaded boxes indicate the construct and/or conditions used as reference. Asterisks denote values not within 20% of each other, with a subsequent repeat of the set.

6.1 Normalised values obtained in set 1.

Set	Number	Type	Repeat	WT (Reference)	Normalised Value
Set1	nr1	Wild type	1	2022.215697	2022.215697
Set1	nr1	Wild type	2	2108.567781	2108.567781
Set1	nr1	Wild type	3	1952.564999	1952.564999
Set1	nr1	ERE	1	2022.215697	3080.16429
Set1	nr1	ERE	2	2108.567781	3258.23215
Set1	nr1	ERE	3	1952.564999	2006.311049
Set1	nr1	ERE/Sp1	1	2022.215697	2759.755569
Set1	nr1	ERE/Sp2	2	2108.567781	2562.880134
Set1	nr1	ERE/Sp3	3	1952.564999	1899.155477
Set1	nr1	VitERE	1	2022.215697	169.5547324
Set1	nr1	VitERE	2	2108.567781	235.530711
Set1	nr1	VitERE	3	1952.564999	175.7790452
Set1	nr1	pGL2	1	2022.215697	2.541827178
Set1	nr1	pGL2	2	2108.567781	2.894229378
Set1	nr1	pGL2	3	1952.564999	317.5238079
Set1	nr2	Wild type	1	707.6997582	707.6997582
Set1	nr2	Wild type	2	337.6230873	337.6230873
Set1	nr2	Wild type	3	891.607708	891.607708
Set1	nr2	ERE	1	707.6997582	1763.2184 *
Set1	nr2	ERE	2	337.6230873	1009.589654 *
Set1	nr2	ERE	3	891.607708	309.0821213 *
Set1	nr2	ERE/Sp1	1	707.6997582	443.1140947
Set1	nr2	ERE/Sp2	2	337.6230873	1429.464225
Set1	nr2	ERE/Sp3	3	891.607708	888.1261639
Set1	nr2	VitERE	1	707.6997582	1.856731263
Set1	nr2	VitERE	2	337.6230873	8.536680949
Set1	nr2	VitERE	3	891.607708	11.40239349
Set1	nr2	pGL2	1	707.6997582	0.403234633
Set1	nr2	pGL2	2	337.6230873	0.531176521
Set1	nr2	pGL2	3	891.607708	0.65154192
Set1	nr3	Wild type	1	1682.849027	1682.849027
Set1	nr3	Wild type	2	2030.898331	2030.898331
Set1	nr3	Wild type	3	2350.250592	2350.250592
Set1	nr3	ERE	1	1682.849027	2162.61996
Set1	nr3	ERE	2	2030.898331	2228.231645
Set1	nr3	ERE	3	2350.250592	2100.285586
Set1	nr3	ERE/Sp1	1	1682.849027	1881.049917
Set1	nr3	ERE/Sp2	2	2030.898331	1591.430702
Set1	nr3	ERE/Sp3	3	2350.250592	1994.783597
Set1	nr3	VitERE	1	1682.849027	169.5547324
Set1	nr3	VitERE	2	2030.898331	235.530711
Set1	nr3	VitERE	3	2350.250592	175.7790452
Set1	nr3	pGL2	1	1682.849027	0.76451415

Set1	nr3	pGL2	2	2030.898331	0.932991381
Set1	nr3	pGL2	3	2350.250592	0.955679655
Set1	nr4	Wild type	1	341.4587878	341.4587878
Set1	nr4	Wild type	2	294.0639323	294.0639323
Set1	nr4	Wild type	3	440.3539735	440.3539735
Set1	nr4	ERE	1	341.4587878	621.5027407
Set1	nr4	ERE	2	294.0639323	669.9157479
Set1	nr4	ERE	3	440.3539735	677.1995328
Set1	nr4	ERE/Sp1	1	341.4587878	661.4857724
Set1	nr4	ERE/Sp2	2	294.0639323	791.7200115
Set1	nr4	ERE/Sp3	3	440.3539735	614.1882724
Set1	nr4	pGL2	1	341.4587878	1.249363664
Set1	nr4	pGL2	2	294.0639323	0.367809811
Set1	nr4	pGL2	3	440.3539735	0.507445426

6.2 Normalised values obtained in set 2.

Set	Number	Type	Repeat	WT (Reference)	Normalised Value
Set2	nr1	Wild type	1	1627.011115	1627.011115
Set2	nr1	Wild type	2	1432.282158	1432.282158
Set2	nr1	Wild type	3	1144.331431	1144.331431
Set2	nr1	ERE	1	1627.011115	1109.215475
Set2	nr1	ERE	2	1432.282158	1049.746643
Set2	nr1	ERE	3	1144.331431	1131.434275
Set2	nr1	ERE/Sp1	1	1627.011115	1048.518863
Set2	nr1	ERE/Sp2	2	1432.282158	1391.538647
Set2	nr1	ERE/Sp3	3	1144.331431	818.9710809
Set2	nr1	VitERE	1	1627.011115	46.13645138
Set2	nr1	VitERE	2	1432.282158	19.03532503
Set2	nr1	VitERE	3	1144.331431	31.15879264
Set2	nr1	pGL2	1	1627.011115	1.237867634
Set2	nr1	pGL2	2	1432.282158	1.015504211
Set2	nr1	pGL2	3	1144.331431	1.998557563
Set2	nr2	Wild type	1	764.131856	764.131856
Set2	nr2	Wild type	2	1198.233708	1198.233708
Set2	nr2	Wild type	3	1207.715769	1207.715769
Set2	nr2	ERE	1	764.131856	960.2010665
Set2	nr2	ERE	2	1198.233708	1373.973091
Set2	nr2	ERE	3	1207.715769	961.833661
Set2	nr2	ERE/Sp1	1	764.131856	1076.382647
Set2	nr2	ERE/Sp2	2	1198.233708	909.9246959
Set2	nr2	ERE/Sp3	3	1207.715769	908.5964454
Set2	nr2	VitERE	1	764.131856	20.77312029
Set2	nr2	VitERE	2	1198.233708	19.01864916
Set2	nr2	VitERE	3	1207.715769	14.8282244
Set2	nr2	pGL2	1	764.131856	1.219990549
Set2	nr2	pGL2	2	1198.233708	0.563898359
Set2	nr2	pGL2	3	1207.715769	0.548865495
Set2	nr3	Wild type	1	42.92999707	42.92999707
Set2	nr3	Wild type	2	59.13456764	59.13456764
Set2	nr3	Wild type	3	57.87336651	57.87336651
Set2	nr3	ERE	1	42.92999707	43.27995472
Set2	nr3	ERE	2	59.13456764	55.5174884
Set2	nr3	ERE	3	57.87336651	54.17905172
Set2	nr3	ERE/Sp1	1	42.92999707	22.90550568

Set2	nr3	ERE/Sp2	2	59.13456764	20.36876666
Set2	nr3	ERE/Sp3	3	57.87336651	32.57888152
Set2	nr3	VitERE	1	42.92999707	2.020115974
Set2	nr3	VitERE	2	59.13456764	0.988167765
Set2	nr3	VitERE	3	57.87336651	0.870805569
Set2	nr3	pGL2	1	42.92999707	0.285288391
Set2	nr3	pGL2	2	59.13456764	0.46823393
Set2	nr3	pGL2	3	57.87336651	0.354275754

6.3 Normalised values obtain in set 3.

Set	Number	Type	Repeat	WT (Reference)	Normalised Value
Set3	nr1	VitERE (estrogen)	1	11794.96142	327.673744
Set3	nr1	VitERE (estrogen)	2	10767.34862	143.5044281
Set3	nr1	VitERE (estrogen)	3	7586.599997	374.6492847
Set3	nr1	Wild type (estrogen)	1	11794.96142	11794.96142
Set3	nr1	Wild type (estrogen)	2	10767.34862	10767.34862
Set3	nr1	Wild type (estrogen)	3	7586.599997	7586.599997
Set3	nr1	pGL2 (estrogen)	1	11794.96142	6.266249516
Set3	nr1	pGL2 (estrogen)	2	10767.34862	2.832526004
Set3	nr1	pGL2 (estrogen)	3	7586.599997	6.270869206
Set3	nr1	VitERE + ER (estrogen)	1	11794.96142	8864.742995
Set3	nr1	VitERE + ER (estrogen)	2	10767.34862	9457.73925
Set3	nr1	VitERE + ER (estrogen)	3	7586.599997	11105.60493
Set3	nr1	Wild type +ER (estrogen)	1	11794.96142	19373.85689
Set3	nr1	Wild type +ER (estrogen)	2	10767.34862	16964.49071
Set3	nr1	Wild type +ER (estrogen)	3	7586.599997	21521.15986
Set3	nr1	Wild type +ER (etoh)	1	11794.96142	17155.15968
Set3	nr1	Wild type +ER (etoh)	2	10767.34862	12392.069
Set3	nr1	Wild type +ER (etoh)	3	7586.599997	17249.17815

5.4 Normalised values obtained in set 4.

Set	Number	Type	Repeat	WT (Reference)	Normalised Value
Set4	nr1	Wild type+ER (estr10^{-6})	1	5161.413535	5651.064706
Set4	nr1	Wild type+ER (estr10^{-6})	2	4481.80454	6221.178843
Set4	nr1	Wild type+ER (estr10^{-6})	3	4425.093057	6300.498858
Set4	nr1	Wild type+ER (estr10^{-8})	1	5161.413535	5684.916364
Set4	nr1	Wild type+ER (estr10^{-8})	2	4481.80454	5664.920498
Set4	nr1	Wild type+ER (estr10^{-8})	3	4425.093057	5798.543525
Set4	nr1	Wild type+ER (estr10^{-10})	1	5161.413535	4683.556736
Set4	nr1	Wild type+ER (estr10^{-10})	2	4481.80454	4808.454008
Set4	nr1	Wild type+ER (estr10^{-10})	3	4425.093057	5536.534901
Set4	nr1	Wild type+ER (estr10^{-12})	1	5161.413535	3720.614974
Set4	nr1	Wild type+ER (estr10^{-12})	2	4481.80454	4564.881163
Set4	nr1	Wild type+ER (estr10^{-12})	3	4425.093057	4518.679137
Set4	nr1	Wild type+ER (etoh)	1	5161.413535	5161.413535
Set4	nr1	Wild type+ER (etoh)	2	4481.80454	4481.80454
Set4	nr1	Wild type+ER (etoh)	3	4425.093057	4425.093057
Set4	nr1	Wild type (etoh)	1	5161.413535	2770.851978
Set4	nr1	Wild type (etoh)	2	4481.80454	3382.74223
Set4	nr1	Wild type (etoh)	3	4425.093057	1672.20383

6.5 Normalised values obtained in set 5.

Set	Number	Type	Repeat	WT (Reference)	Normalised Value
Set5	nr1	Wild type (estrogen)	1	126.4855832	130.1807423
Set5	nr1	Wild type (estrogen)	2	77.76340903	127.8856729
Set5	nr1	Wild type (estrogen)	3	28.39571888	212.999699
Set5	nr1	ERE (estrogen)	1	126.4855832	219.3416251
Set5	nr1	ERE (estrogen)	2	77.76340903	148.307011
Set5	nr1	ERE (estrogen)	3	28.39571888	146.9795548
Set5	nr1	ERE/Sp1 (estrogen)	1	126.4855832	75.31537159
Set5	nr1	ERE/Sp1 (estrogen)	2	77.76340903	112.5614412
Set5	nr1	ERE/Sp1 (estrogen)	3	28.39571888	87.07106626
Set5	nr1	pGL2 (estrogen)	1	126.4855832	0.110985122
Set5	nr1	pGL2 (estrogen)	2	77.76340903	0.079253821
Set5	nr1	pGL2 (estrogen)	3	28.39571888	0.120473488
Set5	nr1	Wild type (etoh)	1	126.4855832	126.4855832
Set5	nr1	Wild type (etoh)	2	77.76340903	77.76340903
Set5	nr1	Wild type (etoh)	3	28.39571888	28.39571888
Set5	nr1	ERE (etoh)	1	126.4855832	38.27420223 *
Set5	nr1	ERE (etoh)	2	77.76340903	119.4357249 *
Set5	nr1	ERE (etoh)	3	28.39571888	79.99246865 *
Set5	nr1	ERE/Sp1 (etoh)	1	126.4855832	124.4396248 *
Set5	nr1	ERE/Sp1 (etoh)	2	77.76340903	70.57922889 *
Set5	nr1	ERE/Sp1 (etoh)	3	28.39571888	45.33953939 *
Set5	nr2	Wild type (estrogen)	1	847.800357	1171.526271
Set5	nr2	Wild type (estrogen)	2	769.2724188	1069.78121
Set5	nr2	Wild type (estrogen)	3	842.8727139	1279.390565
Set5	nr2	ERE (estrogen)	1	847.800357	950.9024867
Set5	nr2	ERE (estrogen)	2	769.2724188	1161.785259
Set5	nr2	ERE (estrogen)	3	842.8727139	1430.960427
Set5	nr2	ERE/Sp1 (estrogen)	1	847.800357	905.305257
Set5	nr2	ERE/Sp1 (estrogen)	2	769.2724188	1012.669589
Set5	nr2	ERE/Sp1 (estrogen)	3	842.8727139	809.5638508
Set5	nr2	pGL2 (estrogen)	1	847.800357	1.714743409
Set5	nr2	pGL2 (estrogen)	2	769.2724188	1.928298563
Set5	nr2	pGL2 (estrogen)	3	842.8727139	1.696703081
Set5	nr2	Wild type (etoh)	1	847.800357	847.800357
Set5	nr2	Wild type (etoh)	2	769.2724188	769.2724188
Set5	nr2	Wild type (etoh)	3	842.8727139	842.8727139
Set5	nr2	ERE (etoh)	1	847.800357	809.8537081
Set5	nr2	ERE (etoh)	2	769.2724188	653.533928
Set5	nr2	ERE (etoh)	3	842.8727139	1055.201849
Set5	nr2	ERE/Sp1 (etoh)	1	847.800357	493.5375362
Set5	nr2	ERE/Sp1 (etoh)	2	769.2724188	631.6290243
Set5	nr2	ERE/Sp1 (etoh)	3	842.8727139	586.88377
Set5	nr3	Wild type (estrogen)	1	283.1234336	1018.242413
Set5	nr3	Wild type (estrogen)	2	273.8712866	428.2672745
Set5	nr3	Wild type (estrogen)	3	301.6722015	395.7907792
Set5	nr3	ERE (estrogen)	1	283.1234336	362.526244
Set5	nr3	ERE (estrogen)	2	273.8712866	486.3277286
Set5	nr3	ERE (estrogen)	3	301.6722015	505.4971129
Set5	nr3	ERE/Sp1 (estrogen)	1	283.1234336	378.4159181
Set5	nr3	ERE/Sp1 (estrogen)	2	273.8712866	360.1149354
Set5	nr3	ERE/Sp1 (estrogen)	3	301.6722015	429.6810032

Set5	nr3	pGL2 (estrogen)	1	283.1234336	0.467225505
Set5	nr3	pGL2 (estrogen)	2	273.8712866	0.862316758
Set5	nr3	pGL2 (estrogen)	3	301.6722015	0.602870787
Set5	nr3	Wild type (etoh)	1	283.1234336	283.1234336
Set5	nr3	Wild type (etoh)	2	273.8712866	273.8712866
Set5	nr3	Wild type (etoh)	3	301.6722015	301.6722015
Set5	nr3	ERE (etoh)	1	283.1234336	300.5897953
Set5	nr3	ERE (etoh)	2	273.8712866	235.6381627
Set5	nr3	ERE (etoh)	3	301.6722015	338.1017325
Set5	nr3	ERE/Sp1 (etoh)	1	283.1234336	215.3034241
Set5	nr3	ERE/Sp1 (etoh)	2	273.8712866	236.3782257
Set5	nr3	ERE/Sp1 (etoh)	3	301.6722015	145.307355
Set5	nr4	Wild type (estrogen)	1	263.7020325	473.524204
Set5	nr4	Wild type (estrogen)	2	278.3149996	371.1112522
Set5	nr4	Wild type (estrogen)	3	301.6722015	459.9691719
Set5	nr4	ERE (estrogen)	1	263.7020325	521.5255238
Set5	nr4	ERE (estrogen)	2	278.3149996	582.9602245
Set5	nr4	ERE (estrogen)	3	301.6722015	482.4186314
Set5	nr4	ERE/Sp1 (estrogen)	1	263.7020325	663.7587101
Set5	nr4	ERE/Sp1 (estrogen)	2	278.3149996	501.7167158
Set5	nr4	ERE/Sp1 (estrogen)	3	301.6722015	412.4540528
Set5	nr4	pGL2 (estrogen)	1	263.7020325	0.467225505
Set5	nr4	pGL2 (estrogen)	2	278.3149996	0.862316758
Set5	nr4	pGL2 (estrogen)	3	301.6722015	0.602870787
Set5	nr4	Wild type (etoh)	1	263.7020325	263.7020325
Set5	nr4	Wild type (etoh)	2	278.3149996	278.3149996
Set5	nr4	Wild type (etoh)	3	301.6722015	301.6722015
Set5	nr4	ERE (etoh)	1	263.7020325	245.5498079
Set5	nr4	ERE (etoh)	2	278.3149996	424.5394573
Set5	nr4	ERE (etoh)	3	301.6722015	284.234301
Set5	nr4	ERE/Sp1 (etoh)	1	263.7020325	295.4988685
Set5	nr4	ERE/Sp1 (etoh)	2	278.3149996	236.3033364
Set5	nr4	ERE/Sp1 (etoh)	3	301.6722015	324.1125948