Molecular characterisation of the gene, *LGALS13*, and its putative involvement in pre-eclampsia

Alisa Postma

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

Supervisor: Dr R. Hillermann-Rebello

Co-supervisor: Prof L. Warnich

March 2009

DECLARATION

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

Date: _____

ABSTRACT

Pre-eclampsia is one of the most common hypertensive disorders of pregnancy in South Africa. Presently, the only cure for pre-eclampsia is delivery, which brings with it, additional complications. As an alternative, clinical management of this disorder relies on timely diagnosis.

The predictive biomarker, Placental Protein 13 (PP13), is currently used for the early diagnosis of pre-eclampsia, in an ELISA-based diagnostic kit, developed by Diagnostic Technologies Limited (DTL)[1]. A decrease in serum PP13 levels has been reported during the first trimester of pregnancy in women who later develop pre-eclampsia. The function of PP13 has not been fully elucidated and it is also not known whether the reduction in PP13 levels is a cause or an effect of the disease. The use of PP13 as a predictive biomarker for pre-eclampsia therefore warrants a comprehensive study of this peptide and the encoding gene, *LGALS13*.

The aim of this study was firstly to characterise *LGALS13* using a range of *in silico* tools. PP13 was found to be most homologous to the predicted protein product of a neighbouring "putative" gene, LOC148003. A gene conversion event between these two genes most likely underlies the so-called "hotspot mutation" in *LGALS13*. Data also demonstrates that the DelT mutation disrupts functionally and structurally important features of the gene and peptide sequences. Through the analysis of the putative promoter region of *LGALS13*, the presence of a Stimulatory protein-1 (Sp1) binding sequence element was predicted, which has implications for regulation of *LGALS13*.

Secondly, the study aimed to establish a study cohort for the investigation of the effect that the *LGALS13* genotype has on the expression of its mRNA and protein products. Serum, plasma and whole

[1] An Israel-based medical diagnostic and biotechnology company. Refer to p 18 for more information.

blood samples were collected and prepared from 316 pregnant women. Placental tissue samples were obtained from a selected group of these subjects for RNA extraction. Once the sampling on the two remaining targeted deliveries has occurred, the collection of samples will be batched and sent to DTL in Israel, for PP13 measurement.

DNA was extracted from the whole blood samples obtained, and all study participants were genotyped for seven sequence variants within the *LGALS13* gene using (i) Multiphor Single Stranded Conformational Polymorphism and Heteroduplex (SSCP/HD) analysis, (ii) restriction enzyme analysis and (iii) DNA sequencing. The genotype data sets will be compared with PP13 levels when they become available, and also with clinical parameters, once the deliveries have all occurred and the database is complete.

This study demonstrated the power of an *in silico* approach to direct the focus of future experimental work. The newly established study cohort will be used for prospective studies aiming at a better understanding of the role which *LGALS13* and PP13 play in the early prediction of pre-eclampsia.

OPSOMMING

Pre-eklampsie is een van die mees algemene hipertensie-verwante swangerskapsiektes in Suid-Afrika. Verlossing is tans die enigste wyse waarop pre-eklampsie genees kan word, wat weer addisionele komplikasies teweegbring. Alternatiewelik berus die kliniese bestuur van hierdie siekte op die vroegtydige diagnose daarvan.

Die voorspellende biomerker, Plasentaproteïen 13 (PP13), word tans gebruik in die vroeë diagnose van pre-eklampsie deur middel van 'n ELISA-gebaseerde diagnostiese stelsel wat ontwikkel is deur *Diagnostic Technologies Limited (DTL)* [2]. Daar is bevind dat serum PP13-vlakke, gemeet tydens die eerste trimester van swangerskap, laer is in vroue wat later pre-eklampsie ontwikkel. Die funksie van PP13 is nog nie vasgestel nie en dit is ook nog nie bekend of die afname in PP13-vlakke 'n oorsaak of 'n gevolg van die siekte is nie. Die gebruik van PP13 as 'n voorspellende biomerker vir die vroeë diagnose van pre-eklampsie regverdig dus 'n omvattende studie van hierdie proteïen en die enkoderende geen, *LGALS13*.

Die doel van hierdie studie was eerstens om *LGALS13* te karakteriseer deur van 'n reeks *in silico*-metodes gebruik te maak. Daar is bevind dat PP13 die meeste homologie toon met die voorspelde proteïenproduk van 'n nabygeleë vermeende geen, *LOC148003*. 'n Geen-omskakelingsvoorval tussen hierdie twee gene is die mees waarskynlike oorsaak van die sogenaamde "warm kol"-mutasie in *LGALS13*. Data dui ook aan dat die DelT-mutasie funksioneel- en struktureel- belangrike eienskappe van die geen en peptiedvolgordes ontwrig. Analise van die vermeende promoterarea van *LGALS13* het die teenwoordigheid van 'n Stimulatoriese proteïen-1 (Sp1) bindings DNS-element voorspel, wat implikasies vir die regulering van *LGALS13* inhou.

[2] 'n Mediese diagnostiese en biotegnologie maatskappy gebasseer in Israel. Meer inligting is beskikbaar op bl 18.

Tweedens het die studie gepoog om ’n studiekohort daar te stel wat gebruik kan word om die effek van die *LGALS13*-genotipe op die ekspressie van die mRNS en proteïenprodukte te ondersoek. Serum, plasma en heelbloedmonsters is van 316 swanger vroue versamel. Plasentaweefsel is verkry van ’n geselekteerde groep uit hierdie kohort, vir latere RNS-ekstraksies. Sodra monsters van die oorblywende twee geteikende bevallings verkry is, sal alle versamelde materiaal na DTL in Israel gestuur word vir PP13 kwantifisering.

DNS is uit bogenoemde volbloedmonsters geëkstraheer en alle studiekohortlede is gegenotipeer vir sewe *LGALS13* variante deur middel van (i) *Multiphor* Enkelstring Konformasie Polimorfisme en Heterodupleks *(SSCP/HD)* analise asook (ii) restriksie ensiem analise en (iii) DNS-volgordebepaling. Die genotipe-datastelle sal vergelyk word met die PP13-vlakke asook kliniese inligting van die studiekohort (sodra dit beskikbaar is).

Hierdie studie toon die vermoë van ’n *in silico*-benadering om die fokus van toekomstige eksperimentele werk te rig. Die nuut-gevestigde studiekohort sal gebruik word vir toekomstige studies wat ’n beter begrip van die rol van *LGALS13* en PP13 in die vroeë diagnose van pre-eklampsie ten doel het.

ACKNOWLEDGEMENTS

1. Dr Renate Hillermann, for supporting me academically and emotionally, as well as financially enabling me to undertake this study, and affording me the opportunity to travel to Israel
2. Erika van Papendorp, clinical research nurse, and Dr Stefan Gebhardt, gynaecologist, for friendly, helpful assistance above and beyond their call of duty
3. Prof Warnich, for co-supervision of this research and her unstinting help in finishing this thesis
4. Friends and colleagues from the university: Petra, Marika, Marguerite, Veronique, Natalie and Jomien for keeping me motivated and smiling
5. Dr Mauritz Venter for sharing his knowledge and technical expertise, as well as Dr George Rebello, for facilitating my academic progress and for his assistance in Israel
6. Staff at TC Newman clinic and Paarl Provincial hospital
7. My parents, for their unconditional support and love
8. Friends and family, for their loyalty and support
9. The NRF for funding
10. DTL for sharing their business and technical skills, and particularly the CEO of the company, Dr Hamutal Meiri for her positive contribution to my scientific growth

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: LITERATURE REVIEW

CHAPTER 3: RESULTS AND DISCUSSION

LIST OF TABLES

CHAPTER 2: MATERIALS AND METHODS

CHAPTER 3: RESULTS AND DISCUSSION

LIST OF ABBREVIATIONS

A	Adenine
AFP	Alpha feto-protein
APS	Ammonium persulphate
Arg	Argenine
Asp	Aspartic acid
ATG	Start codon
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
C	Cytosine
CARES	*Cis*-acting Regulatory Elements
CRD	Carbohydrate recognition domain
Cys	Cysteine
dbSNP: rs	database single nucleotide polymorphism: reference sequence
ddH_2O	double distilled water
dH_2O	distilled water
delT	deletion of a single thymine base
DNA	Deoxyribonucleic acid
dNTP	2'-deoxynucleotide-5'-triphosphate
E	Exon
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ESE	exonic splice elements
EtOH	Ethanol
F	Forward primer
G	Guanine
g	Gram
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
HCG	Human Chorionic Gonadotropin
HDP	Heteroduplex
HELLP	Hemolysis, elevated liver enzymes and low platelets
HIF	Hypoxia Inducible factor
HLA	Human leukocyte antigen
HWE	Hardy-Weinberg equilibrium
IDT	Integrated DNA Technologies
Ile	Isoleucine
IUD	Intrauterine death
IUGR	Intrauterine growth restriction
IVS	intervening sequence
*LGALS*13	galactose-binding, soluble 13
M	Molar
Mg	Milligram
Min	Minute
mL	Millilitre
mM	Millimolar
MMPs	Matrix metalloproteinases
MRC	Medical Research Council
mRNA	messenger ribonucleic acid
NaCl	Sodium Chloride
Na_2HPO_4	disodium hydrogen phosphate
NaOH	Sodium hydroxide
NCBI	National Centre for Biotechnology Information
ng	Nanogram
NH_4Cl	Ammonium chloride
NK-κB	Nuclear factor κ B
ORF	Open reading frame
PA	Plasminogen Activators
PAGE	Polyacrylamide gel electrophoresis
PAPP-A	Pregnancy-associated plasma protein A
PBS	PBS: Phosphate buffered saline
PCR	Polymerase chain reaction
PDA	Piperazine diacrylamide
PECAM	Platelet/endothelial cellular-adhesion molecule
pg	Pictogram
pH	potential of Hydrogen
PI	pulsality index
PIH	Pregnancy-induced hypertension
PP13	Placental protein 13
PPIP	Perinatal Problem Identification Programme
PPROM	Preterm premature rupture of membranes

PROM	Premature rupture of membranes
PTL	Preterm labour
R	Reverse primer
RE	Restriction enzyme
REA	Restriction enzyme analysis
rpm	revolutions per minute
s	Second
SBP	Systolic blood pressure
SDS	Sodium dodecyl sulphate
Ser	Serine
SF2	Splicing factor 2
SNP	Single nucleotide polymorphism
SRp40	Splicing factor, arginine/serine-rich 5
SRp55	Splicing factor, arginine/serine-rich-6
SSCP/HD	Single-stranded conformational polymorphism and heterduplex
T	Thymine
Ta	annealing temperature
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
TEMED	N,N,N',N'-tetramethylethlenediamine
TGA	Stop codon
TGF	Trophoblast Growth Factor
T_{H2}	T-helper cells
Thr	Threonine
Tm	Melting temperature
TNFα	Tumor-necrosis factor α protein
TRIS	Trishydroxymethylaminomethane
Tyr	Tyrosine
U	Unit
uPA	urokinase Plasminogen Activators
V	Volts
VCAM	Vascular cell-adhesion molecule
VEGF	Vascular endothelial growth factor
WHO	World health organization
Wks	Weeks
WT	Wildtype
yrs	Years

CHAPTER 1

LITERATURE REVIEW

1 Pre-eclampsia

1.1 Frequency and symptoms

In South Africa, complications due to hypertension remain the greatest direct cause of maternal death during pregnancy, being responsible for 19.1% of all maternal deaths between 2002 and 2004 (National Committee on Confidential Enquiries into Maternal Deaths in the office of the Minister of Health, 2004). Between 2003 and 2005, hypertensive disorders were also classified as one of the primary causes of stillbirth, perinatal and neonatal death in South Africa (MRC Research Unit for Maternal and Infant Health Care Strategies, PPIP Users and the Saving Babies Technical Task Team, 2006). Hypertensive disorders of pregnancy (HDP) include pre-eclampsia, eclampsia, chronic hypertension, HELLP syndrome and liver rupture, of which pre-eclampsia and eclampsia have the highest incidence. An increase in the number of maternal deaths caused by HDP has been observed in South Africa between 1999 and 2004. Pre-eclampsia is diagnosed when a pregnant woman's blood pressure exceeds 110/90 mmHg at a gestational age greater than 20 weeks, with the presence of proteinuria (Davey and MacGillivray, 1988). A pregnant woman's blood pressure is defined as raised when it measures at 140/90 mmHg or higher, and proteinuria is diagnosed when a protein concentration of 30 mg/mmol is measured in the urine (Duley, 2003). Although some women suffering from pre-eclampsia present with symptoms such as convulsions, headaches, drowsiness and abdominal pain, pre-eclampsia patients are usually asymptomatic (Hayman and Myers, 2004).

Pre-eclampsia is one of the main causes of maternal and fetal mortality and morbidity worldwide, accounting for approximately 8% of maternal deaths worldwide each year (Papageorghiou and Campbell, 2006). The incidence of pre-eclampsia varies on a geographical scale, but has been reported to be responsible for 10-15% of maternal deaths in developing parts of the world such as Africa, Latin-America and the Caribbean

(Duley *et al*. 2007). Of these developing countries, Africa has the highest total mortality associated with hypertensive disorders, which are the most difficult of all pregnancy-related disorders to prevent in both developed and developing countries (Duley *et al*. 2007).

Pre-eclampsia often necessitates preterm delivery, to prevent maternal damage, which can hold devastating consequences for the fetus, such as prematurity or perinatal death (Papageorghiou and Campbell, 2006). Gestational hypertension accounts for one in six stillbirths and one in six abrupt infant deaths (Myers and Brockelsby, 2004). The main causes of neonatal mortality and morbidity are low birth weight (infants born with a weight below 2500g) and immature organ development. Iatrogenic preterm delivery is often responsible for complications such as motor and sensory malfunction, deafness and blindness (Burger *et al*. 2004). The rate of infants born with a low birth weight has increased by 16% since 1990 in the U.S.A. and was estimated to be 8% in 2004. The estimated preterm birth rate in the U.S.A. was 12.5% in 2004 and has increased with 18% since 1990 (Martin *et al*. 2006).

Between 2003 and 2005, the incidence of low birth weight among infants was reported to be 15% in South Africa. Infants weighing more than 500g, had a perinatal mortality rate of 37.5/1000 births and an early neonatal death rate of 13.6/1000 births. Both the perinatal and neonatal mortality rates have remained unchanged over the last six years in South Africa (MRC Research Unit for Maternal and Infant Health Care Strategies, PPIP Users and the Saving Babies Technical Task Team, 2006).

1.2 Placentation and pathogenesis during pre-eclampsia

Pre-eclampsia is diagnosed by an increase in blood pressure and proteinuria. However, it is a very complex, clinically heterogenic, multi-organ syndrome which leads to a cascade of physiological and metabolic changes in the endothelial lining of the maternal blood vessels and

placenta. Although the symptoms of pre-eclampsia only manifest during the second or third trimester, it is thought that the underlying cause of the disorder occurs during placentation, thus in the early stages of pregnancy (Norwitz, 2006). Pre-eclampsia has been previously described by a two-stage model, where a reduction in placental perfusion and abnormal placentation occurs during the first stage which gives rise to the development of the maternal condition during the second stage of the syndrome (Redman, 1991). Before studying the underlying pathogenesis associated with pre-eclampsia, it is necessary to have a basic understanding of the molecular and cellular processes involved in normal placentation.

1.3 Normal versus defective placentation

During the first trimester of pregnancy, cytotrophoblast stem cells within the placenta can either fuse to form the syncytiotrophoblast cell layer, or aggregate, resulting in columns of anchoring villous tissue (Kenny, 2004). These columns of villous tissue establish a physical connection between the uterine wall and the placenta. During early pregnancy, extravillous trophoblast cells invade the uterine wall (interstitial invasion) as well as the maternal myometrium and decidua (endovascular invasion). This leads to transformation of the maternal spiral arteries, involving invasion of the blood vessels and replacement of endothelial and most of the muscoloelastic tissue in the walls of these vessels (Kenny, 2004). This is necessary since vascular cells are more capable of providing the fetoplacental membranes with adequate blood flow necessary for efficient nutrient, oxygen and waste transport (Zhou *et al*. 1997). The small spiral arteries consequently develop into large sinusoidal vessels, creating an intervillous space which transforms the placental vascular system for the accommodation of high blood flow and low blood pressure (Kenny, 2004). It has been shown that these modifications in the maternal vasculature are absent during pre-eclamptic pregnancies, which result in a reduction in placental perfusion. Trophoblast invasion during placentation is notably deeper in humans

than in any of the other primate species studied thus far (Pijenborg *et al*. 2007). It has been shown that, even though there is a sufficient amount of interstitially migrating trophoblast cells present in the placental bed, insufficient trophoblast invasion occurs during pregnancies complicated by pre-eclampsia. Insufficient or shallow trophoblastic and interstitial trophoblast invasion of the maternal spiral arteries lead to uteroplacental insufficiency and subsequent damage to the entire maternal vascular system, which is also associated with placental pathology and the development of pre-eclampsia (Kenny, 2004; Norwitz, 2006). Dysfunctional angiogenesis ultimately leads to reduced blood flow and nutrient delivery to the fetus. Complex molecular mechanisms, involving hormonal and cytokine signalling, enzymatic as well as immunological processes, are involved in trophoblast invasion during placentation. It is crucial that conditions should be optimal and mechanisms occur at the correct stages for implantation to commence. Errors and alterations in underlying placentation processes such as trophoblast invasion, trophoblastic adhesion molecule expression, extracellular matrix digestion and oxygenation can ultimately lead to placental pathologies (Kenny, 2004).

Integrins and cell adhesion molecules (CAM) perform several key functions during cell migration, such as cellular signal transduction and maintenance of tissue integrity. Changes occur in the types of adhesion molecules expressed by extravillous trophoblasts during endovascular migration into myometrial arteries and subsequent transformation of uterine vessels into vascular cells (Kenny, 2004). Invasive trophoblast cells change their adhesion molecule phenotypes to resemble that of the endothelial cells they are invading or replacing. Cytotrophoblast stem cell populations, for example, express E-cadhedrin, while invasive trophoblast cells express VE-cadhedrin. Research and in vitro studies have shown that E-cadhedrin inhibits trophoblast invasion while VE-cadhedrin enhances invasion (Zhou *et al*. 1997). Other examples include changes in the expression of VCAM-1, PECAM-1 and E-selectin, into the adhesion receptor phenotypes of the endothelial cells, which the

invading trophoblast cells ultimately replace (Zhou *et al*. 1997). Abnormal expression of adhesion molecules can result in insufficient trophoblastic invasion and transformation of the maternal spiral arteries, which lead to the development of pre-eclampsia (Kenny, 2004).

Another feature, which might play a role in the development of placental pathology, is defective degradation of the extracellular matrix (ECM) surrounding maternal tissues. The ECM is a matrix consisting of a highly organized network of proteins and polysaccharides, produced by the cells within the matrix. Before trophoblastic invasion of the maternal tissues can commence, this matrix has to be enzymatically digested. Trophoblast cells subsequently express a range of proteinases, activators and inhibitors to achieve successful enzymatic degradation. Some pre-eclamptic pregnancies exhibit defective enzymatic digestion of the ECM. Proteinases studied up to date mostly include matrix metalloproteinases (MMPs), their tissue inhibitors (TIMP), as well as plasmin, plasminogen and the plasminogen activators (urokinase PA and tissue-type PA) which activate these enzymes (Kenny, 2004). The MMPs associated with invasive trophoblast phenotypes are MMP-1, -2, -3, -7, -9 and -11. Research conducted on the cells isolated from pre-eclamptic placentae showed defective modulation of MMP-9 expression. Tissue from pre-eclamptic placentae has also revealed changes in the enzymatic activities of plasminogen inhibitors and urokinase PA (uPA) (Graham *et al*. 1996). This validates further research of the roles which these molecules play during the development and pathogenesis of pre-eclampsia (Kenny, 2004).

1.4 Oxidative stress in the placental environment

Cytotrophoblast invasion and proliferation might partially be influenced by alterations in oxygen tension in the placental environment. Trophoblast proliferation is optimal under slightly hypoxic conditions, while an increase in oxygen tension is believed to trigger the invasive trophoblast differentiation pathway necessary for the invasion of

maternal tissues. Oxygen tension in the intervillous space increases between 8 and 13 weeks of gestation due to this remodelling of the maternal spiral arteries. Alterations in oxygen tension also influence the level of expression of several proteins (Lyall, 2007). Cytotrophoblast cells cultured under hypoxic conditions lack the ability to change their integrin receptor repertoire to accommodate the invasive trophoblast phenotype (Genbacev *et al*. 1997). Reduced oxygen tension also causes trophoblasts to exhibit an increased production of vascular endothelial growth factor (VEGF) as well as inflammatory cytokines (Alsat *et al*. 1996; Benyo *et al*. 1997). Some researches have hypothesized that pre-eclampsia causes the placental environment to be slightly hypoxic, which might lead to impairment of trophoblastic invasion via the above-mentioned changes in adhesion and immunological functions (Kenny, 2004). Under hypoxic conditions, the transcription factor, Hypoxia-inducible factor-1α (HIF-1α), activates the transcription of genes such as growth factor-β_3 (TGF- β_3), which is an inhibitor of trophoblast differentiation. Thus, inhibition of HIF-1α leads to the inhibition of TGF-β_3 expression, causing a reduction in trophoblast proliferation and an increase in the expression of markers of the invasive trophoblastic phenotype such as MMP-9 and α1-integrin (Semenza, 1998; Caniggia, 2000). Research has shown that HIF-1α and HIF-2α are over-expressed during pre-eclampsia (Rajakumar *et al*. 2001) and that TGF- β3 expression can be directly influenced by HIF transcription factors (Schaffer *et al*. 2003). These results have, however, not been sufficient in proving a correlation between hypoxia and the development of pre-eclampsia and further studies are necessary.

The development of pre-eclampsia is dependent only on a placenta and not necessarily a fetus, since this disorder occurs in complete hydatiform mole pregnancies (Moffet and Hiby, 2007). Insufficient uteroplacental circulation leads to oxidative stress and hypoxia in the placenta. It is, however, still unclear how this placental dysfunction might lead to the complications associated with the maternal syndrome. It has been hypothesized that a placental stimulus or factor, released into the

maternal circulation during normal pregnancy, is amplified under hypoxic conditions. Several circulating placental factors and metabolites have been considered as candidate stimuli. Vascular endothelial growth factor (VEGF) has an effect on endothelial function and regulates the permeability of blood vessels (Redman and Sargent, 2007). When VEGF levels are disturbed, vasoconstriction is activated, which may lead to hypertension.

Endothelial cells and blood monocytes are responsible for the synthesis and release of the soluble receptor for vascular endothelial growth factor (VEGFR-1 or sFlt-1). This receptor inhibits VEGF and it has been shown that the infusion of sFlt-1 into rats causes proteinuria and hypertension in these animals. It has also been shown that sFlt-1 levels are increased in pre-eclampsia, and decrease rapidly after delivery (Maynard *et al*. 2003). Another circulating factor, which is present at higher levels in pre-eclampsia patients than in normal pregnancies, is syncytial material (Redman and Sargent, 2000). This is due to an increase in trophoblast turnover in the absence of proliferation, probably resulting in a higher incidence of apoptosis in the placenta. It has also been shown that other factors which might lead to the development of pre-eclampsia, such as hypoxia, also induce trophoblast apoptosis, which strengthens the hypothesis of its role in placental pathogenesis.

The increased level of syncytial material in the maternal circulation may cause endothelial dysfunction and cell death (Crocker, 2004). The removal of syncytial material involves the inflammatory response in normal and pre-eclamptic pregnancies. Circulating placental debris is currently believed to be a key factor in the development of the inflammatory response inherent to both normal and pre-eclamptic pregnancies, since it damages the endothelium, which releases pro-inflammatory substances in response. The syncytial material itself might also elicit an inflammatory response. Whether pre-eclampsia is caused by a defective placenta or increased maternal susceptibility due to a convergence of maternal predisposing factors is an ongoing debate.

Currently, it is thought that a range of maternal and placental dysfunctions cause a cascade of symptoms, initially resulting in decreased placental perfusion and eventually leading to maternal endothelial damage. This damage to the maternal endothelium also has to be considered in the context of the systemic inflammatory response. This response occurs during normal pregnancies, but is more severe in pre-eclampsia. Certain maternal predisposing conditions may also contribute to the severity of the maternal inflammatory response observed in pre-eclampsia (Redman and Sargent, 2007).

1.5 Immunological factors

The pathogenesis of this syndrome may also be based on an immunological process whereby the maternal tissues direct an immune response against the invading fetal cells. Pre-eclampsia is associated with an inappropriately activated endothelium which leads to vasoconstriction in the maternal circulation with subsequent reduced organ perfusion. Blood pressure is controlled at the level of the vascular tree, which is composed of small arteries and arterioles, such as the maternal spiral arteries which supply the intervillious space of the placenta with blood. The endothelium plays a major role in the control of blood pressure, since it releases vasoconstrictors such as endothelin-1 and thromboxane-A2. The increase of vascular endothelial permeability and the activation of the coagulation system directly lead to reduced plasma volume and proteinuria (Wareing, 2004). The hypothesis that a dysfunction in the maternal endothelium underlies the complexity of the pathogenesis of pre-eclampsia (Roberts and Cooper, 2001) can be expanded if the endothelium is perceived as an integral part of the inflammatory network (Redman and Sargent, 2007). The increased endothelial activation associated with pre-eclampsia might be part of a greater cascade of events triggered by an over-activation of the inflammatory response (Redman *et al*. 1999). When studying pre-eclampsia, it is thus crucial to assess changes in agents involved in both endothelial and inflammatory function.

This might be due to the exposure of the maternal system to fetal antigenic material. Proteins produced by fetal cells might be identified as antigens that might trigger maternal cells to elicit an immune response. However, during subsequent pregnancies, the maternal tissues are able to recognise the fetal proteins and an immune response is unnecessary. Another theory which might explain an immunological basis for the pathogenesis of pre-eclampsia is that a short period of sexual cohabitation (and exposure to sperm antigens) poses an increased risk (Robillard *et al*. 1998). While first-time pregnancies in comparison to multiparity, pose a five times greater risk of being affected with pre-eclampsia (Myers and Brockelsby, 2004), epidemiological studies have indicated that primipaternity is the greatest risk factor for the development of pre-eclampsia. (Dekker and Sibai, 1999).

The hypothesis stating that an immune response elicited against paternal sperm antigens explains the immunological basis of pre-eclampsia, has been supported by various studies and findings. Multigravidity (with a single partner) has a protective effect against the development of the disease, however, a change in paternity with subsequent pregnancies increases the risk of pre-eclampsia. While a prolonged period of sexual cohabitation decreases the risk of developing pre-eclampsia, barrier contraception poses an increased risk (Robillard *et al*. 1995). It has also been found that pregnancies resulting from donor insemination or donated oocytes are at an increased risk of developing pre-eclampsia (Need *et al*. 1983). Such donated gametes may elicit an immune response in the maternal tissues (Kenny, 2004). A study focusing on a South-African population found that a sexual cohabitation period shorter than six months is associated with an increased risk of developing pre-eclampsia, and that, within this group, the risk of pre-eclampsia is significantly increased in multigravid women (Verwoerd *et al*. 2002).

Several studies have been aimed at a better understanding of the mechanism of the maternal immune response against invasive fetal

tissues. The Major Histocompatibility Class status of extravillous trophoblast cells determines the type of response elicited by the maternal immune system. Extravillous trophoblasts express three types of HLA class II molecules (HLA-G, HLA-E and HLA-C). It has been hypothesized that HLA-G acts as a signal of placentation by binding to receptors on maternal natural killer (NK) cells and/macrophages, which inhibits these cells to elicit an immune response (Moffett and Hiby, 2007). A debate has formed regarding the involvement of HLA-G in the development and pathogenesis of pre-eclampsia. Some studies have shown that no correlation exists between this disease and polymorphisms present in the genes encoding HLA-G or its receptor, KIR2DL4 (Aldrich *et al*. 2000; Bermingham *et al*. 2000). However, other studies have shown HLA-G expression to be decreased in placental tissue obtained from some pre-eclampsia cases (Le Gal *et al*. 1999). HLA-G expression has also been associated with invasiveness of the trophoblast since scientists have found that, in addition to being more vulnerable to an immune response, trophoblasts with reduced HLA-G expression may also have impaired invasion of the spiral arteries (Goldman-Wohl and Yagel, 2000). Several studies have focused on HLA-C due to its interaction with NK cells which are abundant at the site of implantation. NK cells express Killer Immunoglobulin-like Receptors (KIR) which interact with HLA-C. Both the KIR gene family and the gene encoding HLA-C are highly polymorphic. Another important feature of the HLA-C gene is that it is maternally imprinted, meaning that the paternal allele is always expressed (Hiby *et al*. 1999). Natural Killer cells function through cytokine and chemokine signalling, which has also been shown to influence the degree of trophoblast invasion of maternal tissues (Moffett and Hiby, 2007). Thus, during implantation, the fetal trophoblast cells may express paternally inherited HLA-C molecules for which the maternal NK cells do not possess receptors. The parental genotypes for these genes involved in the immune system might be incompatible and result in inefficient placentation and pre-eclampsia (Hiby *et al*. 2004).

A systemic immune response is evoked during normal pregnancy and reaches its acute phase during the third trimester. Many physiological and metabolic changes associated with pregnancy are in fact the result of this acute-phase inflammatory response (Redman *et al*. 1999; Sacks *et al*. 1998). These changes include an increase in the number of circulating neutrophils (Rebelo *et al*. 1995), inflammatory cytokines such as TNF-α and Interleukin-6 (Melczer *et al*. 2003, Austgulen *et al*. 1995) as well as markers of oxidative stress (Gratacós *et al*. 1998). An increase in the activation of cells involved in the immune response (neutrophils, monocytes and lymphocytes), is also observed during pregnancy (Rebelo *et al*. 1995; Sacks *et al*. 1998). Several metabolic changes of pregnancy, such as insulin resistance and hyperlipidemia, occur in conjunction with (and possibly as a result of) the inflammatory response (Martin *et al*. 1999). Women suffering from pre-eclampsia exhibit a more extreme inflammatory response than that observed in normal pregnancies. All of the above-mentioned changes which occur due to the inflammatory response during normal pregnancy are thus more severe during pre-eclamptic pregnancies (Redman and Sargent, 2007).

1.6 Fetal-Placental-Maternal interactions and responses to endothelial dysfunction

A reduction in intravascular volume, activation of the coagulation cascade and vasoconstriction give rise to the reduced organ blood flow observed during pre-eclampsia. These abnormalities are ascribed to the maternal endothelial dysfunction during pre-eclampsia (Roberts and Lain, 2002). It is thought that this dysfunction of the maternal vascular endothelium is compensated for via a range of maternal, fetal and placental responses such as hypertension, oxidative stress and changes in circulating metabolites, signalling molecules and activation of cellular responses (Crocker, 2004). These physiological and metabolic adaptations might be the effect of a signal generated by the insufficiently perfused placenta. Such a signal might be aimed at alleviating the effect

of the reduction in placental/fetal nutrient availability, and might even affect nutrient transport across the placenta.

Metabolic changes, such as the reduction in HDL cholesterol and an increase in LDL cholesterol, insulin resistance, triglycerides and uric acid (Roberts and Lain, 2002), can be detected very early during pregnancy and also persist several years postpartum (Hubel *et al*. 1998). Pre-eclampsia occurs when a reduction in placental perfusion together with the manifestation of the maternal syndrome is observed. The reduced placental perfusion alone does not result in pre-eclampsia and has also been associated with other pregnancy-related disorders and complications such intra-uterine growth restriction (IUGR) and preterm delivery (Arias *et al*. 1993). This has led to the hypothesis that the development of pre-eclampsia might partially be dependent on a maternal susceptibility to the consequences of a reduction in placental perfusion. Individuals might respond differentially to the resulting metabolic and physiological changes, which could explain the heritability factors in pre-eclampsia. Maternal factors which play a role in the susceptibility of pre-eclampsia include diabetes, hypertension, dyslipidemia, abnormal endothelial function, obesity, hypertension and hyperhomocysteinemia, which are extremely similar to the predisposing factors of cardiovascular disease in later life (Sattar and Greer, 2002).

A very important question in pre-eclampsia research is how the two stages of the above-mentioned model are linked. The mechanism, through which a reduction in placental perfusion leads to the maternal syndrome of pre-eclampsia, remains unknown. It has been postulated that fetal/placental hypoxia might act as a stimulus for this progression of the disorder. Hypoxia-associated products, such as cytokines, have the potential to be transported from the placenta to the maternal circulation where it can induce endothelial dysfunction and increase the inflammatory response (Benyo *et al*. 1997). Another mechanism of crosstalk between the placenta and maternal circulation could be syncytiotrophoblast microparticles which are shed during placental

apoptosis and syncytiotrophoblast necrosis (Huppertz *et al*. 2003). These particles, which have been shown to be increased in the blood of women suffering from pre-eclampsia, might be able to activate the inflammatory response and alter endothelial function (Redman and Sargent, 2000). It is believed that oxidative stress is the most probable stimulus for the development of the maternal systemic response in pre-eclampsia (Roberts and Hubel, 1999). Protein and lipid biomarkers of oxidative stress are increased in the blood and maternal tissues during pre-eclampsia, indicating an increase in systemic oxidative stress. Oxidative stress in the placenta might give rise to the formation of stable oxidation products which could interact with the maternal endothelium. Another promising possible signalling agent is the placental hormone, leptin, which has many functions linked to energy metabolism and adiposity (Teppa, 2000).

1.7 Genetic contribution toward development of pre-eclampsia

Several factors affect the incidence of pre-eclampsia, namely race and ethnicity, genetics, parity, obstetric and medical history. However, differences in the occurrence of pre-eclampsia between racial and ethnic groups might be due to differences in genetic factors (genotype frequencies) linked to this disorder. Personal medical history plays a pivotal role in the risk assessment of the development of pre-eclampsia, since the recurrence risk can be as high as 65%. The severity and time of onset of pre-eclampsia during a previous pregnancy affects the risk of an individual to developing this syndrome again (Reister and Kingdom 2004). Genetic susceptibility to the development of pre-eclampsia was at first believed to be based mainly on the maternal genotype (Myers and Brockelsby 2004), since a family history of the occurrence of this disease seemed to be a predisposing factor (Moffet and Hiby, 2007). However, the risk of developing pre-eclampsia is believed to be influenced by several factors, rendering this syndrome a complex trait which does not

exhibit a typical Mendelian mode of inheritance (Tower, 2004). Thus environmental, lifestyle and genetic aspects should all be considered as predisposing factors for the inheritance of such multifactorial disorders. More recent studies have also shown the importance of the fetal contribution to the development of pre-eclampsia (Esplin *et al*. 2001). It is believed that a combination of maternal and fetal genetic factors might influence the risk of developing pre-eclampsia (Moffet and Hiby, 2007).

Mutations in genes that possibly influence the susceptibility to multifactorial disorders are often common in certain populations, which would explain the higher risk for some populations or ethnic groups of inheriting specific disorders. Even though the candidate genes involved in disease progression often have minor effects, they can be classified as considerably hazardous in some populations. Research into the possible genetic factors impacting on the risk of developing pre-eclampsia has been ongoing for the last twenty years. Population stratification, as well as modifications of Mendelian inheritance patterns (incomplete penetrance, epistasis and genetic heterogeneity), complicates the investigation of the genetic basis of pre-eclampsia. A better understanding of the molecular mechanisms involved in the development of this syndrome will aid in more efficient selection of candidate genes to study further (Tower, 2004).

1.8 Screening and treatment strategies

Currently, the only cure for pre-eclampsia is the delivery of the placenta. Several prevention and treatment strategies have been analysed and studied, of which aspirin treatment seemed to be the most promising treatment of pre-eclampsia. Women identified as being at high risk of developing pre-eclampsia due to abnormal Doppler ultrasound readings (an indication of abnormal trophoblast invasion), have been shown to benefit from low-dose aspirin treatment at 14-16 weeks of gestation (Ebrashy *et al*. 2005; Duley *et al*. 2007). However, the development of severe pre-eclampsia leading to preterm labour has been shown to

drastically decrease with regular aspirin administration from the first trimester onwards (Nicolaides *et al*. 2006; Vanio *et al*. 2002).

In order to provide women at high risk of developing pre-eclampsia with the necessary antenatal care and treatment strategies before the onset of the disease, it is thus crucial to diagnose them during the first trimester (Papageorghiou and Campbell, 2006). Extensive research has been conducted in the search for maternal or fetal biomarkers which are present at an early gestation and are informative of underlying placental or other pregnancy-related pathologies. Such markers could aid in the detection of those individuals who would respond well to preventative treatment strategies such as dietary calcium supplementation and the administration of antiplatelet agents (Gonen *et al*. 2006). Early intervention could help prevent hypertensive emergencies and permanent damage of the maternal vascular system. Antioxidant vitamins and anticoagulation treatment might also aid in reducing the severity of pre-eclampsia (Nicolaides *et al*. 2006).

Different early signs of pre-eclampsia can be identified during the first and second trimesters. Bleeding during the first trimester has been associated with a twofold-increased risk for the development of pre-eclampsia. Impaired plasma volume expansion during the first trimester may also aid in the identification of women with an increased risk to developing pre-eclampsia. However, the non-invasive identification of such candidates has not been perfected (Reister and Kingdom, 2004). During the second trimester, impaired reduction in haemoglobin concentration and an increase in blood pressure are indicative of an increased risk to developing pre-eclampsia. Impaired utero-placental perfusion and damage to the maternal vasculature may disrupt the maternal-fetal barrier. This allows fetal molecules to enter the maternal circulation. An increase in certain fetal molecules such as α-fetoprotein (AFP) and human chorionic gonadotropin (hCG), is associated with an increased risk of presenting with pre-eclampsia (Reister and Kingdom, 2004). A preliminary diagnosis of pre-eclampsia based on elevated blood

pressure and proteinuria, can be further substantiated by certain accompanying signs. HELLP syndrome (hemolysis, elevated liver enzymes and low platelets) complicates 4-12% of severe pre-eclampsia cases. Certain biochemical tests exist which could be predictive of pre-eclampsia, such as an increase in plasma fibronectin or the levels of specific maternal serum markers. Liver function tests may also be important in detection of the progression of pre-eclampsia (Hayman and Myers, 2004).

An effective diagnostic tool for pre-eclampsia is Doppler ultrasound measurement, which represents the degree of impedance to utero-placental blood flow during pregnancy. It has been found that the combination of a Doppler ultrasound reading with the measurement of certain maternal serum markers is the most effective means of diagnosing pre-eclampsia (Papageorghiou and Campbell 2006, Spencer *et al*. 2007). Impedance to blood flow in the uterine arteries decreases with gestation in normal pregnancies while it increases during pre-eclamptic pregnancies. Doppler ultrasound measurement is a non-invasive method of predicting the risk of developing pre-eclampsia during the first and second trimesters. This prediction is, however, associated with a high false positive rate (Spencer *et al*. 2007).

Although Doppler detects only half of all pre-eclampsia patients (early onset of syndrome and/or delivery before 34 weeks of gestation), its success increases to 90% detection rate when used in conjunction with the promising predictive serum marker, Placental Protein 13 (PP13) (Nicolaides *et al*. 2006). This protein is expressed on the apical membrane of the placenta and is released into the maternal circulation throughout the pregnancy. During normotensive pregnancies, PP13 levels gradually increase from the first to the third trimester. Recent studies have shown that the expression patterns of PP13 are significantly altered in pre-eclamptic pregnancies. The concentration of PP13, measured in the maternal serum during the first trimester, is markedly reduced in women who develop severe pre-eclampsia later

during pregnancy when compared to healthy pregnancies (Nicolaides *et al*. 2006). However, the third trimester serum PP13 levels of women either suffering from pre-eclampsia or who will later develop this disorder, are 50-70% higher than that of normal pregnancies (Burger *et al*. 2004; Than *et al*. 2007). It has also been found that syncytiotrophoblast expression of PP13 mRNA and protein is severely decreased in third trimester pregnancies which develop pre-eclampsia (Than *et al*. 2008). Diagnostic Technologies Limited (DTL) has developed a diagnostic kit for the prediction of pregnancy-related disorders such as pre-eclampsia. The method implemented in this kit is also based on the measurement of serum PP13 concentration during pregnancy [3].

Women at the highest risk of developing pre-eclampsia can be identified by initially screening serum PP13 levels and subsequently performing Doppler ultrasound screening on the selected individuals. If this method is followed, the detection rate of pre-eclampsia is 90% with a false positive rate of 6% (Nicolaides *et al*. 2006; Papageorghiou and Campbell, 2006, Spencer *et al*. 2007). The severity and time of onset of pre-eclampsia have also been correlated with PP13 levels during the third trimester. Early onset, severe pre-eclampsia is associated with a significant increase in maternal serum PP13 concentrations while late onset, less severe cases of pre-eclampsia showed only a slight increase (Than *et al*. 2007). The possible reasons for the difference between normal and pathological pregnancies in serum PP13 levels could be abnormal primary protein structure, dysfunctional protein synthesis and impaired migration from the placenta to the bloodstream, or a combination of these factors. It has also been suggested that the gene encoding PP13 might be down-regulated in pregnancies which require early delivery due to pre-eclampsia (Burger *et al*. 2004).

[3] An Israel-based medical diagnostic and biotechnology company. DTL's main focus is the development and production of diagnostic kits, using Placental Protein 13 as a predictive biomarker, to screen for pregnancy-related disorders such as pre-eclampsia and preterm labour .

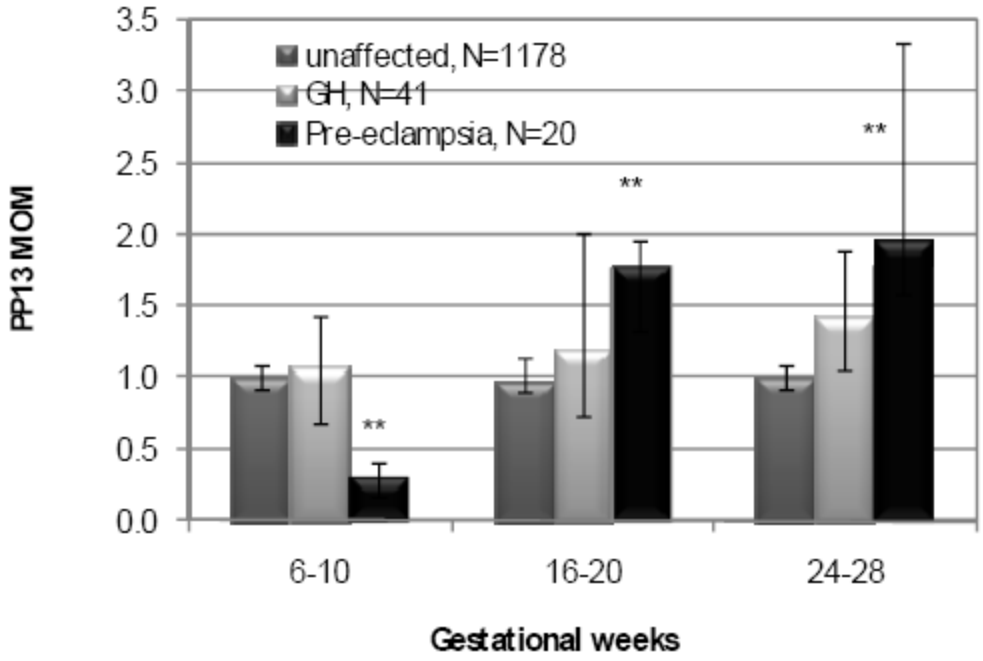


Figure 1: Comparison of maternal serum PP13 levels (expressed as multiples of the normal median, MoM, for each gestational period) from unaffected (healthy), pre-eclamptic and gestational hypertensive (GH) pregnancies. (Source: Gonen *et al*. 2006)

Figure 1 (Gonen *et al*. 2006) clearly depicts the alteration in PP13 levels during pregnancy, as well as the difference in PP13 concentration between unaffected, gestational hypertensive and pre-eclamptic pregnancies. During the first trimester of pregnancy, women who subsequently developed pre-eclampsia, exhibited significantly lower levels of PP13 (0.30 MoM, 95% CI 0.15-0.40) than unaffected women ($P<0.001$). However, during the second and third trimesters, PP13 levels in pre-eclamptic pregnancies were found to be significantly higher (1.78 MoM, 95% CI 1.36-1.94 and 1.97 MoM, 95% CI 1.55-4.32) than that measured in unaffected pregnancies ($P<0.001$).

2 Placental Protein 13

Placental Protein 13 (PP13) is a relatively small protein with an approximate weight of 15.6kDa and consists of 139 amino acids (Burger *et al*. 2004; Than *et al*. 2004). Sequence and structural homology has shown that PP13 is a member of the galectin protein family, which

classically has a strong binding capability to sugar residues via a conserved carbohydrate recognition domain (Visegrády *et al*. 2001; Yang *et al*. 2002). PP13 is also referred to as Galectin 13 and forms a homodimer via disulphide bonds, similar to other galectins. The sugar-binding function of PP13 seems to rely on its secondary structure since reducing agents disrupt the disulphide bonds of the homodimer structure, decreasing its sugar-binding activity (Visegrády *et al*. 2001). PP13 has been shown to specifically bind to sugar residues expressed in the placenta, such as mannose, *N*-acetyl-lactosamine and *N*-acetyl-glucosamine (Burger *et al*. 2004; Than *et al*. 2004). During pregnancy, PP13 is specifically expressed in trophoblasts and on the brush border membranes of syncytiotrophoblasts, which indicates that it might play a role during implantation (Burger *et al*. 2004; Than *et al*. 2004). As a galectin, PP13 probably aids in trophoblast migration into and invasion of the maternal spiral arteries via the induction of calcium-driven catalytic reactions in these placental and fetal membranes (Burger *et al*. 2004).

The effect of PP13 on trophoblasts was studied by Burger *et al*., and it was found that the protein induces calcium depolarisation of the cell which leads to an increase in fatty acid and prostaglandin release. Women suffering from pre-eclampsia, who have lower than normal PP13 levels during the first trimester, might thus exhibit abnormal ratios of vasodilatory and vasoconstrictory agents such as thromboxane and prostacyclin (Burger *et al*. 2004). This might lead to impaired oxygenation in the placenta, thereby causing pregnancy pathologies. Since the exact function of PP13 during placentation is unknown, it is unclear whether a dysfunctional PP13 is the result of abnormal implantation, or whether it is the cause of the pathology (Burger *et al*. 2004).

PP13 is encoded by the gene, *LGALS13* (lectin, galactoside-binding, soluble 13), which is located on the long arm of the human chromosome 19, in close proximity to another four known galectin coding genes and several putative genes (Visegrády *et al*. 2001; Yang *et al*. 2002).

LGALS13 is highly homologous to *LGALS10*, which encodes the Charcot Leyden Crystal (CLC) protein. These two proteins also exhibit the same secondary structure, folding into five - and six stranded β-sheets as well as two α-helices (Visegrády *et al*. 2001). *LGALS13* has four exons and the part of the sequence which translates into the carbohydrate recognition domain (CRD) is located in the third exon (Visegrády *et al*. 2001). The CRDs of PP13 and CLC are very similar due to the substitution of three of the seven highly conserved amino acids in their CRDs (Yang *et al*. 2002). A characteristic unique to these two galectins and absent in the other family members, is the ability to bind to mannose and *N*-acetyl-galactosamine (Than *et al*. 2004). The tertiary structure of PP13 has not yet been fully elucidated, however, one can infer that putative disulfide bonds hold together the two identical subunits of which the protein is composed (Visegrády *et al*. 2001; Than *et al*. 2004). The location of Cysteine (Cys) residues in the peptide aids in predicting the location of these disulfide bonds. It is thought that the tertiary structure of PP13 is crucial for its ability to bind to sugar residues, as is the case with other galectins (i.e. Galectin-1and Galectin-2). This will explain the decrease in sugar-binding capability in the presence of reducing agents (Than *et al*. 2004). The Cys residues thought to be involved in disulfide binding are also found in the highly homologous putative gene sequences but are absent in CLC. Homologous genes existing in close proximity of another on the same chromosome and which share certain functions often derive from an evolutionary gene duplication event (Yang *et al*. 2002). This might be the case with the cluster of galectin genes on human Chromosome 19. Certain diversions from this sequence homology, such as the position of Cys residues, suggest possible function diversion events taking place in this protein family. Changes in the environment of PP13 could impact on the tertiary structure and biological activity of this protein (Than *et al*. 2004).

Putative phosphorylation sites have also been identified on the outer surface of PP13 (Yang *et al*. 2002) and it has been shown that PP13 expressed *in vivo* is phosphorylated (Than *et al*. 2004). Phosphorylation

regulates sugar-binding by modulation of the CRD of Galectin-3, which suggests that the putative Serine and Tyrosine kinase residues might play a similar role in PP13 (Yamazaki *et al*. 2001). PP13 possibly cross links with Annexin II and beta/gamma actin which have been shown to co-localise with PP13 to the placenta and fetal hepatic cells. This gives an indication that PP13 is transported out of the cell via ectocytosis, contained in a vesicle with annexin II and beta/gamma actin. PP13 exhibits a weak lypophospholipase activity which is thought to aid in the slow, steady release of PP13, from the vesicle by which it is transported, to the external surface of the syncytiotrophoblasts' plasma membrane (Than *et al*. 2004). The fact that no signalling peptide for transporting PP13 across cellular membranes is detected in the amino acid sequence, supports the hypothesis of ectocytosis (Yang *et al*. 2002). The subsequent gradual release of PP13 might aid in the prevention of blood clotting in the low blood flow organs where it is expressed. The co-localisation of Annexin II and beta-gamma Actin with PP13 on the outer surface of the syncytiotrophoblasts is also significant as these proteins are thought to be involved in the haemostatic functions of the placenta during implantation and pregnancy (Than *et al*. 2004).

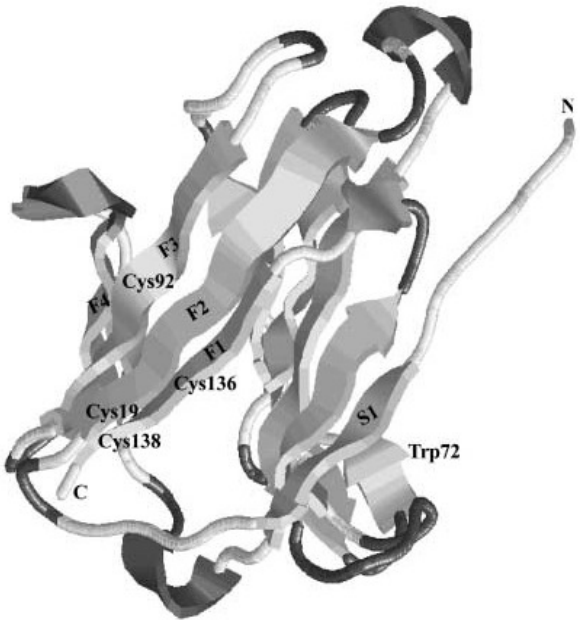


Figure 2: Image depicting the tertiary structure of PP13. Some of the highly conserved amino acid residues, beta sheets S1 and F1-F4, as well as the C- and N-termini of the protein are shown. Trp72 on beta-sheet S6a is situated in the putative carbohydrate recognition domain.

The Cysteine residues (Cys 136, 138, 19 and 92), which are putatively involved in dimerisation of the protein, are also shown. (Source: Than *et al*. 2004).

3 Galectins

3.1 The evolution of multigene families

The genes encoding galectin proteins are classified as a multigene family. The main event leading to the formation of multigene families is gene duplication, whereby several copies of an ancestral gene might come to exist within a genome. The redundant copies of a gene often undergo nucleotide substitution, insertions and deletions. Selective pressures act on these functional gene duplicates which can either result in the attainment of novel functions and/or the loss of the original functions of the ancestral gene with possible formation of a pseudogene. In this way, clusters of homologous genes have evolved into extended gene families with differential biological functions (Papadakis and Patrinos, 1999).

3.2 Structural and functional classification

Lectins form a large group of proteins that specifically interact with oligosaccharide molecules. Two classes of lectins have been characterised in vertebrates, namely the C-type lectins and the S-type lectins. Pentraxins and selectins are examples of C-type lectins, which are calcium dependent for carbohydrate interaction to occur. S-type lectins, also known as galectins, are calcium independent and are found in a wide range of species (Perillo *et al*. 1998). The galectin protein family plays important roles in a vast range of molecular and cellular functions, including immunological activities, inflammation, cell proliferation, migration and adhesion, RNA splicing, remodelling apoptosis activation/inhibition and tissue differentiation (Cooper and Barondes, 1999; Visegrády *et al*. 2001; Yang *et al*. 2002).

A tissue-specific expression pattern is an important feature of galectins, which are usually regulated on a developmental level (Cooper *et al*. 2002). It has also been postulated that one galectin might perform different functions in different tissues depending on the cell type and available ligands (Barondes *et al*. 1994). For example, both Galectins-1 and 3 can inhibit cell adhesion when bound to the polylactosamine chains on laminin. Galectin-1 can, however, also promote cell-matrix adhesion by cross-linking glycoconjugates with a cell surface. Galectin-1 is expressed in a wide range of tissues such as thymus, placenta, kidney, muscle and neurons, while Galectin-2 seems to be mainly expressed in hepatomas. Both Galectins-3 and 4 are expressed in intestinal epithelial cells (Barondes *et al*. 1994). However, Galectin-3 is mainly expressed in basophils, activated macrophages, mast cells and epithelial cells, hence its function during inflammatory responses. Galectin-9, 10 and 14 are also believed to be involved in immunological functions, especially during allergenic responses (Rabinovich *et al*. 2002). Galectin-10 is specifically expressed in basophils and eosinophils and has been isolated from a range of tissues such as heart, spleen, bone marrow, colon and testis. Some galectins (Galectin-1 and 3) may also have distinct intracellular and extracellular functions (Cooper and Barondes, 1999). However, it is thought that galectin functions might not be absolutely crucial to sustain life, but that biological processes would be carried out more efficiently with galectins. This notion is substantiated by the fact that knock-out mice lacking both Galectins-1 and 3 were alive (Cooper and Barondes, 1999). The great number of galectins involved in pathogenesis and immunological functions might suggest that these proteins play a vital role in rate limiting functions during disease progression (cancer and inflammation). Galectins thus might also prove to be of great therapeutic and diagnostic value in the future (Cooper *et al*. 2002).

Galectins are characterised by their affinity for β-galactoside, which enables them to connect components in the extracellular environment with the cells in which they reside. Due to their di- or multivalency,

galectins are capable of interacting with, and subsequently cross linking, multiple sugar residues at the same time. This characteristic might provide a mechanism by which galectins mediate cell adhesion via cell-cell and cell-matrix interactions (Barondes *et al*. 1994; Cooper *et al*. 1994; Cooper and Barondes, 1999). The locations of galectins are either cytosolic, submembranous or nuclear (Visegrády *et al*. 2001; Cooper *et al*. 2002; Yang *et al*. 2002). Another feature shared by most known galectins is the absence of a secretory signal peptide sequence which has led to the hypothesis that these proteins are transported across the cell membrane via a non-classical mechanism (Cooper *et al*. 2002; Yang *et al*. 2002). Many theories exist concerning the reason for this mechanism of transport. One is that a non-classical secretory pathway provides a mechanism for cells to selectively secrete specific galectins, as opposed to all galectins being secreted at the same time via the recognition of a single transport peptide. Another reason might be to enable the galectins to separate from their bound carbohydrate ligands, which will ensure that these interactions only take place outside the cells (Barondes *et al*. 1994, Cooper *et al*. 1994).

Galectins are divided into three main structural groups, namely the prototype galectins (Galectins-1, 2, 5, 7, 10, 11, 13, 14), tandem-repeat galectins (Galectins-4, 6, 8, 9, 12) and the chimera galectins (Galectin-3) (Cooper and Barondes, 1999; Rabinovich *et al*. 2002; Baum *et al*. 2002; Yang *et al*. 2002). The origin of these distinct galectin classes is believed to be based upon the different cross-linking properties of each distinct structure. Prototype galectins generally exist as homo-dimers via the self-association of two identical monomer subunits. This enables cross-linking of different ligands since the two CRD pockets of the monomeric subunits face to the outside of the dimer. Tandem-repeat galectins are characterised as two non-identical subunits joined with or without a linker peptide. Galectins belonging to this class are able to interact with multivalent ligands since their CRDs can simultaneously bind to several structures. This might enable tandem-repeat galectins to crosslink different ligands as opposed to the dimeric properties of prototype

galectins. Chimeric galectins have one galectin domain attached to a distinct N- or C-terminal domain (Cooper *et al*. 2002).

3.3 Carbohydrate recognition domain (CRD)

A characteristic core sequence of approximately 130 amino acids, usually encoded by one exon, is found in all galectins. These core residues fold into two antiparallel β-sheets, forming a globular tertiary structure (Cooper and Barondes, 1999). The adjacent β-strands form a pocket that enables carbohydrate binding. This core sequence thus confers carbohydrate-binding capability, which explains its high conservation in mammalian and other galectins. The concave side of the pocket is formed by six β-strands (designated S1-S6) while the five β-strands form the convex side (F1-F5). The carbohydrate ligand fits into the groove of the pocket, at the concave side (Cooper 2002). Scientists are still uncertain of the number of ligands galectins are able to interact with at a given time and of the impact the galectin structure has on its specificity. It is crucial to study the different types of ligands galectins interact with, in order to gain a better understanding of the cell and time-specific functions galectins carry out (Perillo *et al*. 1998). The intracellular function of the CRD is also still unclear and further research has to elucidate this (Cooper 2002). Although a small number of residues (the core region of the CRD) are conserved in all galectins, other residues within the CRD might be conserved within certain species or subgroups of galectins.

Due to evolutionary processes, certain galectin relatives (such as CLC and PP13) exhibit amino acid substitutions in this core region, resulting in a loss of L-galactoside binding activity. However, these galectin relatives show specificity to other sugar residues which might give them novel functions (Cooper and Barondes, 1999; Yang *et al*. 2002). Several putative galectin genes have been identified in close proximity to the genes encoding PP13 and CLC on chromosome 19 (Cooper and Barondes, 1999; Yang *et al*. 2002). This cluster of putative genes on the

long arm of human chromosome 19, are surrounded by pseudogenes and very high conservation is observed in this area. The order of the known and putative genes in this area is: *LGALS10*, *LGALS14*, LOC400696 (AC005515-I), LOC148003, *LGALS13*, LOC390930. *LGALS4* and *LGALS7* are located further down the chromosome, toward the centromere (Cooper *et al*. 2002). Standard intron/exon boundaries are present in all these genes and absent from the pseudogene sequences (Cooper *et al*. 2002; Yang *et al*. 2002). Of these genes, *LGALS13* and *LGALS14* and LOC400696 are expressed in the placenta while *LGALS10* is expressed mainly in eosinophils and basophils (Cooper *et al*. 2002).

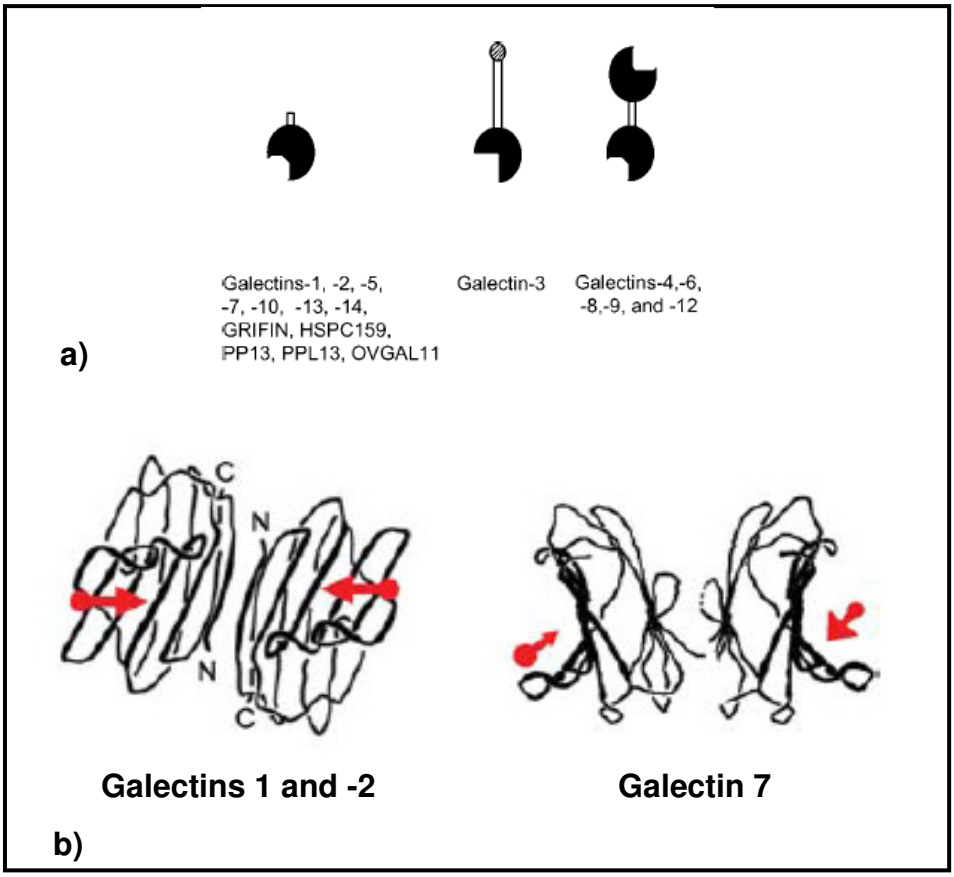


Figure 3
a) A graphical representation of the different subunit types according to which galectins are classified. Black shaded areas depict the carbohydrate recognition domains (CRDs) of these peptides. Galectins with single CRDs (1, 2, 3, 5, 7, 10, 13 and 14) can form monomers or dimers and Galectin-3 has been shown to form oligomer structures.

b) Structural representations of two types of galectins with different dimer interfaces. The region and direction of dissacharide bonding is indicated by red arrows which point from the reducing to the non-reducing end of each structure.
(Source: Leffler *et al.*2004)

It has been suggested that a Galectin-10 (CLC) subfamily of galectins, with distinct features, has evolved. CLC differs from other galectins in that it cannot bind to β-galactose, but seems to have an affinity for mannose instead (Cooper *et al*. 2002). The CRD of CLC has a similar structure to that of other galectins. The unique sugar binding interactions of this protein are the result of specific amino acid changes and different chemical bond formations in this region. For instance, the glutamic acid in position 75 of most galectins is replaced by a glutamine, which interacts with mannose instead of galactose (Cooper and Barondes, 1999). These subtle changes seem to be conserved in this new galectin-10 subfamily, although some of these putative genes might have more conserved CRDs than CLC (Cooper and Barondes, 1999). However, the preference for mannose binding is likely to be a key attribute of this new subfamily. This new subfamily seems to have evolved quite recently due to the absence of orthologues to these putative genes in non-primates and the tissue-specific expression profiles of the members (Cooper and Barondes, 1999). Orthologues of these subfamily members have been described and some of these sequences are available in databases such as GenBank and Ensembl. The mouse CLC-like sequence is very similar to one of the putative galectin-10 subfamily members, LOC400696 (Cooper, 2002). The sheep CLC-like protein, ovgal11, seems to be up regulated during inflammation but has not been shown to exhibit lectin activity. The tissue specificity of the new galectin-10 subfamily members might be the reason for their evolution. The fact that the majority of these proteins (PP13, Galectin-14 and LOC400696) are expressed in the placenta, leads to hypotheses around their functions during pregnancy. It is also noteworthy that these new family members are only found in primates, with very specific placental anatomies.

3.4 Regulation of galectin expression

Galectins exhibit differences in the timing and location of their expression (Chiariotti *et al*. 2004). Some galectins are expressed only in specific tissues while others are only expressed during certain developmental

stages (Cooper and Barondes, 1999). PP13 is a good example of the tissue and developmental stage specific regulation of galectins in that it is only expressed in the placenta, fetal spleen and liver during pregnancy. The main molecular mechanism governing tissue and developmental specific expression of proteins is based on the activity of upstream sequence elements in the promoter areas of the genes encoding these proteins. Regulation of the sub-cellular localisation of different galectins might also be a mechanism that modulates their activity. The expression of galectins-1 and 3 are induced by, amongst others, phorbol esters. *Cis*-elements which putatively bind to phorbol ester sensitive transcription factors, AP-1 and -2, were identified in the promoters of the genes encoding galectin-1 and 2 (Gaudin *et al*. 1997). Epigenetic mechanisms of gene regulation are also important for the correct timing and location of galectin expression. Methylation has been shown to be an important mechanism of regulating the expression of certain galectins (Galectin-1) (Benvenuto *et al*. 1996; Cooper, 2002). Comparison of the new galectin-10 family members reveals high similarity in their promoter areas which suggests that, if these putative genes are translated, their proteins are most likely co-expressed (Cooper and Barondes, 1999). Three consensus cap sites, which are known to regularly act as transcription start sites, are present in the same locations of all the members of this subfamily. The promoters of all the new family members seem to be TATA-less and this is substantiated by a CCAAT box approximately 35bp upstream from one of the putative transcription start sites. The new family members do not, however, contain all the granulocyte-specific *cis*-elements present in the galectin-10 promoter area. It is possible that similar sites have simply not been identified in the new members. A placenta-specific *cis*-element, which recognises TEF-5, has been identified in PP13, AC005515 and galectin-10 (Jacquemin *et al*. 1997). The fact that the galectin-10 subfamily seems to contain both immune-response specific and placental *cis*-elements, might indicate that the new subfamily's functions combine the immunological functions of granulocytes with the placentation mechanisms of trophoblasts (Cooper *et al*. 2002).

4 Eukaryotic gene structure and expression

4.1 Transcriptional regulation of expression, promoter structure and analysis

The central dogma of molecular genetics is that a gene, composed of DNA, is transcribed into RNA which is translated into a chain of amino acids, forming a polypeptide or protein. Correct RNA and protein synthesis requires regulatory mechanisms to govern the time, localisation and level of expression of different genes. Eukaryotic gene expression is regulated at the transcriptional and translational level. Transcriptional regulation of genes is of great importance in almost all eukaryotic cellular processes (Wingender *et al*. 2000), and is mainly governed by the interaction of *cis*- and trans-acting elements. *Cis*-regulatory elements are situated in the promoter regions of eukaryotic genes and direct binding of regulatory proteins (trans-acting elements), thereby governing gene expression. Promoter regions are usually located upstream from the gene which they control and are identified by the presence of certain consensus DNA sequences. The consensus sequences of certain regulatory DNA elements commonly found in the promoter regions of eukaryotic genes have been extensively described and studied. For instance, in eukaryotic genes, the transcription start point is usually the central Adenine nucleotide base in a PyAPy sequence element, where Py represents a pyrimidine base (Cytosine or Thymine) and A refers to the Adenine nucleotide base. A conserved DNA sequence element, called the TATA box, is usually located at position -25 to -30 of a eukaryotic mRNA gene. The basal transcription complex, which consists of a number of proteins which interact during transcription, assembles around this region. One of the transcription factors comprising the basal transcription complex, TATA-binding protein (TBP), recognizes this consensus sequence on the DNA molecule and specifically binds to the region where the TATA box is situated. The remaining transcription factors subsequently assemble at this position by binding to TBP and each other, which eventually initiates transcription.

Other consensus sequences, such as the CAAT and the GC boxes, as well as the Kozak sequence (Kozak, 1981) may also be present in the promoter regions of eukaryotic genes (Fairbanks and Andersen, 1999).

Promoters can be divided into proximal and distal regions, the proximal region usually contains the TATA-box and transcription start site (TSS) and is subsequently responsible for the assembly of the basal transcription complex (Fairbanks and Andersen, 1999; Ohler, 2006). Distal promoter regions usually regulate time, location and level of expression of genes. The distal parts of promoters contain conserved sequences called *cis*-acting regulatory elements (CAREs) upstream of the core promoter. Transcription factors recognize these sequence elements, bind to them and either enhance or repress the rate of transcription initiation by RNA polymerase. CAREs are usually between 5 and 20 nucleotides in length and can be situated up- or downstream of the TSS (Molina and Grotewold, 2005). The detection of these elements is complicated by their short sequences. Eukaryotic enhancer and repressor elements are often organised into domains according to their activating or silencing properties, but they can also occur dispersed and repeated throughout a promoter region (Molina and Grotewold, 2005). Promoters differ significantly between different genes and species, and many exceptions to the above-mentioned consensus sequences and locations of regulatory elements are found. Some genes have, in addition to a promoter region, enhancer elements within the DNA sequence. These elements' direct binding of certain proteins (activators) usually act via enhancement of the transcription process. These enhancer elements can be situated upstream, downstream or inside a gene. The termination of transcription is also governed by consensus sequence elements. The RNA molecules produced by RNA Polymerase II are usually cleaved 11-30 nucleotides downstream from the highly conserved sequence, AAUAAA.

DNA packaging also plays a pivotal role in regulation since transcription factors cannot bind DNA which is organized into a condensed chromatin

structure (Molina and Grotewold, 2005). DNA packaging is influenced by DNA methylation and transcription factors and co-activators, which can alter the chromatin structure (Wolffe, 1998; Saha *et al*. 2006).

The detection of core promoter sequences and *cis*-regulatory elements is challenging due to their complex nature and the short length of the sequence elements involved. Promoter prediction differs from gene prediction in the fact that promoters are not structurally well-defined (Rombauts *et al*. 2003). Nevertheless, structural and sequence-based algorithms are currently used to identify regulatory sequences. Sequenced-based algorithms, such as PromoterInspector, compare the sequences of proposed promoters to those of non-regulatory regions (Scherf *et al*. 2000).

Double stranded DNA exists in three differents conformers namely, B-, Z- and A-DNA. B-DNA is the most common conformation found *in vivo*. GC-rich areas may fold into a Z-DNA structure, which, unlike B-DNA, is a left-handed helix conformation. A-DNA may arise when dehydration occurs, under non-physiological conditions. Another difference between B-and Z-DNA is that B-DNA has similar major and minor grooves whereas Z-DNA has deep minor and shallow major grooves. The bases in B-DNA are almost perpendicular to the helical axis (Dickerson *et al*. 1982). Bases are, however, not always coplanar with respect to each other. Propeller twist gives an indication of the degree to which two bases are twisted (El Hassan and Calladine, 1996).

Structure-based algorithms distinguish promoter regions according to specific structural characteristics such as the chemical properties of Z-DNA, the curvature and twist of B-DNA, DNA propeller twist, bendability, the stability of the duplex DNA structure and the presence of CpG islands. An example of such an algorithm is McPromoter, which integrates these different structural features in order to predict promoter regions (Rombauts *et al*. 2003).

Genes with similar functions or expression patterns are usually co-regulated and co-expressed, which suggests that possible regulators might be identified by searching promoter sets of such co-expressed genes for frequently occurring nucleotide patterns (motifs) (Rombauts *et al*. 2003). Such over-represented and biologically relevant nucleotide patterns have in previous studies been identified with the aid of bioinformatic algorithms and other tools designed for this purpose. POCO is an example of such a computer program which is able to identify over-represented nucleotide patterns in a set of sequences (Kankainen *et al*. 2006). When used to test two unique, oppositely expressed promoter sets, POCO cross-validates the identified sequence motifs, aiding in the identification of motifs that are unique for different gene sets or pathways. POCO assumes that both gene sets are regulated by transcription factors and that the same transcription factor cannot up-regulate one gene set while down-regulating another. It is thus expected to identify different transcription factors in differentially expressed gene sets (Kankainen *et al*. 2006). Other programs which are used to detect over-represented nucleotide patterns include MotifSampler (Thijs *et al.* 2001) and Weeder (Pavesi *et al*. 2006). MotifSampler includes a matrix-based algorithm which identifies nucleotide motifs. Weeder evaluates each motif based on the number of sequences in which it occurs (Pavesi *et al*. 2006).

Conserved promoter sequences can also be identified using phylogenetic footprinting, a process which compares the promoter areas of a set of orthologous genes (from different species) to identify conserved motifs. Algorithms based on this method discard sequences that are too similar and correct sequences which do not contain the conserved motif. When the phylogenetic distance between two organisms being compared is either too great or too small, their sequences are less informative and less efficient in detecting conserved motifs. Phylogenetic footprinting can successfully be carried out using multiple sequence alignments generated with algorithms such as ClustalW (Thompson *et al*. 1994). Motifs should be conserved over long

evolutionary periods and alignments should contain few mismatches (Rombauts *et al*. 2003). MEME is a probalsitic motif detection program which disregards noise from those parts of the sequences lacking the conserved motif (Rombauts *et al*. 2003). PromH uses orthologous gene sequences to improve promoter prediction accuracy and quality (Solovyev and Shahmuradov, 2003).

Once over- or differentially represented sequence motifs are identified, it is possible to compare these to international *cis*-element databases to determine their identity, given that these motifs have been previously documented and submitted to the databases. Several databases with known promoter and other regulatory sequence information are available for this purpose. TRANSFAC contains the sequences of transcription factors, their DNA-binding sites as well as the sequence profiles of these sites (Wingender, 2000).

In conclusion, there are several intrinsic characteristics of promoter and *cis*-regulatory sequences which complicate their unambiguous identification. *In silico* prediction programs are useful tools for the detection of genomic signals and elements involved in gene regulation (Wingender, 2000). These detection tools and software have to be altered for species-specific investigation of promoter regions. *In silico* prediction tools are efficient for the simultaneous analysis of great numbers of promoter regions. This aids in phylogenetic footprinting of orthologous promoters from different species (Rombauts *et al*. 2003).

4.2 Post-transcriptional regulation and alternative splicing

Before the initiation of translation into protein products, pre-mRNA molecules are processed into mRNA molecules. Eukaryotic mRNA processing involves polyadenylation of the 3’ end of the molecule, the addition of a methylguanine cap to the 5’ end and the removal of the

introns. Eukaryotic genes are composed of short exons interspersed with longer intron sequences. After transcription, introns are spliced out and the exons are joined to produce mature mRNAs which are finally translated to proteins. This process is mediated by the action of a set of five small nuclear RNA molecules (snRNAs) and various polypeptides, responsible for the accurate recognition of intron/exon boundaries. These consensus sequences direct the correct splicing of exons, which is called the GT-AG rule (describing the nucleotide sequence of the sense DNA strand). Other consensus sequences have also been identified around the exon-intron boundaries, but are not as commonly found as the GT-AG rule (Fairbanks and Andersen, 1999).

One of the reasons for the extensive complexity of the proteome is alternative splicing. Some genes can be differentially spliced to create different isoforms from the same transcription product, which partly explains the vast amount of proteins being produced from the seemingly small number of genes (~40 000) contained in the human genome (Cartegni *et al*. 2002). Such genes contain numerous possible intron/exon boundaries which complicate the correct recognition of these sites. Alternative splicing is often regulated in a tissue- or developmental stage specific manner, as is the case for the expression of either calcitonin or tropomyosin isoforms (Strachan and Read, 2004). Alternative splicing can also govern protein location and function. This is achieved via *cis*-acting elements, such as enhancers and silencers, in the sequences of alternatively spliced genes. These *cis*-acting elements are recognised by specific trans-acting factors and either enhance or antagonise splice site recognition (Strachan and Read, 2004). Mutations in these regulatory motif sequences can impact significantly on the alternative splicing mechanism of eukaryotes and have been shown to be the cause of several genetic disorders in humans (Cartegni *et al*. 2002; Hims *et al*. 2007; Wang and Cooper, 2007).

Two models for exonic splice element function exist, namely the antagonist and the recruiting models. Serine/Arginine-rich (SR) proteins

(splicing factors) bind specifically to exonic splice enhancers (ESE) on RNA molecules. This leads to either the recruitment of splicing machinery to the site in question, or the inhibition of nearby silencer elements. These models may function simultaneously, as has been shown for the splicing of immunoglobulin M pre-mRNA. It is thought that splicing factors can recognise a vast range of ESEs (Cartegni *et al*. 2002).

Although not as well defined and understood to date, silencer elements are thought to be as prevalent as ESEs. Most silencing sequences occur in intronic regions, however, some exonic splicing silencers (ESS) have been identified. It is thought that some exons are held at a silenced state unless neutralised via the action of splicing signals and enhancer elements. Silencers probably interact with heterogeneous nuclear ribonucleoproteins (hnRNP), which act as negative regulators. Three hypotheses have been formulated for the mechanism by which silencers function. Several hnRNPs might bind to sites around a silenced exon and cause a section of the premRNA to loop outward, subsequently preventing splicing from occurring. It is also possible that inhibitory and stimulatory factors compete for overlapping enhancer and silencer sites. If an inhibitory factor blocks the stimulatory factor from binding to its corresponding enhancer element, that motif might be silenced. The third hypothesis is that binding of hnRNPs to high-affinity exonic sites, recruits inhibitory hnRNP complexes to coat the premRNA, which disables spliceosome assembly and antagonises enhancer activity (Cartegni *et al*. 2002).

The mechanisms involved in alternative splicing are complex and involves extensive crosstalk between different enhancer and silencer sites as well as the splicing factors they interact with. It is clear that sequence variations which disrupt these elements can cause major alterations in the splicing mechanism of genes. Splicing mutations can either cause exon skipping or activation of cryptic splice sites. Since splice enhancing and silencing elements occur both in coding and non-

coding regions of genes, intronic mutations, which usually do not directly affect protein sequences, should be taken into account when studying splicing defects. Seemingly harmless exonic silent mutations may also affect mRNA splicing if, for example, the point mutation occurs in an ESE. It has been postulated that approximately 60% of all disease-causing mutations has affected splicing (Wang and Cooper, 2007). Neurofibromatosis type 1 (Ars *et al*. 1999) and ataxia telangiectasia (Teraoka *et al*. 2000) are examples of genetic disorders caused by exon skipping or cryptic splice site activation, due to exonic point mutations. These sequence variants might have been classified as nonsense or frameshift mutations, had the presence of an ESE not been considered. This highlights the importance of studying mutations not only on a genomic, but also on the RNA level (Cartegni *et al*. 2002).

5 Problem statement

Despite extensive research that has been done on pre-eclampsia, the exact cause of this disorder has not been established. Pre-eclampsia is a multi-organ syndrome and sufferers exhibit a range of symptoms and variable degrees of severity. Currently, no cure for pre-eclampsia exists and prevention of the maternal syndrome is being used to manage the disease in the developed world. PP13 has been shown to be an effective biomarker for the early prediction of pre-eclampsia. The concentration of PP13, measured in the maternal serum during the first trimester, is reduced in women who develop severe pre-eclampsia later during pregnancy when compared to healthy pregnancies (Nicolaides *et al*. 2006). Moreover, the third trimester serum PP13 levels of women either suffering from pre-eclampsia or who will later develop this disorder, are 50-70% higher than that of normal pregnancies (Burger *et al*. 2004; Than *et al*. 2007).

A diagnostic kit which measures the concentration of PP13 in the sera of pregnant women has been developed by a diagnostic company in Israel, Diagnostic Technologies Ltd. (DTL). In collaboration with this company,

this study focuses on the gene (*LGALS13*) encoding this biomarker. PP13 is expressed in the trophoblasts and on the brush border of syncytiotrophoblasts during pregnancy. The function of PP13 has not been elucidated, but it has been found that it binds to sugar residues found in placental tissue, such as mannose, *N*-acetyl-lactosamine and *N*-acetyl-glucosamine (Burger *et al*. 2004; Than *et al*. 2004). It has been postulated that PP13 plays a role during implantation in pregnancy (Burger *et al*. 2004; Than *et al*. 2004). *LGALS13* belongs to a multigene family which most likely formed via evolutionary gene duplication events.

PP13 shares a high degree of sequence homology with other galectins such as Galectin-14 and Galectin-10 (CLC). This high degree of homology might have implications for the specificity of the diagnostic kit for pre-eclampsia. If the PP13-specific antibody binds to another, highly homologous galectin, the test result will be compromised. It is thus crucial to conduct an extensive analysis of the galectin protein family, specifically the subfamily into which PP13 has been categorised. The reason for the reduction of PP13 levels in women who subsequently develop pre-eclampsia has not been found. The variable expression of this protein leads one to look at the DNA and RNA level of regulation and expression. It is possible that some of the sequence variants previously identified in *LGALS13* are responsible for the reduction in protein levels. To date, none of these variants has been associated with a risk of developing pre-eclampsia.

The above-mentioned issues are important to address in order to gain a better understanding of the factors that can be used for early diagnosis of pre-eclampsia. A broader knowledge of the function, time and level of expression of PP13 will aid in understanding its possible role in pregnancy and the development of pre-eclampsia. To gain such knowledge, it is crucial to understand the underlying biological mechanisms involved in PP13 production. It is important to gain a broader knowledge of the molecular mechanisms at work on the genome, transcriptome and proteome levels.

6 Aim and Objectives

The aim of this study was firstly to perform a comprehensive *in silico* characterisation of the gene, *LGALS13*, and secondly to initiate a large-scale project by establishing a study cohort with a comprehensive data set designed to facilitate future molecular and functional studies.

This was achieved by carrying out the following steps

1. Studying the evolutionary conservation of the subfamily of galectins to which PP13 belongs
2. Analysing the genomic and peptide sequences of *LGALS13* and PP13 to establish the location of the amino acids and corresponding nucleotide bases which are important for the functional and structural properties of the protein
3. Evaluating whether any of the known variants in *LGALS13* might influence the level of expression or splicing mechanism of this gene
4. Studying the regulation of expression of *LGALS13* at the DNA and epigenetic levels, using various tools such as phylogenetic footprinting and *in silico* analyses
5. Obtaining clinical data and samples for experimental use from subjects of a study cohort based in the Paarl region of the Western Cape of South Africa
6. Genotyping individuals within the study cohort for future correlations with RNA levels and PP13 concentrations

CHAPTER 2

MATERIALS AND METHODS

1 Bioinformatic analyses

Information on the galectin genes was gathered from the NCBI database and Ensembl Genome Browser. Information on the galectin protein family was obtained from the Uniprot/Swissprot database (http://au.expasy.org/uniprot). Identity scores for different Galectins were calculated using the Basic Local Alignment Tool (BLAST) on the NCBI web site (http://blast.ncbi.nlm.nih.gov/Blast.cgi). All information gathered from web-based sources was regularly updated. A structural representation of the PP13 peptide was generated using Swissprot. A phylogenetic tree depicting the evolutionary relationships between different galectin proteins was generated using the Mr Bayes software package (Huelsenbeck, 2001).

The expression profile of *LGALS13* in humans was obtained from the Genomics Institute of the Novartis Research Foundation (GNF 2003) web site (Su *et al*. 2002). Putative exonic splice enhancer motifs were predicted using ESEfinder v 3.0 (Cartegni *et al*. 2003). All ESE matrices were searched at there default threshold levels (set as the median of the highest score for each sequence in a set of randomly-chosen sequences of 20 nucleotides in length).

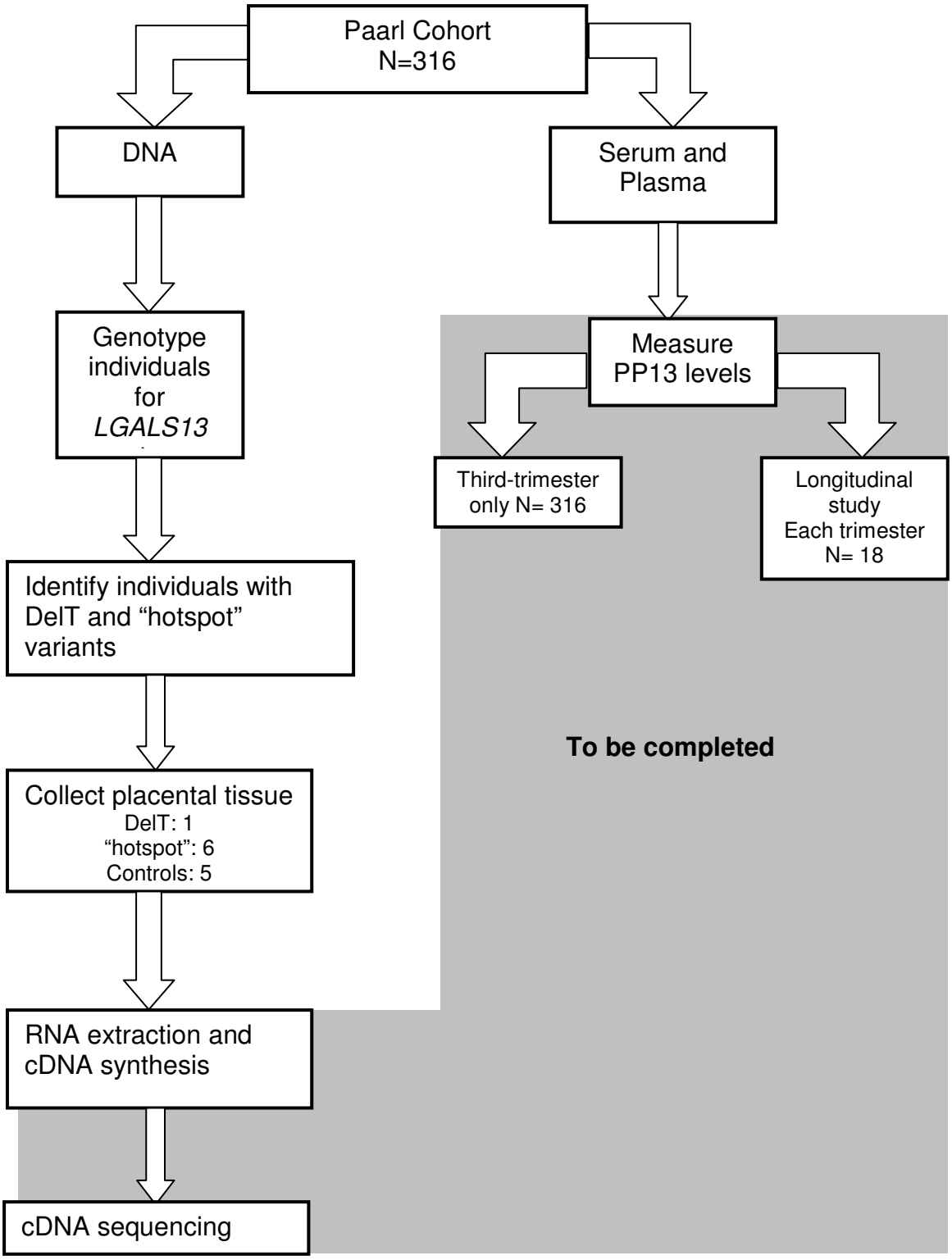
The proximal and distal promoter areas of *LGALS13* were studied using various bioinformatic tools and databases. CpGplot (European Bioinformatics Institute 2006-2008, EMBL) was used to identify CpG islands in the region upstream of the promoter area of *LGALS13.* CpG islands between 50 and 200 basepairs long were searched for within windows of 50 to 100 basepairs. The step length was also altered between 1 and 5.

MotifSampler (Thijs *et al.* 2001) and WeederH, v 1.0 (Pavesi *et al*. 2007) were used to identify putative transcription factor binding sites in the promoter area of *LGALS13*. Sequence alignments of *LGALS13* with its homologues and orthologues were made using ClustalW software in the BioEdit package (Ibis Biosciences, Carlsbad, CA). Orthologues and homologues of *LGALS13* were imported as fasta files into the MotifSampler and WeederH software

packages. The Eukaryotic Default setting was selected for *in silico* searches for regulatory elements in *LGALS13*.

2 Experimental procedures

The following graphic presentation summarises the experimental procedures used in this study and also indicates the future work related to this project.



2.1 Study Cohort

This project was approved by the Institutional Review and Ethics Board (N05/07/122) (Appendix 1). Written proof of informed consent was obtained from all adult study participants (Appendix 2). To minimise genetic heterogeneity, participants were largely restricted to the South African Coloured and Black (Xhosa speaking) populations of the Western Cape. The Coloured population of the Western Cape is an established ethnic group of South Africa with San, Khoi, Madagascan, Javanese and European ancestry (Loubser *et al*., 1999). The Xhosa population of South Africa is an innate African ethnic group, with the majority of Xhosa-speaking individuals based in the Eastern and Western Cape provinces (Statistics South Africa, 2001). The study cohort comprised 316 pregnant females and four babies from these pregnancies. Eighteen of the 316 subjects participated in a longitudinal study running parallel to this one. These subjects were required to donate blood during their regular clinical visits throughout their pregnancies. Serum was thus obtained from these subjects during each trimester of pregnancy. Sampling was mostly carried out in the Paarl region of the Western Cape, specifically at the T.C. Newman Clinic in Paarl East and Paarl Provincial Hospital.

2.2 Sample collection and preparation

For plasma collection and DNA extraction, 5-10 ml whole blood was collected in 7 ml EDTA-containing vacutainer tubes. For serum collection 5-10 ml whole blood was collected in 9 ml gel vacutainer tubes, without heparin. Serum samples were mainly collected during the third trimester of pregnancy (298 members of the study cohort). First, second and third trimester serum samples were collected from 18 of the total number of 316 study participants. A total number of 12 placental tissue samples and four cord blood samples were collected at delivery. Cord blood samples were stored in EDTA tubes. Placental samples were snap frozen in liquid nitrogen and stored in tissue collection tubes or aluminium foil at -80ºC. Placental samples were also

collected in tubes containing RNAlater (Applied Biosystems/Ambion, Austin, U.S.A.).

2.2.1 Serum and Plasma preparation

Whole blood samples, collected in either EDTA or heparin free gel vacutainer tubes, were left to stand at room temperature for 1-2 hours. The samples were subsequently centrifuged at 2500-3500rpm for 10-15 minutes whereafter the top phases were removed and pipetted into 1.5ml Eppendorf (Eppendorf, Hamburg, Germany) tubes. These samples were stored at -20ºC until required for measurement of PP13 concentration.

2.2.2 DNA extraction

Genomic DNA was extracted using a salting out technique as described by Miller *et al*. (1988) (Appendix 3) from whole blood and cord blood samples collected in EDTA tubes. Approximately 5ml of blood was used for each individual extraction. DNA was resuspended in TBE buffer, whereafter the concentration was measured using a NanoDrop™ ND-1000 Spectrophotometer (NanoDrop Technologies Inc., USA). All DNA samples were diluted to 50 ng/µl and stored at 4°C.

2.3 Mutation Screening

DNA samples from 316 pregnant females as well as four neonates were screened for previously identified mutations in exons 1 and 3 of the *LGALS13* gene.

2.3.1 PCR amplification

a) Oligonucleotide Primers

The *LGALS13* gene and LOC148003 putative gene were annotated using Locuslink on the NCBI web site (http://www.ncbi.nlm.nih.gov/) (Appendix 4).

PCR primer pairs were designed for sequence specific amplification of each exon of interest, including flanking intronic and putative promoter regions (Table 1). *LGALS13* and LOC148003 specific primers were designed to include exonic and intronic regions for discrimination between PCR products obtained from cDNA and genomic DNA templates (Table 2). Primers were analysed for hairpin, homo- and heterodimer structures using the OligoAnalyzer software available on the IDT® web site (http://www.idtdna.com) (Integrated DNA Technologies Inc., Coralville, IO, USA). Sequence specificity of each primer was verified using NCBI Basic Local Alignment Search Tool (BLAST) (http://www.ncbi.nlm.nih.gov/BLAST/). All primers were synthesised by the Department of Molecular and Cell Biology, University of Cape Town.

Table 1: Primers used for PCR amplification of exons 1 and 3 of *LGALS13*

Primer name	Sequence (5'-3')	Amplicon size (bp)	Tm (ºC)	Ta (ºC)
LGALS13-E1F (new)	GTTTCTTCCTAACAACTAAACCTG	324	52	52
LGALS13-E1R (new)	TATTCCACACCTCAATAGCTCTA		53	
LGALS13-E3.1F	GGCCATCAGTATTATCTGGGAG	213	55	50
LGALS13-E3.1R	GATTGCCAAAGTGCACTCGG		54	
LGALS13-E3F2	GGATGAGGATTCAGATATTGCC	265	53	50
LGALS13-E3R2	CCCTGACGGACTACTGAG		53	

Table 2: Primers used for specific amplification of *LGALS13* and the putative gene, LOC148003, from genomic DNA as well as cDNA

Primer name	Sequence (5'-3')	Amplicon size (bp)		Tm (ºC)	Ta (ºC)
		Genomic	cDNA		
LGALEx3F	CGAGTGCACTTTGGCAATCA	2088	255	52	47
LGALEx4R	ATTGCAGACACACACTGAGGT			52	47
RNALOCE3F	AGTGCACTTAGGCCGTCG	1979	255	53	50
RNALOCE4R	CCATTGTTGACAAGCACTGAGTC			53	50

b) PCR conditions

For amplicons smaller than 1kb, reactions were performed in a total volume of 50 µl comprising: 10 pmol each forward and reverse primer (University of

Cape Town, South Africa), 0.25 mM of each dNTP, with either 0.5 U GoTaq® DNA polymerase, 1x reaction buffer (Promega, Madison, WI, USA) or 0.5 U GoTaq® Flexi DNA Polymerase, 1x Flexi reaction buffer and 0.25-2 mM $MgCl_2$ (Promega, Madison, WI, USA). For each reaction, ~50 ng genomic DNA was used as template.

Amplicons greater than 1 kb in length were generated using Kapa LongRange® PCR reagents (Kapa Biosystems, Inc., Massachusettes, USA) according to manufacturer's specifications. ~50 ng genomic DNA was used in 25 µl total reaction volumes.

DNA amplification was performed using the GeneAmp®PCR System 2700 (Applied Biosystems, California, USA) and cycling conditions were adjusted according to the nature of each expected amplicon (Table 3).

Table 3: PCR cycling conditions used to generate each amplicon

Gene	Amplicon		Primer set		Initial denaturation temperature and time	Final extension temperature and time)	PCR Cycle			Number of cycles
	Name	Size (bp)	Forward	Reverse			Denaturation temperature and time	Annealing temperature and time	Extension temperature and time	35
LGALS13	Exon1	324	LGALS13-E1F(new)	LGALS13-E1R(new)	95 ºC, 3 min	72 ºC, 5 min	95 ºC, 20s	52ºC, 30s	72ºC, 30s	35
LGALS13	Exon3.1	213	LGALS13-E3.1F	LGALS13-E3.1R	95 ºC, 3 min	72 ºC,5 min	95ºC, 30s	50ºC, 30s	72ºC, 30s	35
LGALS13	Exon3.2	265	LGALS13-E3F2	LGALS13-E3R2	95 ºC, 3 min	72 ºC, 5 min	95ºC, 30s	50ºC, 30s	72ºC, 30s	35
LGALS13	LGALS 13	Genomic: 2088 cDNA: 255	LGALEx3F	LGALEx4R	95 ºC, 5 min	72ºC, 4 min	95ºC, 20s	46ºC, 20s	68ºC, 2 min	35
LOC148003	LOC	Genomic: 1979 cDNA: 255	RNALOCE3F	RNALOCE4R	95 ºC, 5 min	72ºC, 4 min	95 ºC, 20s	50ºC, 20s	68ºC, 2 min	35

2.3.2 Agarose gel electrophoresis

Successful amplification was confirmed via agarose gel electrophoresis of the PCR products obtained. 2-5 µl of each PCR product was added to an equal volume of DNA loading dye (95% Formamide, 20 mM EDTA. 0.05% Xylene Cyanol, 0.05% Bromophenol Blue), and resolved on a 1% agarose gel, containing Ethidium Bromide in a total concentration of 0.5 µg/ml. Electrophoresis was performed in 1x TBE buffer (90 mM Tris-HCl, 90 mM Boric acid and 1 mM EDTA, pH 8.0) for 30-60 minutes at 120 Volts. PCR products could subsequently be visualised under ultraviolet light (260 nm) using the MULTI GENIUS Bio Imaging System (Syngene, MD, USA).

2.3.3 Multiphor Single Stranded Conformational Polymorphism (SSCP) and heteroduplex analysis

Approximately 3 µl of PCR product was added to an equal volume of SSCP loading buffer (95% Formamide, 100mM NaOH, 0.25% Bromophenol Blue and 0.25% Xylencyanol) and dentatured at 95°C for 5 minutes, followed by a rapid quenching on ice for three minutes. The samples were resolved on a 12% non-denaturing polyacrylamide gel (99% Acrylamide, 1% Piperazine diacrylamide [PDA], TEMED and 10% Ammonium persulfate). Electrophoresis was performed for 90-120 minutes at 4-11ºC and 355 Volts, using a Pharmacia LKP 2117 Multiphor II Electrophoresis Unit (Pharmacia, Uppsala, Sweden). Single stranded conformational polymorphisms and heteroduplex conformation patterns were visualized via 0.1% Silver staining (Liechti-Gellati *et al*. 1999) (Appendix 6).

2.3.4 Restriction enzyme analysis

Sequence variants that created or abolished a particular restriction enzyme recognition site were identified and characterized using the BioEdit Sequence Alignment Editor (Isis Pharmaceutical, Carlsbad, California, USA). These variants were genotyped via restriction enzyme analysis according to manufacturer's recommendations (New England Biolabs Inc., Massachusettes, USA). For restriction enzyme digestion reactions, 10-15 µl of

PCR products were added to ddH_20, 1-2 units of enzyme and an appropriate buffer in a total reaction volume of 20-25 µl. All reactions were performed at 37ºC for 16 hours. A list of restriction enzymes and sites as well as the conditions used in this study is provided in Table 4. Restriction enzyme reaction products and appropriate size standard ladders were resolved on 2-3% Ethidium Bromide containing agarose gels by electrophoresis in 1x TBE buffer at 100 Volts for 2-4 hours.

Table 4: Restriction enzymes (RE) and allele description used to confirm variations identified in this study.

Amplicon	dbSNP	Genomic position and nucleotide change	Optimal reaction temperature	Recognition site 5'-3'
Exon1	rs: 3764843[1]	-98 (A/C), 5'UTR	*Ava*I (37℃)	C*YCGRG
Exon3.2	rs: 2233708[1]	IVS3+72 (T/A)	*Stu*I (37℃)	AGG*CCT

1. Bruiners, N. 2007, MSc Thesis, University of Stellenbosch

2.3.5 DNA sequencing

PCR products were purified using SureClean® (Bioline Inc., USA) and sequenced by the University of Stellenbosch Central DNA Sequencing Facility. The generated sequences were submitted to BLAST for confirmation of the identity and orientation of the amplicons obtained via PCR.

2.4 RNA extraction

RNA was extracted from placental tissue using the Invitrogen Trizol protocol (Chomczynski and Sacchi, 1987). Placental samples weighing less than 30 mg were homogenised in 600-800 µl Trizol reagent inside 2 ml Eppendorf tubes using a Polytron® Kinematica PT 2100 homogeniser (Kinematica Inc., Switzerland). Extracted RNA samples were resolved on a 1% agarose gel for assessment of the integrity and purity of these samples. All glassware and equipment were cleaned with deionized, diethylpyrocarbonate (DEPC) treated milliQ H_2O. RNA yield was quantified using a NanoDrop™ ND-1000 Spectrophotometer (NanoDrop Technologies Inc., USA).

3 Statistical Analysis

Statistical analysis was performed using the Hardy-Weinberg Equilibrium Chi^2 test to compare the genotype and allelic frequency distribution in the study groups. Genotypes and alleles were considered to be in Hardy-Weinberg equilibrium if the observed frequencies did not differ significantly from the expected ($p>0.05$). The study cohort was divided into groups with (1) good pregnancy outcome (controls) and (2) adverse pregnancy outcomes (patients) for each population group (Coloured and Black). Statistically significant associations between genotype and allele frequencies and pregnancy outcomes were determined using 2 × 2 contingency table analyses and the Chi^2 test for independence using the EpiInfo 6 statistical software package (http://www.cdc.gov/epiinfo/Epi6/ei6.htm). Odds ratio (OR) and relative risk (RR) at 95% confidence intervals (CI) were also determined to describe the strength of association. Any association demonstrated was confirmed by Mantel-Haenzel and Fisher's exact test p-values. A p-value of <0.05 was considered statistically significant.

CHAPTER 3

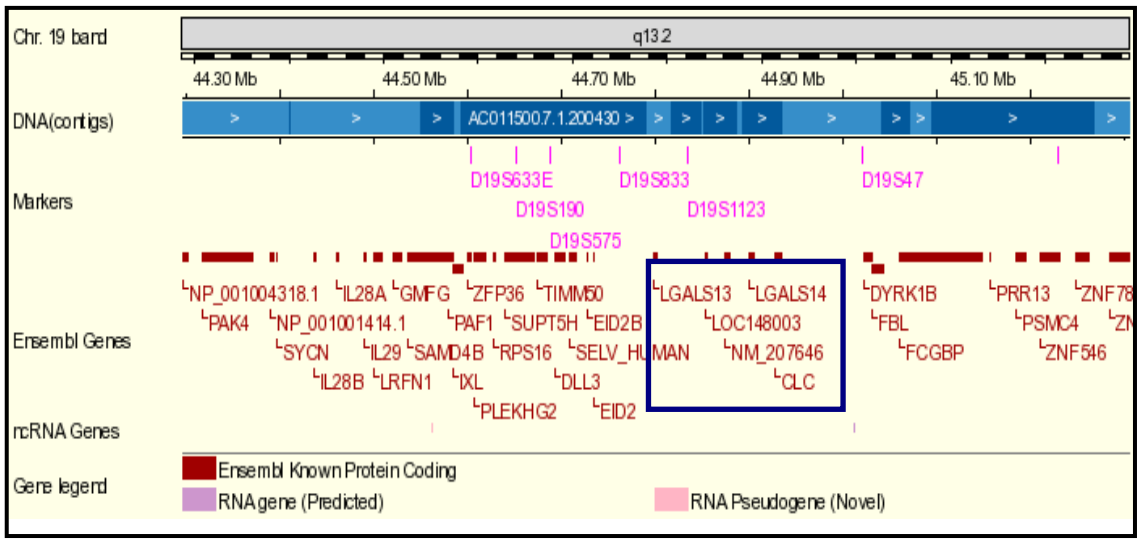
RESULTS AND DISCUSSION

1 Bioinformatic analyses

1.1 Galectin protein family

The galectin protein family plays important roles in a vast range of molecular and cellular functions (Cooper and Barondes, 1999; Visegrády *et al*. 2001; Yang *et al*. 2002). It has been postulated that a unique subfamily of galectins, which shares an affinity for mannose instead of β-galactose as carbohydrate ligand, has evolved (Cooper and Barondes, 1999). Since Placental Protein 13 (PP13) is categorised as a member of this subfamily, this section focuses on members of this subfamily of galectins, such as PP13, Galectin 10/Charcot-Leyden-Crystal protein (CLC), Galectin 14 and Galectin 7.

Figure1 shows that the gene encoding PP13, *LGALS13*, is located on the long arm of the human chromosome 19, in close proximity to the genes encoding the above-mentioned members of the galectin subfamily and several putative genes. The gene order in this chromosomal region (from the telomere towards the centromere) is: *LGALS10*, *LGALS14*, LOC400696 (AC005515-l), LOC148003, *LGALS13* and LOC390930.



Note: This image is subject to change as new information becomes available on the Ensembl database.

Figure 1: Diagrammatic representation of the long arm of human chromosome 19 according to the Ensembl database, release 50 (Ensembl Genome Browser, September 2008 http://www.ensembl.org/Homo_sapiens/geneview?gene=ENSG00000105198) A blue block is drawn around the cluster of genes belonging to the above-mentioned galectin subfamily.

In this study, the Ensembl and NCBI databases were used to investigate the gene order on chromosome 19. This order is graphically shown in Figures 1 and 2, and corresponds to that published by Cooper *et al*., which is the most recent publication of these results. The information obtained from the two databases used to investigate the genomic location of *LGALS13*, differed, which complicated this analysis. Some of the pseudogenes published on the NCBI database are not published on Ensembl, as can be seen when comparing Figures 1 and 2. Figure 2 shows that five of the six genes within the genomic region between pair number 44913735 and 44722657 are in the same orientation. This may complicate primer design for gene-specific amplification and have implications in gene conversion or mechanisms involving the generation of diversity within the Galectin protein family.

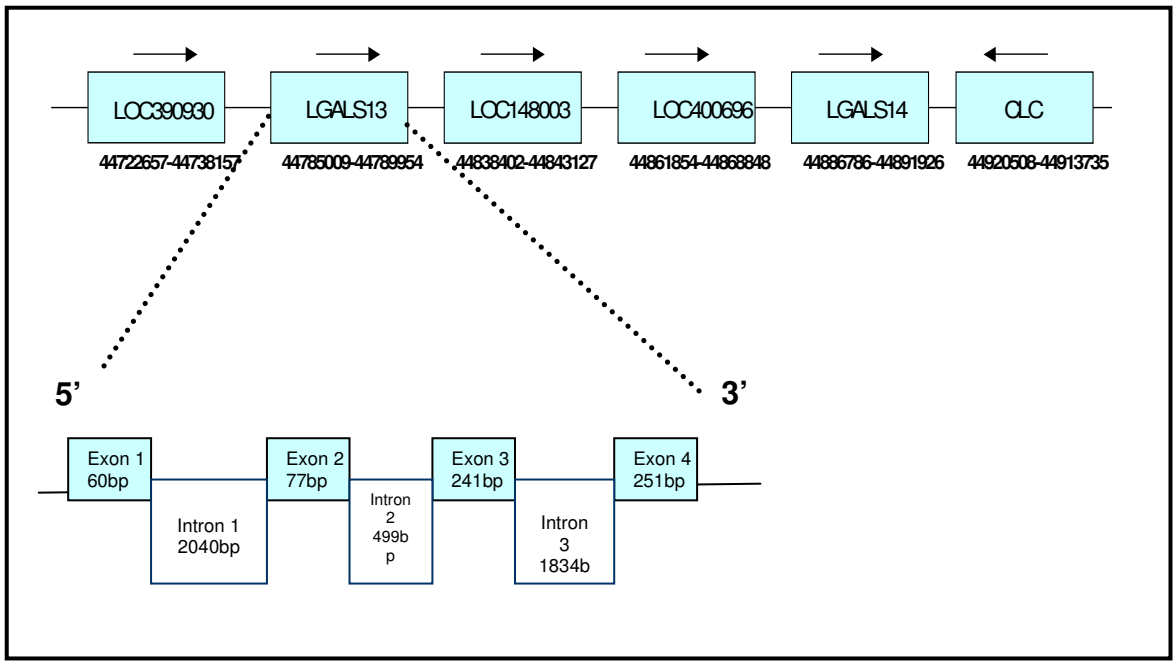


Figure 2: Order of Galectin genes on the long arm of chromosome 19, according to NCBI, September 2007. Gene orientation is indicated by arrows for each gene name. A zoomed-in view of *LGALS13* shows that it has four exons (size range 60-251bp) and spans a genomic region of 4950 bp. Intronic regions vary between 499bp and 1834bp in length (Not drawn to scale).

Table 1: *LGALS13* and other galectin family members with the highest degree of homology and structural and functional similarities

Gene	Genomic location	Protein	Synonyms	Amino Acid sequence homology to PP13	Primary Accession Number
LGALS 4	Chromosome 19: 43,984,152-43,995,425	Galectin-4	Lactose-binding lectin 4 L-36 lactose-binding protein L36LBP Antigen NY-CO-27	34% Identity*	P56470
LGALS 7	Chromosome 19: 43.97m	Galectin-7	Gal-7 HKL-14 PI7 p53-induced gene 1 protein	30% Identity	P47929
LGALS 10	Chromosome 19: 44.91m	Eosinophil lysophospholipase	Charcot-Leyden crystal protein Lysolecithin acylhydrolase CLC Galectin-10	55% Identity	Q05315
LGALS 13	Chromosome 19: 44.79m	Galactoside-binding soluble lectin 13	Placental tissue protein 13 Placenta protein 13 PP13 Galectin-13	100% Identity	Q9UHV8
LGALS 14	Chromosome 19: 44.89m	Placental protein 13-like	Charcot-Leyden crystal protein 2 CLC2 Galectin-14	68% Identity	Q8TCE9

Source: Uniprot/SWISSPROT (http://au.expasy.org/uniprot)
*: Identity scores calculated using the Basic Local Alignment Tool (BLAST) on the NCBI web site (http://blast.ncbi.nlm.nih.gov/Blast.cgi)

The high degree of sequence homology in the galectin multigene family complicates gene-specific PCR amplification. According to the identity scores calculated in this study, the galectin subfamily member sharing the most sequence homology with PP13 is Galectin 14 (68% identity, Table 1). However, the predicted protein product of the putative gene, LOC148003, is 75% identical to PP13 (Table 2). Another homologue of *LGALS13, LGALS10*, is also located on the long arm of chromosome 19 (Figure 1) and encodes the Charcot Leyden Crystal (CLC) protein. These two proteins are 55% identical (Table 1) and have been reported to exhibit similar secondary structures (Visegrády *et al*. 2001). Table 1 shows that PP13 is the least homologous to

Galectin-7 (30% identity) and Galectin-4 (34% identity). The genes encoding these proteins, *LGALS7* and *LGALS4*, are located further along chromosome 19, closer to the centromere, and are not shown in Figures 1 and 2.

In this study, GNF SymAtlas v 1.2.4 (http://symatlas.gnf.org/SymAtlas/) was used to investigate the tissue-specific expression of the different galectins belonging to the subgroup of the above-mentioned galectin protein family. According to bioinformatic analyses done in this study, *LGALS13* and *LGALS14* are mainly expressed in the human placenta while *LGALS10* is expressed in whole blood, bone marrow and CD 34+ cells of humans (Figure 6 and Supplementary data).These findings correlate to those of Cooper, where it was also stated that *LGALS10* is expressed in eosinophils and basophils (Cooper 2002). The differences in expression patterns within the galectin subfamily could be a consequence of the levels of conservation and functional diversification in the protein family. PP13 and Galectin-14 exhibit similar expression patterns and share greater homology than PP13 and Galectin-10. This could indicate that PP13 and Galectin-14 have functional similarities. The expression pattern of *LGALS13* will be further discussed in section 1.5.1.

1.2 Conservation of the galectin protein family

The abundance of amino acids involved in the structural and functional properties of PP13 is most likely the reason for the high degree of conservation of this region, both between different galectin genes (Figure 3 and Table 1) and among *LGALS13* orthologues from different primate species (Table 2 and Supplementary data). Homologous genes existing in close proximity of one another on the same chromosome and which share certain functions often derive from an evolutionary gene duplication event (Yang *et al*. 2002). This is most likely the case with the cluster of galectin genes on human chromosome 19.

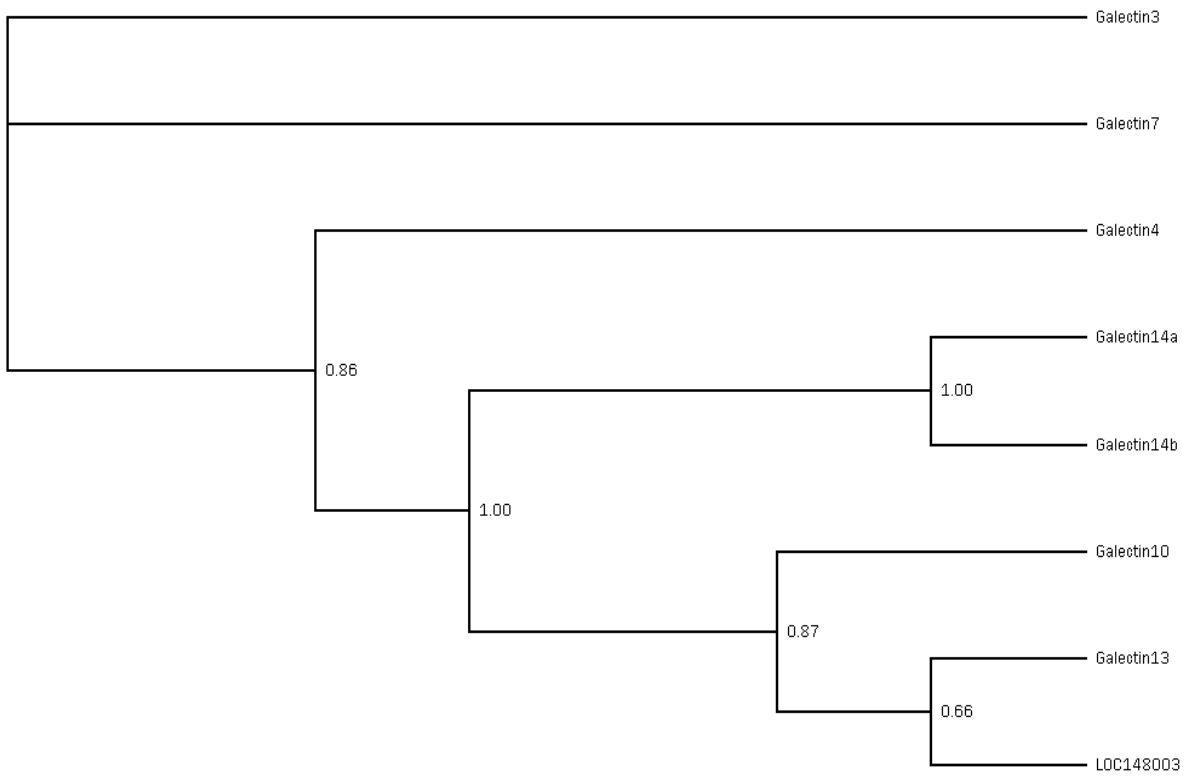


Figure 3: Maximum likelihood tree (cladogram with fixed increases in branch length) depicting the evolutionary relatedness of the genes within the subgroup of galectins situated on chromosome 19. This phylogenetic tree was generated using MrBayes v.3 software (Huelsenbeck *et al*. 2001).

The evolutionary relatedness of the different galectin genes is indicated in Figure 3. The maximum likelihood tree shown supports the data in Tables 1 and 2. The bootstrap values given on some of the branches of this tree indicate the level of confidence that this tree could be reproduced. A bootstrap value of 1 indicates that the data generated supports the particular branching at that point in the tree in 100% of the bootstrap replicates obtained. Galectins 3 and -7 are the outgroups in this phylogenetic tree. Galectin 3 is not located on the same chromosomal region as the other galectins in the tree. Of the galectin genes located on chromosome 19, Galectin 7 is the least identical to PP13 (30%). According to this phylogenetic tree, the diversion of Galectins 10 and 13, as well as LOC148003, from Galectin 14 is supported in 100% of the bootstrap replicates of this tree. Galectin 10 diverged from Galectin 13 and LOC148003 at this point in the tree in 87% of the bootstrap replicates obtained by this analysis. Galectin 13 and LOC148003 were found to be most

closely related, with their branches separating at this point in the tree in 66% of the replicates. Branching within this cladogram and clustering of galectins observed in Figure 3 are interesting since Galectin 10 seems to be more closely related to Galectin 13 than Galectin 14. Table 1 showed that, on peptide level, Galectin 14 is more silimar to Galectin 13 (68%) than Galectin 10 is (55%). These discrepencies might explain the relatively low bootstrap values given on the branches separating Galectins 10 and 13 in the phylogenetic tree. The differences might also be due to differences in DNA and protein conservation. Although the Galectin 13 and 14 peptides are more similar, it is possible that the non-coding regions within the genes are more conserved between Galectins 10 and 13. This might have implications for the regulation and expression of these genes, since the conservation of untranslated regions are often associated with similar expression patterns in homologues (Pennacchio and Rubin, 2001). Evidence that Galectins 13 and 14 share a similar expression pattern in humans exists (Section 1.5).

Table 2: Identity scores (in %) of PP13 from primates in comparison to two other galectin peptide sequences, CLC and LOC148003, the DNA sequences of which are found in close proximity to *LGALS13* on human chromosome 19 (Protein level comparison).

	PP13 Human	**PP13 Chimp**	**PP13 Rhesus monkey**	**LOC148003**	**CLC**
PP13 Human	100%	98%	97%	75%	55%
PP13 Chimpanzee	98%	100%	97%	76%	55%
PP13 Rhesus monkey	97%	97%	100%	76%	53%
LOC148003	75%	76%	76%	100%	52%
CLC	55%	55%	53%	52%	100%

It has previously been reported that PP13 is most homologous to CLC. However, the results in Figure 5 and Table 2 show that the putative protein encoded by LOC148003 would be the most homologous to PP13, in the event that it is expressed. These proteins are also the most closely related of the subgroup of galectins under investigation (Figure 3).

Table 2 shows that PP13 is highly conserved in primates, the human peptide being 98% identical to the Chimpanzee orthologue and 97% identical to the Rhesus monkey orthologue. PP13 has, to date, not been identified in any eutherian species other than the human, chimpanzee, orang-utan and rhesus monkey. All of the primate genomes have not been fully sequenced and would be helpful to investigate the evolution and function of this protein. This protein most likely has an important role in the development and sustenance of the primate placenta during pregnancy. Furthermore, the putative protein of LOC148003 exhibits a slightly greater homology to the Chimpanzee and Rhesus monkey PP13 orthologues than that of the human protein. It would be interesting to investigate whether Chimpanzee and Rhesus monkey orthologues of the LOC148003 putative gene exist. This could be informative of conservation of this putative gene, within the galectin protein family as well as within each species. It might also give an indication of the evolution within this protein family.

1.3 Predicted gene-conversion event

Gene conversion occurs when a section of DNA is transferred from one strand to another. This conversion is non-reciprocal, and the strand from which the section of DNA originates remains unchanged. Such events can occur during base mismatch repair in recombination and may lead to non-Mendelian inheritance. Gene conversion contributes to the exchange of genetic material between homologues and is an important aspect of the evolution of multigene families (Posada *et al*. 2002).

One of the DNA sequence polymorphisms present in *LGALS13*, results in a six base pair change within the 11bp region between base pair numbers 137 and 148 in exon 3. This polymorphism has been named the “hotspot mutation”, as it was originally thought that the above-mentioned genomic region had a relatively high mutation rate. It was, however, later found that this sequence variant is 100% identical to the corresponding genomic region of a predicted gene, LOC148003 (Figures 1 and 2), located downstream from *LGALS13* (LOC148003 is classified as a predicted gene on the NCBI

database). One explanation for this observation is that the so-called “hotspot mutation” is the result of an evolutionary gene conversion event between *LGALS13* and LOC148003.

The multiple sequence alignment in Figure 4 demonstrates the close evolutionary relationship between the putative gene and the primate *LGALS13* orthologues in exon 3. Figure 4b focuses on the genomic region which could hypothetically have been exchanged during the evolutionary gene-conversion event. This deduction was made after analyses of the sequence alignment revealed a block of 100% sequence identity between *LGALS13* and LOC148003. This region is located between base pair positions 107 and 177 of exon 3.

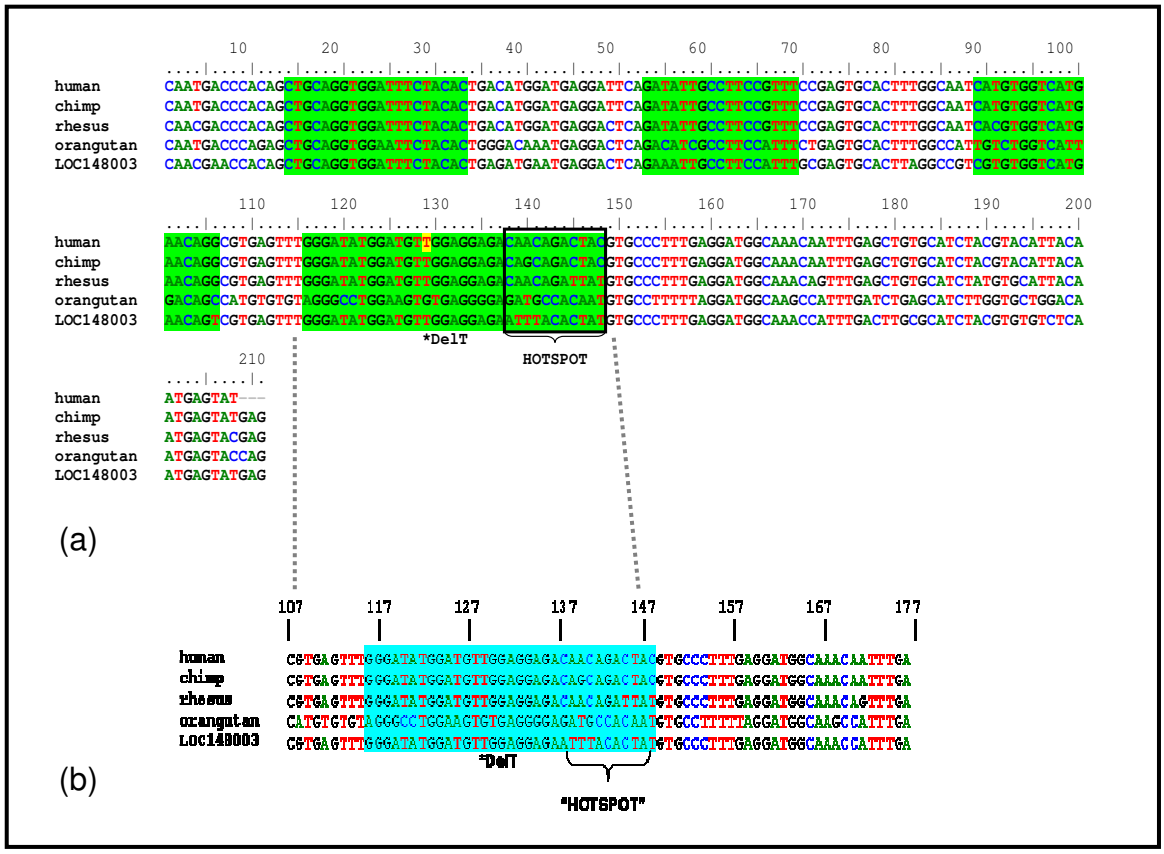


Figure 4 (a): Multiple sequence alignment of the highly conserved exon 3 in *LGALS13* from human, chimpanzee, rhesus monkey, orang-utan as well as the corresponding region in the putative human gene, LOC148003. Areas highlighted in green depict the conserved sequence regions which are important for the secondary structure of PP13. The 229delT (DelT on the figure) and “hotspot” sequence variants are indicated by their names below their positions within the sequence. The hotspot sequence shown here is the

wild type. The mutant sequence is identical to the corresponding region of the LOC148003 sequence. del229T is a sequence variant which has been identified in the human *LGALS13* sequence (Bruiners, 2007) and will be further discussed in Section 2. **(b)** A zoomed-in view of the region between base pair numbers 107 and 177. These base pair numbers represent base pair positions within the exon. The region highlighted in blue is believed to have undergone an evolutionary gene conversion event between *LGALS13* and LOC148003. The so-called "Hotspot mutation" is indicated between base pair numbers 138 and 148. The highlighted region depicts a part of the sequence encoding one of the beta sheets which are responsible for the formation of the carbohydrate recognition domain of the related protein, PP13. A high level of conservation between the *LGALS13* and LOC148003 sequences is evident from this alignment and indicates the positions where an evolutionary crossover event might have occurred

LGALS13 and LOC148003 exhibit 75% sequence identity (Table 2). The high degree of homology between these two DNA sequences, as well as their close proximity to one another on chromosome 19 (Figures 1 and 2), supports the theory that an evolutionary gene conversion event might have resulted in the presence of the so-called "hotspot mutation" in the *LGALS13* gene. As evident from Figure 5, exons 2, 3 and 4 are highly conserved in *LGALS13* and LOC148003, with 95%, 86% and 89% identity respectively shared between these homologues.

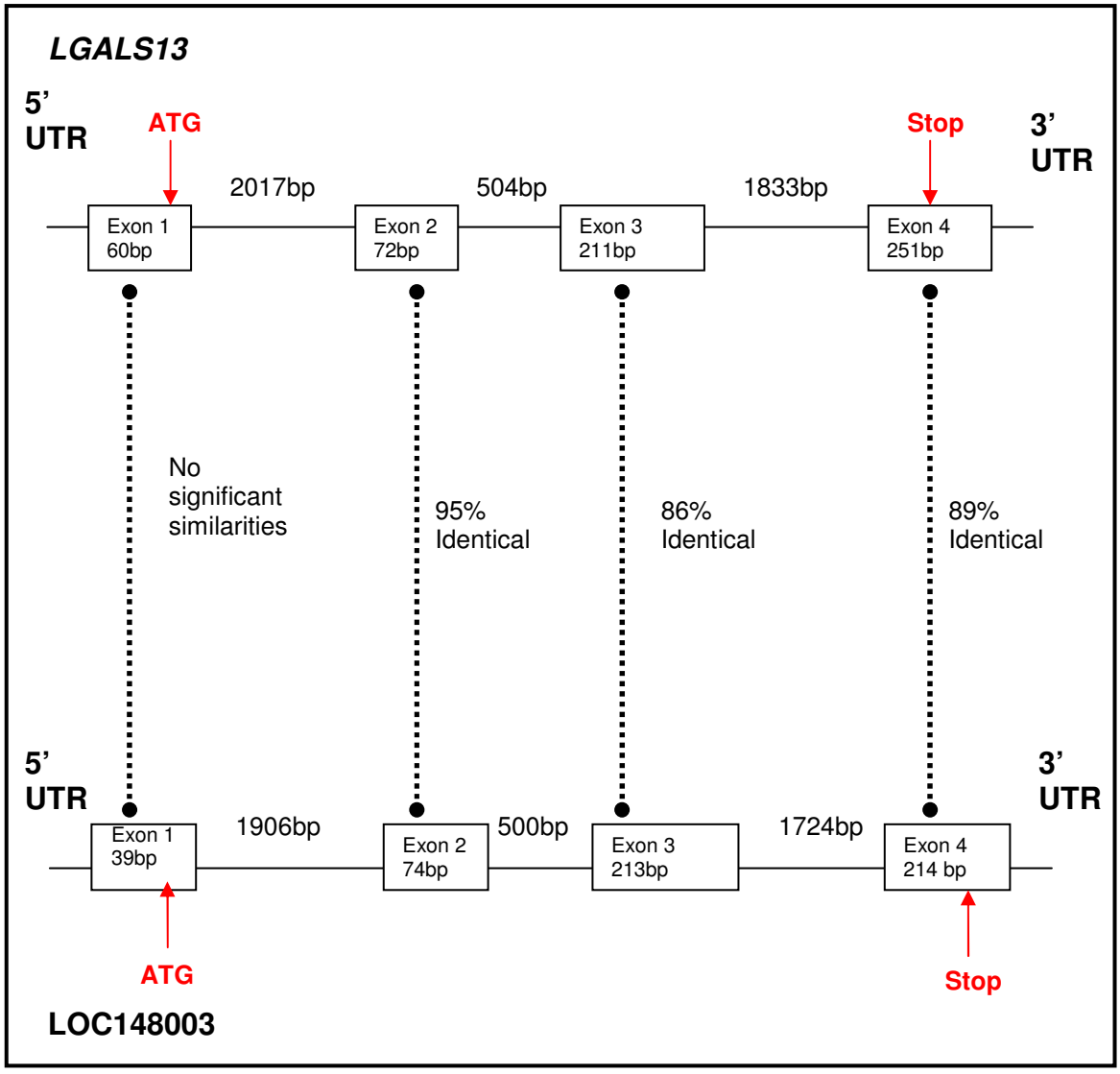


Figure 5: Comparison of *LGALS13* and LOC148003, with regard to the genetic architecture, exonic structure and homology.

The high level of conservation of exons 2-4 (Figure 5) might be due to the fact that all of the amino acids important for the functional and structural properties of the protein product are encoded by these exons (Figure 6). On the protein level, the Cysteine residues which are predicted to be involved in disulfide bonding in PP13 (Cooper, 2002), are not conserved in the LOC148003 peptide (Figure 6). Only Cys 19 is present in the LOC148003 peptide. Cys 92 is replaced by an Arginine, Cys 136 by a Leucine and Cys138 by an Asparagine residue (Figure 6). This might have significant implications for the secondary and tertiary structures of this predicted protein. The SWISSMODEL

protein modelling server was used to generate an image of the putative tertiary structure of the predicted LOC148003 peptide (Figure 7b). This predicted structure is similar to the predicted PP13 model, with 11 antiparallel β-sheets (Figure 7a). The high level of homology between the putative LOC148003 peptide and PP13 implies that these proteins might have similar functions. If this protein is expressed, it might be functionally redundant. However, the subtle differences on peptide level might suggest unique functional and structural properties.

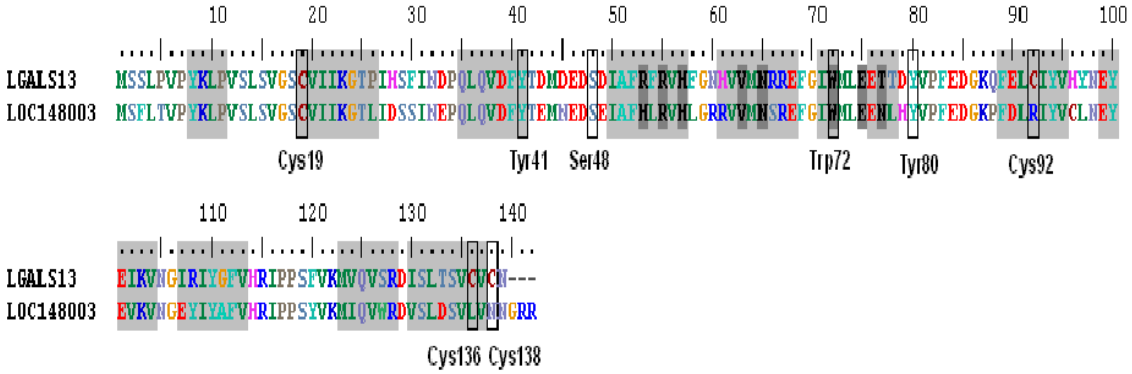


Figure 6: Alignment of the peptides encoded by the *LGALS13* and the putative gene, LOC148003. Highly conserved amino acids are indicated by grey shading. Amino acid residues possibly involved in disulfide bonding and phosphorylation are indicated by black boxes with their amino acid name abbreviations and positions below each box. Amino acids which putatively are important for specific binding of carbohydrate ligands, and form part of the carbohydrate recognition domain, are indicated by dark grey shading and black typing.

The fact that PP13 is used as a predictive biomarker for the early diagnosis of pre-eclampsia validates analyses of its homologues. It is very important that the diagnostic kit used to measure PP13 levels should be sensitive to PP13 alone, and none of its homologues. If LOC148003 is expressed as a protein in humans, the PP13 specific antibody in the ELISA kit might bind to this 'new' addition to the galectin protein family, which would result in the inaccurate measurement of PP13 levels.

Despite the high level of conservation and homology between the predicted gene, LOC148003, and the other galectins, it has not been fully classified as a

galectin. LOC148003 is also not recorded on the Ensembl database. The reference sequence for LOC148003, available on the NCBI web site, was predicted using computational software (GNMOM). Two mRNA transcripts (XM_086001.8 and XM_001721341.1) are documented on the NCBI web site. These predicted transcripts exhibit a high degree of sequence similarity to previously documented ESTs available on the NCBI web site, which supports the existence of these predicted sequences (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=retrieve&dopt=full_report&list_uids=148003&log$=databasead&logdbfrom=nuccore). The similarity in exonic structure between LOC148003 and *LGALS13*, the EST data and fact that there is no evidence to prove that LOC148003 is not expressed as a functional protein, suggest that this could possibly not be a pseudogene.

It is not often that six base pairs in an eleven-base pair region within a coding region of gene are simultaneously mutated, even if these mutations always occurred together due to linkage disequilibrium. The gene-conversion hypothesis provides a sound solution for this observation. Further experimental work could be performed to analyse and substantiate the gene-conversion hypothesis. The gene-specific amplification of *LGALS13* and LOC148003 fragments from DNA and cDNA templates obtained from individuals with and without the so-called “hotspot” variant would be informative. These fragments could be sequenced to confirm the presence or absence of the variant. The presence of the “hotspot” variant in the mRNA of individuals bearing the mutation would imply that the variant is expressed. It would also be interesting to confirm that the protein product (with a three amino acid change) is functional. Gene-specific primers for the amplification of *LGALS13* and LOC148003 were designed for this purpose. These primers were designed to include intronic and exonic regions of these genes to enable amplification from DNA and cDNA templates, as well as discrimination (based on PCR product size differences) between products from different templates.

These experiments were initiated in this project but could unfortunately not be completed due to time constraints. The putative gene, LOC148003, was

successfully amplified from genomic DNA. Total RNA was also extracted from placental tissue (Supplementary data).

1.4 Structural and functional classification of galectins

Figure 7a shows the predicted tertiary structure of one subunit of the PP13 homodimer. This image was generated using the SWISSMODEL protein modelling server. This structure correlates with that described in the literature, 11 antiparallel β-sheets can be distinguished and are numbered on the figure. The secondary structure of PP13 is similar to that of CLC, both proteins fold into five- and six-stranded β-sheets as well as two α-helices (Visegrády *et al*. 2001).

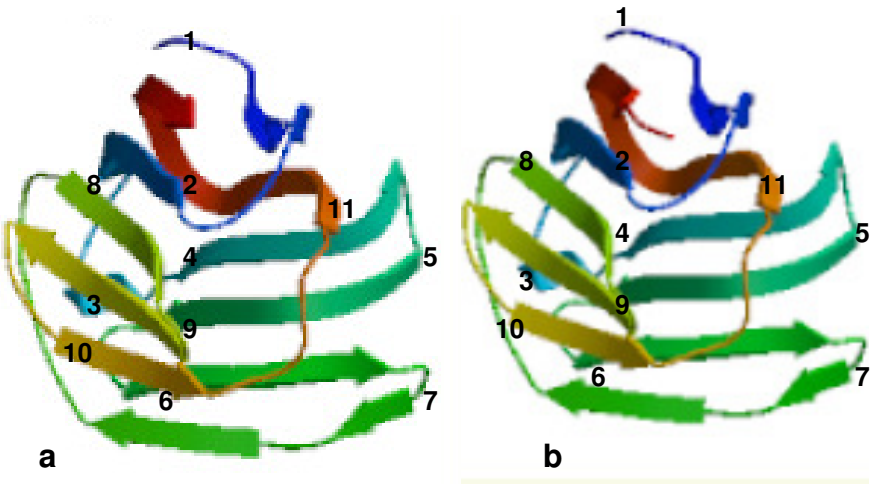


Figure 7: Three-dimensional predicted model of **a)** PP13 **b)** and the peptide encoded by the predicted gene, LOC148003. The 11 anti-parallel β-sheets are indicated by numbers 1-11 (SwissProt: swissmodel.expasy.org/workspace).

Figure 8 indicates that exons 2 to 4 encode the amino acid residues which form part of the galectin-specific carbohydrate recognition domain. This galectin-specific domain is usually 130 amino acids in length and encoded by one or two exons. This region is functionally important and thus highly conserved. Figure 8b indicates the positions of amino acid residues which are putatively involved in phosphorylation of PP13. NetPhos 2.0 was used to predict the location of possible phosphorylation sites within PP13 (Blom *et al*. 1999) This software utilises neural networks to predict the location of

phosphorylation sites and gives a prediction score between 0 and 1 for each site. The threshold was set to 0.5 for this analysis, meaning that scores above 0.5 were significant and indicated the presence of a predicted phosphorylation site. All of the residues indicated had scores > 0.5 (Supplementary data). It has previously been reported that Serine 48, Tyrosine 41 and Tyrosine 80 are involved in phosphorylation of PP13. This study has found that Serine 13, 29 and 131 as well as Threonine 78 are also possibly phosphorylated in this peptide.

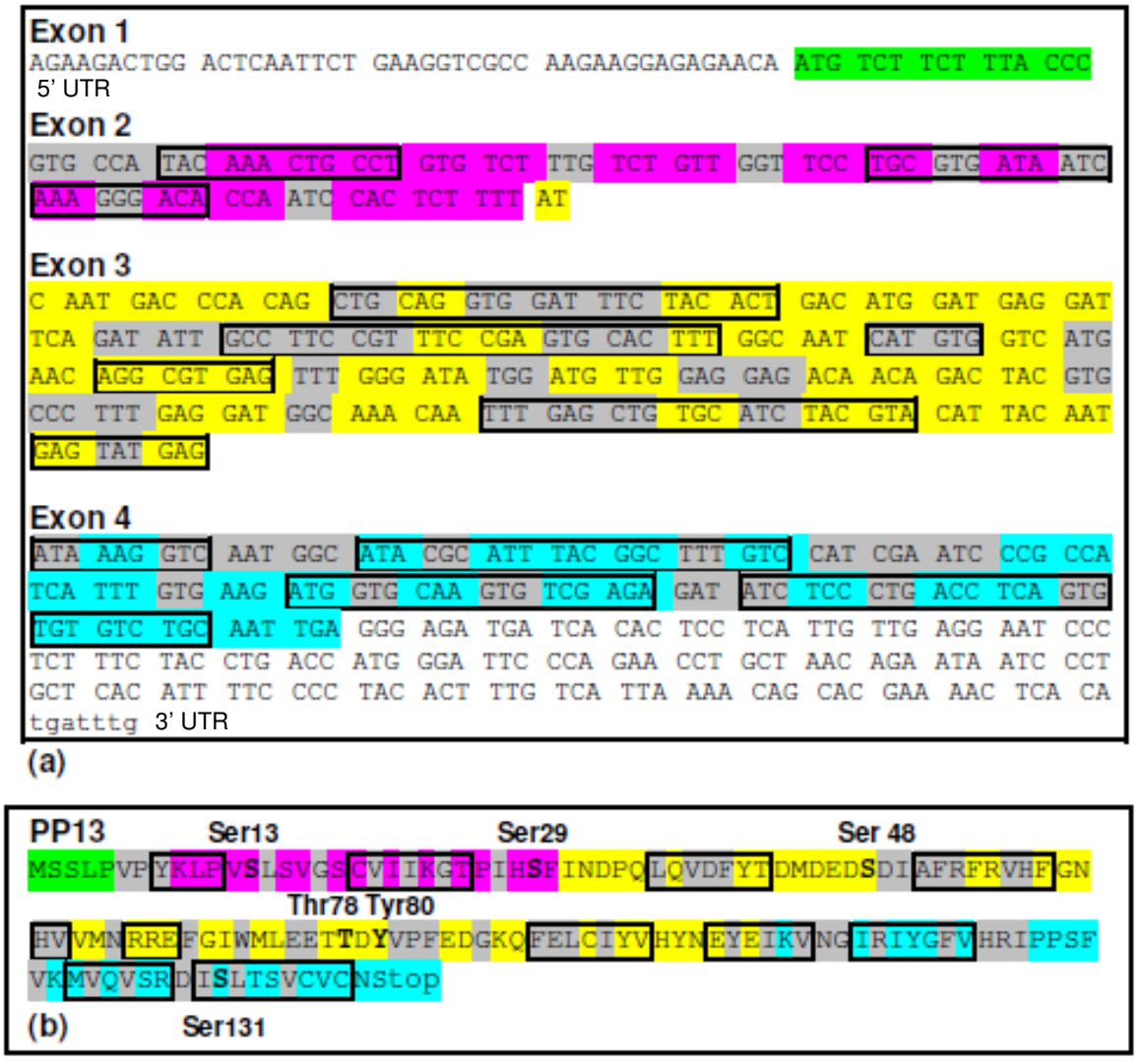


Figure 8: (a) Breakdown of exonic structure within the nucleotide sequence of *LGALS13* and **(b)** the corresponding amino acids of the PP13 peptide, as encoded by each codon. Highly conserved, galectin protein family-specific regions within the nucleotide and amino acid sequences of *LGALS13* are indicated by grey shading in a and b. Amino acid residues that are predicted as being involved in phosphorylation are indicated by bold typing in b. The amino acid residues, folded into the eleven beta sheets that make up the

secondary structure of PP13, are indicated by black boxes (a and b). The 5' and 3' UTR's (unshaded nucleotides) are indicated in a.

Although a small number of residues (the core region of the CRD) are conserved in all galectins, other residues within the CRD might be conserved within certain species or subgroups of galectins. The carbohydrate recognition domains of PP13 and CLC are very similar due to the substitution of three of the seven amino acids which are highly conserved among most galectins (Yang *et al*. 2002). Such evolutionary processes could result in functional diversification within large protein families such as the galectins. The amino acid substitution in CLC and PP13 results in a loss of L-galactoside binding activity. However, these galectin relatives show specificity to other sugar residues which might give them novel functions (Cooper and Barondes, 1999; Yang *et al*. 2002). Functional diversification might also entail differential expression patterns of proteins. Such differences are usually detected and predicted with greater difficulty than those involved in tertiary structure. Expression patterns of genes are governed by the regulatory regions, usually located in the promoter area. This will be further discussed in section 1.5.3b.

1.5 *LGALS13* expression

1.5.1 *LGALS13* expression profile in humans

From Figure 9, it can be inferred that *LGALS13* is mainly expressed in the placenta in humans. This figure depicts *LGALS13* expression levels over a range of organs and tissue types, such as the hypothalamus, adrenal gland, ovary, uterus, heart, as well as the fetal liver and lungs.

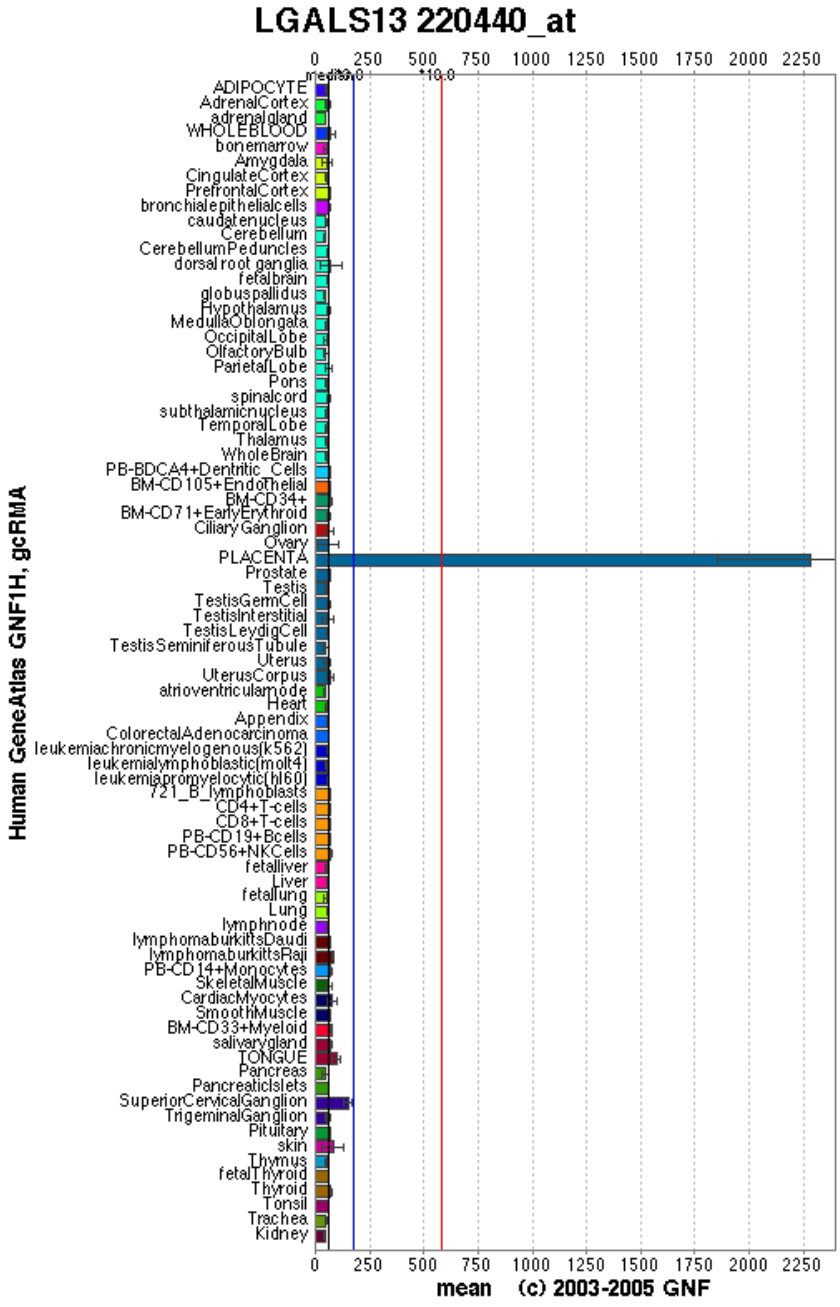


Figure 9: Expression profile of PP13 in a range of human tissues. (Source: Genomics Institute of the Novartis Research Foundation (GNF), 2003, Su *et al*. 2002).The X-axis represents the mean level of expression of *LGALS13* and the Y-axis represents different human tissue types. The black line is drawn at the mean value of *LGALS13* expression (M) measured across all tissues (M ≈ 60), the blue line represents a level of expression which is 3-fold greater than M (3M) while the red line represents an expression level 10-fold greater than M (10M).

Figure 9 depicts the expression profile of PP13 in numerous tissue types. In humans, *LGALS13* is almost exclusively expressed in the placenta. This

expression pattern restricts access to *LGALS13* mRNA, thereby complicating studying the temporal expression pattern of this gene. The fact that placental tissue can only be sampled after delivery makes it virtually impossible to assess whether *LGALS13* is expressed at differential levels throughout pregnancy. PP13 levels have been shown to differ, depending on the gestation of a pregnant woman. The placental-specific expression pattern of *LGALS13* inhibits investigations into the cause for these differential protein levels. It is possible that *LGALS13* is expressed at differential levels throughout pregnancy. It is also possible that PP13 is expressed at a constant rate and that the differences observed in serum PP13 levels could be attributed to the transport of the peptide, from the placenta into the maternal circulation. The expression profiles of two *LGALS13* homologues (Galectins 10 and 14) have also been studied (Figures 4 and 5 in Supplementary material). Galectin 10 has been found to be expressed in a wider range of tissues, with the greatest level of expression in the bonemarrow, wholeblood and CD34+ cells in humans. Galectin 14 exhibits a similar expression pattern to Galectin 13, with exculsive expression in the placenta. This is interesting due to the conflicting results from protein and DNA level homology studies as discussed previously (Section 1.1).

1.5.2 *LGALS13* transcripts and isoforms

To date, no human PP13 isoforms have been documented on the NCBI web site (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=retrieve&dopt=full_report&list_uids=29124&log$=databasead&logdbfrom=nucest, December 2008). Two *LGALS13* transcripts (ENST 00000221797 and ENST00000392053) have been published on the Ensembl Genome Browser (http://www.ensembl.org/Homo_sapiens/Transcript/ProteinSummary?db=core;g=ENSG00000105198;r=19:44784556-44789953;t=ENST00000392053, December 2008). These transcripts are similar in that both have four exons and encode proteins of 139 amino acids in length. The sizes of the first exons differ between these two transcripts. Exon 1 of ENST00000392053 is 15 base pairs in length while the corresponding exon in ENST00000221797 is 60 base

pairs long. The exons also differ considerably on DNA sequence level. This difference causes the ENST00000392053 transcript to be 554 base pairs long while ENST00000221797 is 599 base pairs in total length. These two transcripts thus exhibit the use of alternative transcription start sites in *LGALS13*. Exons two to four are similar in length and sequence in both transcripts.

The above-mentioned transcripts have not been documented by any other genomic or gene-related database. The conflicting information given by different databases, concerning transcripts and alternatively spliced isoforms of *LGALS13* and PP13, complicates the characterisation of this gene. To shed more light on the above-mentioned observations, the genomic sequence of *LGALS13* was analysed for the presence of splice elements. This was achieved using the ESEfinder program (Cartegni *et al*. 2003).

Although often neglected, the coding regions of genes should also be analysed for the presence of DNA sequence motifs with splice factor recognition capabilities. In this study, ESEfinder was used and predicted the presence of 33 such sequence motifs within exon 3. These results are summarised in Table 3. These results are of interest due to the location of the DNA sequence polymorphisms focused on in this study. Other exons were also analysed for such motifs but produced no significant results. All the splice factors listed in Table 3 belong to the Serine/Arginine-rich (SR-rich) family of proteins. These factors specifically recognise exonic splice enhancer elements and aid in the activation of weak splice sites during RNA processing (Baralle and Baralle, 2005; Cartegni *et al*. 2002). This has implications for alternative splicing. Although none of the identified ESEs are disrupted by the DNA sequence variants in exon 3, the presence and exact location of these sites are merely predicted. Further experimental procedures would have to be conducted to confirm the presence and function of these motifs in *LGALS13*. One method of achieving this is the minigene-system (Cooper, 2005). This assay involves the construction of minigenes containing intronic regions with sequence variants putatively involved in alternative splicing. A minigene construct is designed to contain the genomic region of interest. Transient

transfection of an appropriate cell line with this construct enables the researcher to assess whether alternatively spliced mRNA is produced due to the presence of a sequence variant (Cooper, 2005).

Table 3: Summary of the exonic splice-enhancing motifs predicted to be present in the *LGALS13* gene sequence with the trans-acting factors which recognise these elements.

Splice factor	Motifs *present in LGALS13*
SF2/ASF	CACAGCT
	TGCAGGT
	CTGACAT
	ATGAGGA
	CAGGCGT
	TGGAGGA
	GAGACAA
	CAGACTA
SC35	GACCCACA
	AGCTGCAG
	GATTTCTA
	GGATTCAG
	GCCTTCCG
	GGATGTTG
SRp40	CCACAGC
	CTGCAGG
	TTTCTAC
	CTACACT
	TGACATG
	AGACAAC
	AGACTAC
	TTTGAGG
	TTACAAT
SRp55	CACAGC
	TGCAGG
	TGAGGA
	TATGGA
	TACGTG
	TGAGGA
	TGCATC
	TACGTA
	TGAGTA

Data generated using ESEfinder v 3.0: http://rulai.cshl.edu/tools/ESE (Cartegni *et al*. 2003)

1.5.3 Regulation of expression

a) Epigenetic regulation of expression

It is important to study the methylation status of the promoter region of a gene since this epigenetic regulatory mechanism plays a pivotal role in the regulation of gene expression (Fang *et al*. 2006). Some genes have clusters of CG nucleotide pairs occurring at a frequency of approximately 10-fold higher than the rest of the genome. Such pairs are named CpG islands and

the abundance of CpG islands in the promoter area of a specific gene can indicate whether such a gene is epigenetically regulated via methylation. In this study, the promoter region of *LGALS13* was analysed for the presence of CpG islands.

Figures 10 and 11 show the results obtained from the prediction software, CpGplot, used for the *in silico* prediction of CpG islands in this study (http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html). This program estimates CpG content and island location using three sets of parameters, namely: percentage GC content, the ratio of observed CpG islands to expected CpG islands and the length of the islands. The output of the program is a graphical representation of the distribution and borders of predicted CpG islands in the sequence being analysed. Three graphs are produced for each sequence analysis, with the base pair position (3'-5') given on the x-axes and the above-mentioned parameters represented by the y-axes. Figure 10 clearly shows that the *LGALS13* homologue, *LGALS3*, has a CpG island stretching between 50 and 1250 base pairs upstream from the gene (base pair positions -50 and -1250). Figure 11 shows that *LGALS13* lacks predicted CpG islands, from which one can infer that this gene is not regulated on an epigenetic level via methylation of the proximal promoter region. It is, however, possible that CpG islands are present further upstream from the gene.

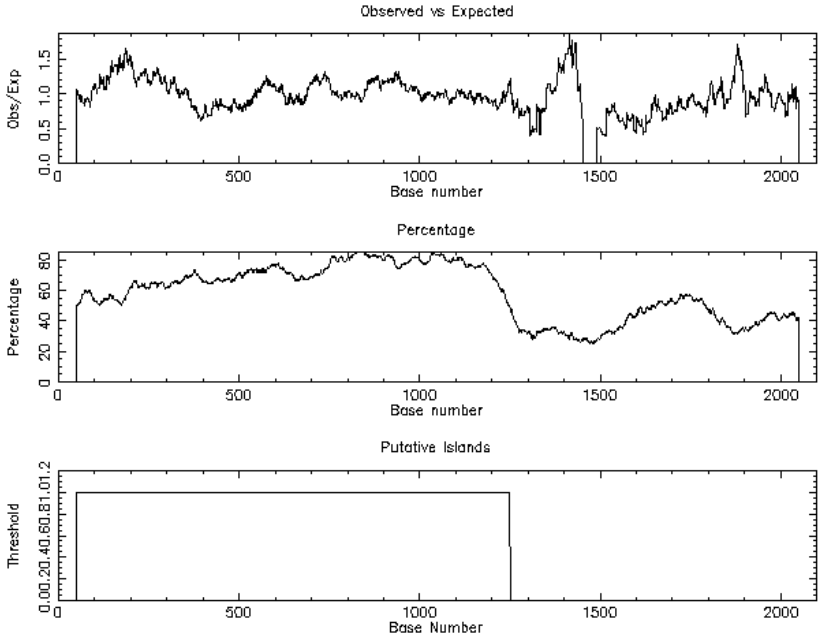


Figure 10: Results obtained via CpGplot, using the putative promoter region (2000 base pairs upstream) of the *LGALS3* gene sequence. The figure clearly indicates a significant cluster of CpG islands around positions –50 and –1250 upstream of the gene. The Observed/Expected ratio of this result is > 0.60.

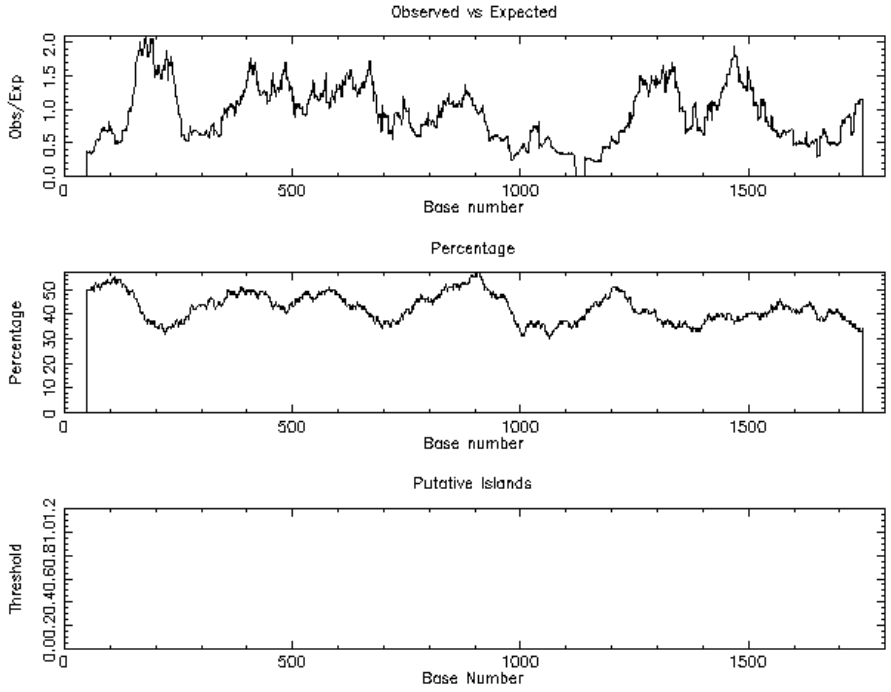


Figure 11: CpGplot results for the putative promoter region (1500 base pairs upstream) of *LGALS13*. This figure indicates a lack of CpG islands in this region of the gene.

The detection of a CpG island in *LGALS3* (Figure 10) and not in *LGALS13* (Figure 11), suggests that different galectins are uniquely regulated. Different modes of regulation might exist in this gene family due to tissue-specific expression. One study, using *in silico* based techniques, reported the existence of CpG islands, at varying densities, in the promoter areas of Galectins 1-4, 7-10, 12 and 13 (Chiarotti *et al*. 2004). According to these results, Galectins 1, 3 and 8 had the highest CpG island densities. These results were generated using the program, GeneWorks by IntelliGenetics Inc. Chiarotti *et al*. also reported that the up - and down regulation of Galectins 1, 3, 7 and 8 are associated with tumour formation. Incorrect control of genes via methylation often plays an important role in the development of cancer (Fang *et al*. 2006). This supports the hypothesis that methylation is involved in tissue-specific expression of certain galectins. More experimental data is, however, required to substantiate these *in silico*-based findings. To date only Galectin-1 has, however, been experimentally shown to be epigenetically regulated via methylation (Benvenuto *et al*. 1996)

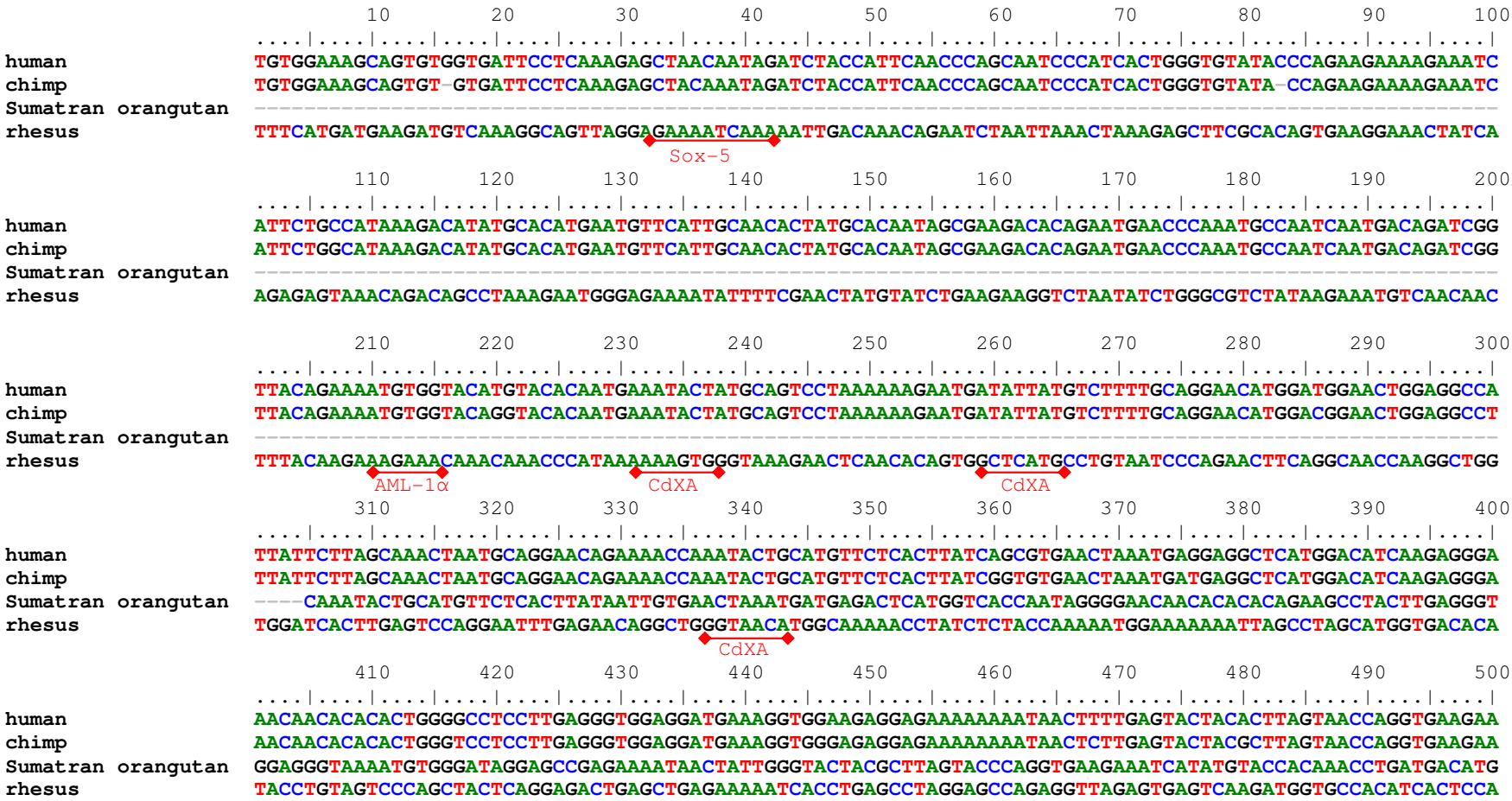
b) *cis* and *trans*-acting regulatory mechanisms

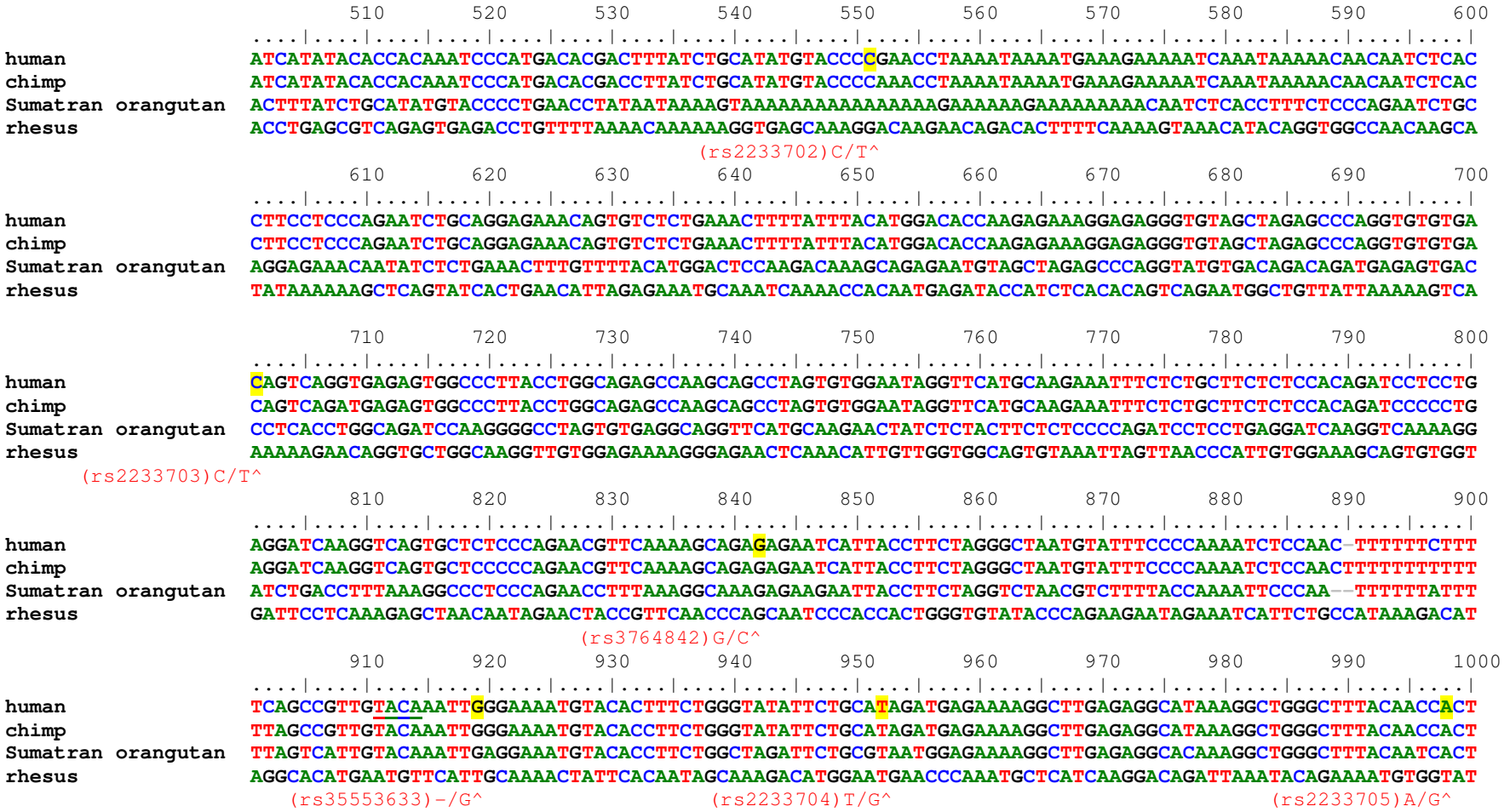
To date, the promoter region of *LGALS13* has not been fully characterised. In this study, a range of *in silico* tools were used to analyse a genomic region 1500 base pairs upstream from *LGALS13*. The programs MotifSampler and WeederH were used to identify known *cis*-elements in the putative promoter region of *LGALS13*. This investigation was aimed at gaining a better understanding of the regulation and control of this gene. Table 4 and Figure 12 show that the putative promoter region of *LGALS13* exhibits a non-classical architecture and lacks classical promoter sequence motifs such as TATA and CAAT boxes. The location of the promoter region of this gene was, thus, not predicted by the presence of classical promoter motifs.

An alignment of the putative promoter areas (1300 base pairs upstream) of the human, chimpanzee, orang-utan and rhesus monkey *LGALS13* orthologues is shown in Figure 12. These orthologues were used for the identification of conserved putative regulatory elements or regions within the

proximal promoter of *LGALS13.* The choice of species was based on the availability of genomic sequence information as well as the known expression profile of the gene of interest. *LGALS13* has to date not been reported to be expressed in any other eutherian mammals but those used in this study. The limited genomic sequence information available for the *LGALS13* orthologues from primates restricted the promoter area to be analysed to 1300 base pairs upstream from the gene of interest. It is interesting to note the level of conservation between the different orthologues in Figure 12. The human and chimp sequences are highly conserved throughout the alignment. The orang-utan and human sequences are conserved up to approximately 400 base pairs upstream of the gene. The rhesus monkey and human sequences are conserved up to approximately 100 base pairs upstream of the gene. This degradation of sequence conservation portrays the evolutionary relatedness between the above mentioned species, based on several DNA and protein sequence conservation studies (Wu and Su, 2008; Wang and Su, 2004; Hendy *et al*. 1994).

The multiple sequence alignment shown in Figure 12 reveals many interesting features of the putative promoter area of *LGALS13*. The predicted location of known *cis*-elements (Sox-5, CdXA, AML-1α, AP-1 as well as several putative E-boxes and a CAP site) are mapped onto Figure 12 and listed in Table 4.


10 20 30 40 50 60 70 80 90 100
human TGTGGAAAGCAGTGTGGTGATTCCTCAAAGAGCTAACAATAGATCTACCATTCAACCCAGCAATCCCATCACTGGGTGTATACCCAGAAGAAAAGAAATC
chimp TGTGGAAAGCAGTGT-GTGATTCCTCAAAGAGCTACAAATAGATCTACCATTCAACCCAGCAATCCCATCACTGGGTGTATA-CCAGAAGAAAAGAAATC
Sumatran orangutan
rhesus TTTCATGATGAAGATGTCAAAGGCAGTTAGGAGAAAATCAAAAATTGACAAACAGAATCTAATTAAACTAAAGAGCTTCGCACAGTGAAGGAAACTATCA
Sox-5
110 120 130 140 150 160 170 180 190 200
human ATTCTGCCATAAAGACATATGCACATGAATGTTCATTGCAACACTATGCACAATAGCGAAGACACAGAATGAACCCAAATGCCAATCAATGACAGATCGG
chimp ATTCTGGCATAAAGACATATGCACATGAATGTTCATTGCAACACTATGCACAATAGCGAAGACACAGAATGAACCCAAATGCCAATCAATGACAGATCGG
Sumatran orangutan
rhesus AGAGAGTAAACAGACAGCCTAAAGAATGGGAGAAAATATTTTCGAACTATGTATCTGAAGAAGGTCTAATATCTGGGCGTCTATAAGAAATGTCAACAAC
210 220 230 240 250 260 270 280 290 300
human TTACAGAAAATGTGGTACATGTACACAATGAAATACTATGCAGTCCTAAAAAAGAATGATATTATGTCTTTTGCAGGAACATGGATGGAACTGGAGGCCA
chimp TTACAGAAAATGTGGTACAGGTACACAATGAAATACTATGCAGTCCTAAAAAAGAATGATATTATGTCTTTTGCAGGAACATGGACGGAACTGGAGGCCT
Sumatran orangutan
rhesus TTTACAAGAAAGAAACAAACAAACCCATAAAAAAGTGGGTAAAGAACTCAACACAGTGGCTCATGCCTGTAATCCCAGAACTTCAGGCAACCAAGGCTGG
AML-1α CdXA CdXA
310 320 330 340 350 360 370 380 390 400
human TTATTCTTAGCAAACTAATGCAGGAACAGAAAACCAAATACTGCATGTTCTCACTTATCAGCGTGAACTAAATGAGGAGGCTCATGGACATCAAGAGGGA
chimp TTATTCTTAGCAAACTAATGCAGGAACAGAAAACCAAATACTGCATGTTCTCACTTATCGGTGTGAACTAAATGATGAGGCTCATGGACATCAAGAGGGA
Sumatran orangutan ----CAAATACTGCATGTTCTCACTTATAATTGTGAACTAAATGATGAGACTCATGGTCACCAATAGGGGAACAACACACACAGAAGCCTACTTGAGGGT
rhesus TGGATCACTTGAGTCCAGGAATTTGAGAACAGGCTGGGTAACATGGCAAAAACCTATCTCTACCAAAAATGGAAAAAAATTAGCCTAGCATGGTGACACA
CdXA
410 420 430 440 450 460 470 480 490 500
human AACAACACACACTGGGGCCTCCTTGAGGGTGGAGGATGAAAGGTGGAAGAGGAGAAAAAAAATAACTTTTGAGTACTACACTTAGTAACCAGGTGAAGAA
chimp AACAACACACACTGGGTCCTCCTTGAGGGTGGAGGATGAAAGGTGGGAGAGGAGAAAAAAAATAACTCTTGAGTACTACGCTTAGTAACCAGGTGAAGAA
Sumatran orangutan GGAGGGTAAAATGTGGGATAGGAGCCGAGAAAATAACTATTGGGTACTACGCTTAGTACCCAGGTGAAGAAATCATATGTACCACAAACCTGATGACATG
rhesus TACCTGTAGTCCCAGCTACTCAGGAGACTGAGCTGAGAAAAATCACCTGAGCCTAGGAGCCAGAGGTTAGAGTGAGTCAAGATGGTGCCACATCACTCCA


510 520 530 540 550 560 570 580 590 600
human ATCATATACACCACAAATCCCATGACACGACTTTATCTGCATATGTACCCCGAACCTAAAATAAAATGAAAGAAAAATCAAATAAAAACAACAATCTCAC
chimp ATCATATACACCACAAATCCCATGACACGACCTTATCTGCATATGTACCCCAAACCTAAAATAAAATGAAAGAAAAATCAAATAAAAACAACAATCTCAC
Sumatran orangutan ACTTTATCTGCATATGTACCCCTGAACCTATAATAAAAGTAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAAAAACAATCTCACCTTTCTCCCAGAATCTGC
rhesus ACCTGAGCGTCAGAGTGAGACCTGTTTTAAAACAAAAAAGGTGAGCAAAGGACAAGAACAGACACTTTTCAAAAGTAAACATACAGGTGGCCAACAAGCA
(rs2233702)C/T^
610 620 630 640 650 660 670 680 690 700
human CTTCCTCCCAGAATCTGCAGGAGAAACAGTGTCTCTGAAACTTTTATTTACATGGACACCAAGAGAAAGGAGAGGGTGTAGCTAGAGCCCAGGTGTGTGA
chimp CTTCCTCCCAGAATCTGCAGGAGAAACAGTGTCTCTGAAACTTTTATTTACATGGACACCAAGAGAAAGGAGAGGGTGTAGCTAGAGCCCAGGTGTGTGA
Sumatran orangutan AGGAGAAACAATATCTCTGAAACTTTGTTTTACATGGACTCCAAGACAAAGCAGAGAATGTAGCTAGAGCCCAGGTATGTGACAGACAGATGAGAGTGAC
rhesus TATAAAAAAGCTCAGTATCACTGAACATTAGAGAAATGCAAATCAAAACCACAATGAGATACCATCTCACACAGTCAGAATGGCTGTTATTAAAAAGTCA
710 720 730 740 750 760 770 780 790 800
human CAGTCAGGTGAGAGTGGCCCTTACCTGGCAGAGCCAAGCAGCCTAGTGTGGAATAGGTTCATGCAAGAAATTTCTCTGCTTCTCTCCACAGATCCTCCTG
chimp CAGTCAGATGAGAGTGGCCCTTACCTGGCAGAGCCAAGCAGCCTAGTGTGGAATAGGTTCATGCAAGAAATTTCTCTGCTTCTCTCCACAGATCCCCCTG
Sumatran orangutan CCTCACCTGGCAGATCCAAGGGGCCTAGTGTGAGGCAGGTTCATGCAAGAACTATCTCTACTTCTCTCCCCAGATCCTCCTGAGGATCAAGGTCAAAAGG
rhesus AAAAAGAACAGGTGCTGGCAAGGTTGTGGAGAAAAGGGAGAACTCAAACATTGTTGGTGGCAGTGTAAATTAGTTAACCCATTGTGGAAAGCAGTGTGGT
(rs2233703)C/T^
810 820 830 840 850 860 870 880 890 900
human AGGATCAAGGTCAGTGCTCTCCCAGAACGTTCAAAAGCAGAGAGAATCATTACCTTCTAGGGCTAATGTATTTCCCCAAAATCTCCAAC-TTTTTTCTTT
chimp AGGATCAAGGTCAGTGCTCCCCCAGAACGTTCAAAAGCAGAGAGAATCATTACCTTCTAGGGCTAATGTATTTCCCCAAAATCTCCAACTTTTTTTTTTT
Sumatran orangutan ATCTGACCTTTAAAGGCCCTCCCAGAACCTTTAAAGGCAAAGAGAAGAATTACCTTCTAGGTCTAACGTCTTTTACCAAAATTCCCAA--TTTTTTATTT
rhesus GATTCCTCAAAGAGCTAACAATAGAACTACCGTTCAACCCAGCAATCCCACCACTGGGTGTATACCCAGAAGAATAGAAATCATTCTGCCATAAAGACAT
(rs3764842)G/C^
910 920 930 940 950 960 970 980 990 1000
human TCAGCCGTTGTACAAATTGGGAAAATGTACACTTTCTGGGTATATTCTGCATAGATGAGAAAAGGCTTGAGAGGCATAAAGGCTGGGCTTTACAACCACT
chimp TTAGCCGTTGTACAAATTGGGAAAATGTACACCTTCTGGGTATATTCTGCATAGATGAGAAAAGGCTTGAGAGGCATAAAGGCTGGGCTTTACAACCACT
Sumatran orangutan TTAGTCATTGTACAAATTGAGGAAATGTACACCTTCTGGCTAGATTCTGCGTAATGGAGAAAAGGCTTGAGAGGCACAAAGGCTGGGCTTTACAATCACT
rhesus AGGCACATGAATGTTCATTGCAAAACTATTCACAATAGCAAAGACATGGAATGAACCCAAATGCTCATCAAGGACAGATTAAATACAGAAAATGTGGTAT
(rs35553633)-/G^
(rs2233704)T/G^
(rs2233705)A/G^

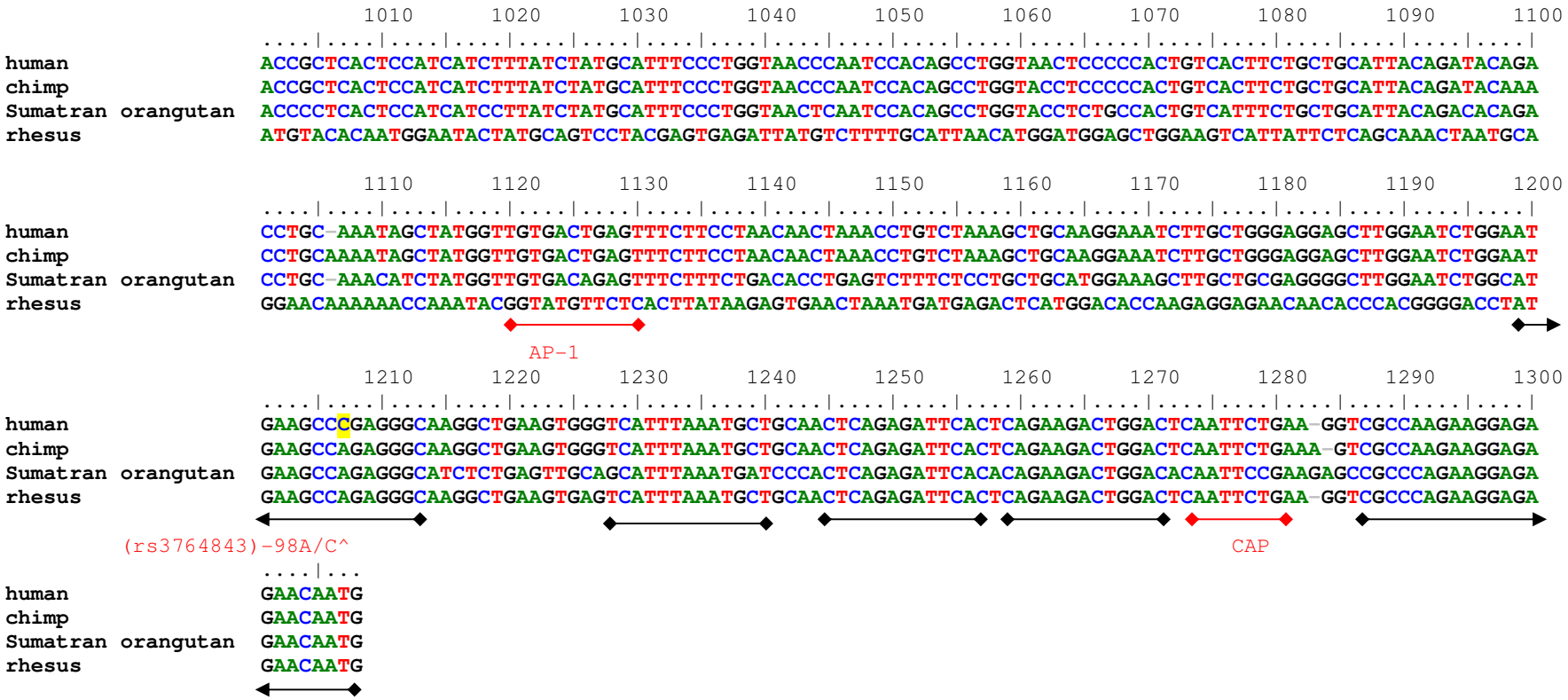


Figure 12: Multiple sequence alignment of 1300 base pairs upstream of the human *LGALS13* gene and its orthologues in chimpanzee, Sumatran orang-utan and rhesus monkey. The alignment ends with the start codon (ATG) of each gene. Numbers above the alignment follow the 5'-3' orientation, towards the start codon. Position -1 is thus indicated by 1305 and position -98 by 1207 on this alignment. Blocks of highly conserved regions in all sequences are indicated by black lines below the alignment. Red lines below the alignment indicate the position of known sequence elements which may play a role in gene expression and control. The names of these elements are also given below each red line. Single nucleotide polymorphisms (SNP's) occurring in the human *LGALS13* gene sequence are indicated by the name of the SNP in red, followed by the base pair position and the nucleotide change which occurs.

Table 4: Predicted conserved sequence elements with the names of their putative corresponding *trans*-acting factors identified by bioinformatics tools.

Putative *cis*-acting element	**Consensus sequence**	**Motif**	**Putative *trans*-acting factor**	**Position in sequence**		
				Human	**Chimpanzee**	**Rhesus monkey**
CAAT box	GG(T/C)CAATCT	CCAAT	c/EBP CTF/NF1	None	None	-1580 to -1576
TATA box	TATAAAA	TATAA	Transcription Factor IID (TFIID)	None	None	-171 to -167
						-704 to -698
GC box	GGGNGG	GGGC/AGGG	Stimulatory protein 1 (Sp1)	None		
TRE	RSTGACTMANN	GTGAGT TGTGACTGAGT	Ap1 family	-185 to -175	-185 to -175	-8434 to -8429, -8228 to -8223, -8147 to -8142, -8135 to -8130, -8131 to -8126, -8088 to -8083, -6235 to -6230, -833 to -828, -82 to -77
CRE	GTGACGT	GTGACGT	CREB/ATF family	None		
Cap	NCANNNNN	TCATTACC		-456 to -449	-458 to -451	None

Data obtained using WeederH: Pavesi *et al*. 2007 and MotifSampler: Thijs *et al.* 2001 and the multiple sequence alignment of the promoter region of *LGALS13* (Figure 12)

Classical eukaryotic promoter sequence motifs such as the TATA and CAAT boxes were predicted to be present in the rhesus monkey orthologue of *LGALS13* by the bioinformatic tools, WeederH and MotifSampler (Table 4). These motifs were not predicted to be present in the chimpanzee or human orthologues of *LGALS13*. The TATA box is recognised by TBP (TATA-binding protein), which is a key component in the basal transcription complex. The correct assembly of this complex of transcription factors is crucial for initiating transcription in eukaryotes. The CAAT box is often used to identify the promoter region of a eukaryotic gene and is recognised by the transcription factor, NF-γ. Moreover, no CRE-elements or GC boxes, motifs which are often found in eukaryotic promoter regions, were identified in the *LGALS13* orthologues.

A motif associated with a TRE-element (TPA-responsive element) was, however, found in the human, chimpanzee and rhesus monkey orthologues of the gene of interest. The AP-1 (Activator protein 1) family of transcription factors bind specifically to the AP-1 or TRE-elements. This protein family can be divided into Fos and Jun subfamilies. Proteins such as c-Jun, JunB, c-Fos, Fra-1 and Fra-2, belong to these families and are found in mammals. AP-1 proteins bind in dimer formation, with very high affinity to TRE, and with lower affinity to CRE (cAMP Response Element). CRE motifs are usually recognised by members of the activating transcription factor (ATF) or CRE binding (CREB) protein families, which also bind their target sequences as dimers. The sequence motif for a cap-site was also identified in the human and chimpanzee orthologues of *LGALS13*. The cap motif signals the initiation of transcription (Thijs *et al.* 2001).

1.6 Summary of findings from Section 1

In summary, *LGALS13* is a member of the multigene family encoding galectin proteins. *LGALS10* and -*14* are situated in closest proximity to *LGALS13* of all the members of the subfamily of galectins they fall under, and share a high degree of sequence homology (55% and 68% respectively) with the gene of interest. *LGALS13* is flanked by two putative genes, LOC148003 and

LOC390930. The protein product of LOC148003 is 75% identical to PP13, and exhibit a highly similar predicted tertiary structure. These findings might have implications for the use of PP13 as a predictive biomarker for pre-eclampsia. PP13 is highly conserved in primates and not found in any other mammalian species, which indicates that it might have a role in placentation. In humans, *LGALS13* is almost exclusively expressed in the placenta.

The core sequence encoding the CRD spans exons 2, 3 and 4. Many functionally important amino acid residues (involved in disulfide bonding and phosphorylation) are also located in these exons.

On the DNA level, *LGALS13* has a non-classical promoter architecture and lacks TATA and CAAT boxes. A number of *cis*-elements, including an element associated with the Ap-1 family of transcription factors, are predicted to be present in the putative promoter area, which exhibits blocks of sequence conservation.

The putative promoter region analysed in this study exhibited no predicted CpG islands, indicating that this gene is probably not regulated via methylation. The intron-exon boundaries of *LGALS13* conform to the GT/AG rule. Sequence elements specific for members of the SR-rich family of splice factors are putatively present in and around exon 3.

2 Sequence variation in *LGALS13*

The human gene encoding Placental Protein 13, *LGALS13* exhibits a high level of sequence variation in general, and particularly in the study population. In December 2008, a total of 41 SNPs had been localised in the human *LGALS13* gene and documented on the NCBI database, as shown in the annotation of the gene sequence (Supplementary data).

This study reports on the genotype frequencies at seven of the polymorphic loci in *LGALS13*, 4 of which are novel and have not been documented on the SNP database (Bruiners, 2007). The polymorphisms focused on in this study include one SNP in the promoter area, four intronic SNPs and two exonic sequence variants (Figure 13 and Table 5).

To date, no conformational variants have been identified in the amplicons representing exons 2 and 4 of *LGALS13*. This might be due to the high degree of conservation of these regions which might imply that these regions are important for the structural and functional properties of PP13. Figure 13 clearly shows that the majority of the *LGALS13* variants investigated in this study cluster around exon 3. The fact that exon 3 seems to be less conserved than exons 2 and 4 is interesting since exon 3 has been shown to be functionally relevant (See Section 1). Table 5 summarises the sequence variants investigated in this study, as well as their possible effects on *LGALS13* expression and function.

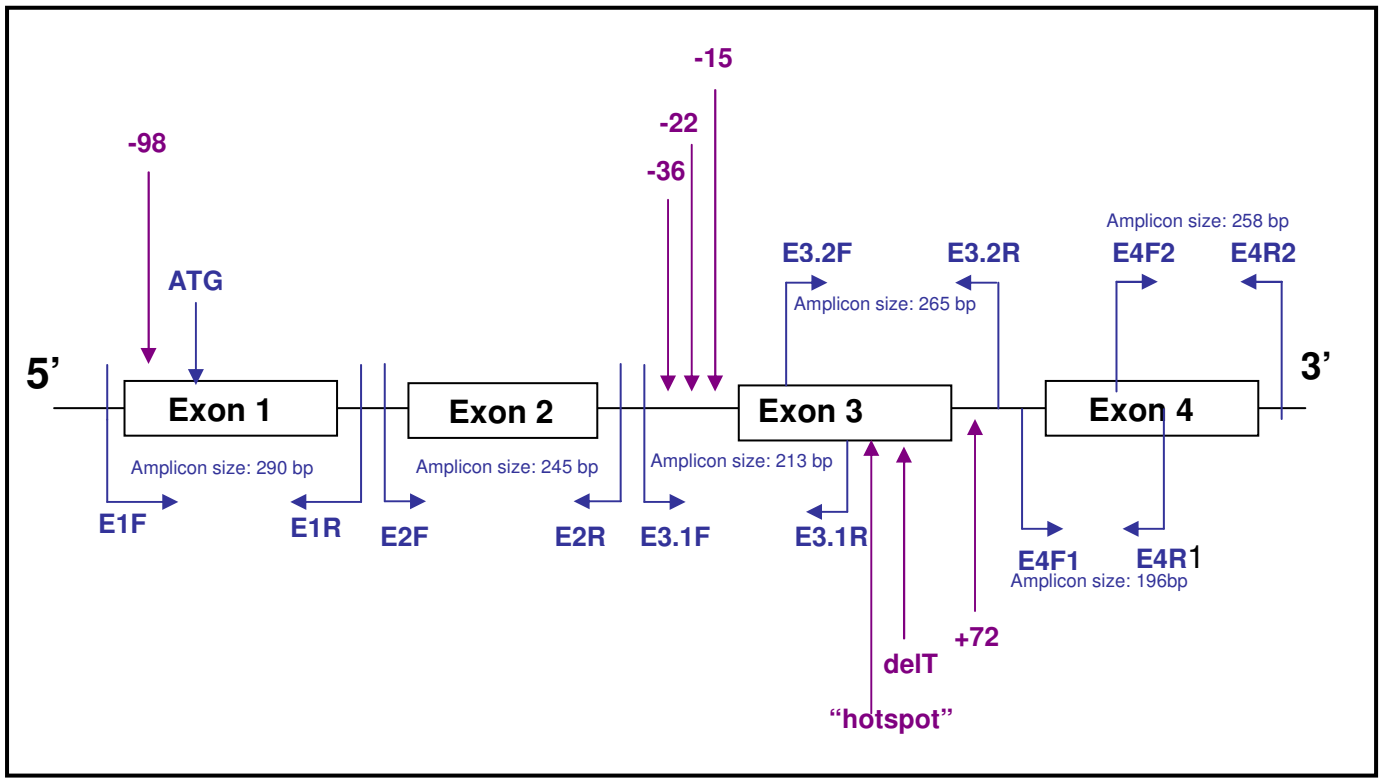


Figure 13: Graphical representation of the intron/exon structure of *LGALS13*. The positions of the sequence variants identified are indicated by purple arrows. -98 Refers to the position 98 base pairs upstream from the start codon. The intronic variants are named relative to their position to exon 3, thus the -36 variant is 36 base pairs upstream from exon 3 and the +72 variant is 72 base pairs downstream from exon 3. The positions of the primers and sizes of the PCR amplicons generated, which are used to type these variants, are shown in blue.

The pregnancy-related protein encoded by the gene of interest, *LGALS13*, is currently being used as a biomarker for the early detection of pre-eclampsia. This warrants the characterization of *LGALS13* as well as the analysis of its sequence variants and their possible effects on gene expression. The DNA variants this study focuses on can be divided into two main categories, i) variants which might influence the transcriptional regulation of the gene (-98 A/C) and ii) variants which might generate isoforms (intronic and exonic variants) (Figure 13 and Table 5).

Table 5: Summary of polymorphisms identified in *LGALS13* with the possible functional effects they might have on protein expression and function

Name	Description	Possible effect
-98 A/C	A/C change 98bp upstream of ATG, dbSNP: rs3764843	Alter expression profile of PP13
IVS2 -36 (G/A)	G/A change in intron 2, novel (Bruiners, 2007)	Create alternative splicing
IVS2 -22 (A/G)	A/G change in intron 2, dbSNP: rs2233706	
IVS2 -15 (G/A)	G/A change in intron 2, novel (Bruiners, 2007)	
IVS3 +72 (T/A)	T/A change in intron 3, dbSNP: rs2233708	
229delT	Deletion of single T in exon 3 (Bruiners, 2007)	Truncated/defective protein or decrease in protein expression
"Hotspot"	6bp variation within a 11bp region in exon 3 (Bruiners, 2007)	Alteration of secondary and tertiary protein structure, functional effects

Study participants were genotyped for the above-mentioned DNA sequence polymorphisms using the multiphor SSCP/HDP system. The study cohort used for this investigation will be discussed in more detail in Section 3. Sequence variants were identified by different conformations on the multiphor polyacrylamide gels. In some instances, genotypes were confirmed using Restriction enzyme digestion and semi-automated DNA sequencing. Multiphor SSCP/HDP proved to be a sensitive and accurate tool for the identification of sequence variants in this study. Hardy Weinberg equilibrium (HWE) was determined at each locus for each individual group. The

significance of the differences observed in genotype and allele frequencies between different groups was validated using Fisher's exact test and the Chi^2 test.

Genotype frequencies at the -98, IVS3 +72T/A and IVS2 -22A/G loci were found to be in Hardy-Weinberg equilibrium. The IVS2 -15 G/A polymorphism was absent from this study cohort which reflects how rare this mutation is. The IVS2 -36G/A polymorphism was only found in the homozygous wild type and heterozygous states and not in the homozygous mutant state. This might indicate that the mutant allele is selected against. Another study has reported on the genotyping results of these sequence variants in a similar study cohort. In this study, all of the intronic variants were found to be in Hardy-Weinberg equilibrium (Bruiners, 2007).

The exonic variants found in *LGALS13*, namely the 229delT and the so-called "hotspot mutation", were only detected in the heterozygous state in this study. This might indicate that these variants have some degree of negative effect on individuals who bear them, and that these effects might be intensified if these variants were to occur in both *LGALS13* alleles of an individual.

2.1 Variation within the putative promoter region

One of the known SNPs in *LGALS13* is located in the putative promoter area. This polymorphism results in a change from A to C 98 base pairs upstream from the start codon. Individuals in the study cohort were genotyped for this polymorphism via Restriction enzyme digestion of PCR products amplified from exon 1 of *LGALS13*. The C allele creates a recognition site for the enzyme, *Ava*l. Figure 14 depicts the three genotypes for this locus: an A/A genotype yields no digestion and only one fragment (292 base pairs in size) is visualised on an agarose gel. The A/C genotype yields digestion of the C allele producing and two fragments are visible on an agarose gel. The CC genotype yields a single fragment, which is smaller than that of the A/A genotype, and this is visible on an agarose gel (242 base pairs in size).

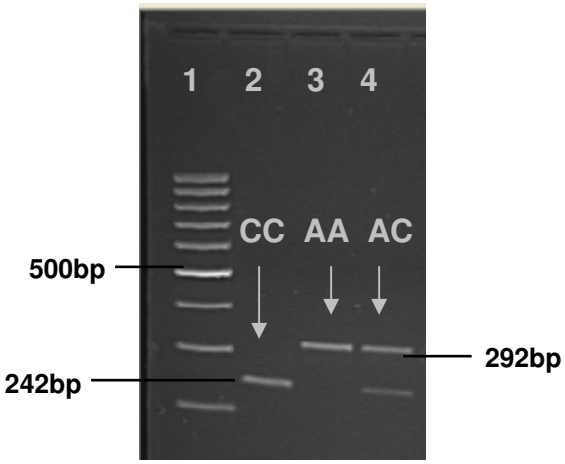


Figure 14: Agarose gel electrophoresis image depicting DNA fragments resulting from differential restriction enzyme digestion. PCR products of *LGALS13* exon 1 were digested using *Ava*I according to manufacturers' specifications. The single fragment in lane 2 represents the CC genotype. The larger single band in lane 3 represents the AA genotype and two fragments in lane 4 depict the heterozygous AC genotype. Lane 1 shows a 100bp ladder as size standard (Fermentas Gene O'ruler 100bp DNA ladder).

Table 6: Observed genotype and allele frequencies at the -98 A/C locus in the total cohort. P-values below the table indicate the statistical significance of any correlations observed.

		Total Cohort N=302	**Coloured N=214**	**Black N=88**
Genotype frequency	**A/A**	153 (0.51)	103 (0.48)	50 (0.57)
	A/C	127 (0.42)	97 (0.45)	30 (0.34)
	C/C	22 (0.07)	14 (0.07)	8 (0.09)
Allele frequency	**A**	433 (0.72)	303 (0.71)	130 (0.74)
	C	171 (0.28)	125 (0.29)	46 (0.26)

(Genotype frequency:Black vs Coloured Chi^2 p-value=0.186; Allele frequency: Black vs Coloured Fisher's exact test two-tailed p-value= 0.487).
Note: Due to population stratification and technical constraints, sample sizes differ from the total cohort for each genotype frequency analysis reported in this study.

The -98 A/C variant occurs in the Coloured and Black populations of the Western Cape of South Africa (Table 6). All genotype distributions were found to be in Hardy Weinberg equilibrium ($p>0.05$). Statistical analyses were performed and no statistically significant difference was found between the genotype/allele frequencies of the two population groups studied ($p>0.05$).

Table 7: Observed genotype and allele frequencies at the -98 A/C locus in the study participants with known pregnancy outcomes. P-values below the table indicate the statistical significance of any correlations observed.

Population group		**Coloured** N=92		**Black** N=42	
Pregnancy outcome		**Adverse** N=23	**Controls** N=69	**Adverse** N=8	**Controls** N=34
Genotype frequency	**A/A**	9 (0.39)	38 (0.55)	5 (0.63)	19 (0.56)
	A/C	14 (0.61)	25 (0.36)	1 (0.13)	12 (0.35)
	C/C	0 (0.00)	6 (0.09)	2 (0.25)	3 (0.09)
Allele frequency	**A**	32 (0.70)	101 (0.73)	11 (0.69)	50 (0.74)
	C	14 (0.30)	37 (0.27)	5 (0.31)	18 (0.27)

(Adverse outcome vs Controls Coloured population Genotype frequencies Chi^2 p-value= 0.07, Allele frequencies Fisher's exact test two-tailed p-value= 0.704; Black population Genotype frequencies Chi^2 p-value= 0.278, Allele frequencies Fisher's exact test 2-tailed p-value= 0.758).

Data from the pregnancies with known outcomes was analysed for association between genotype/allele frequencies and adverse pregnancy outcome. No association between the genotype/allele distribution at the -98 locus and pregnancy outcome was found in either of the population groups studied ($p>0.05$).

Table 7 shows that the A/A genotype frequency of the Coloured population differs significantly (39% vs 55%) between the two subgroups (Adverse pregnancy outcome and Control groups). This frequency does, however, not seem to differ significantly (63% vs 56%) between the two subgroups in the Black population of the Western Cape. The A/C genotype seems to be distributed differently between the Adverse pregnancy outcome and Control groups for the different population groups. In the Coloured population, the A/C genotype frequency of the Adverse group is almost double that of the Control group (61% vs 36%). However, in the Black population, the A/C genotype frequency of the Control group is approximately double that of the Adverse group (35% vs 13%). The allele frequencies are similar across the population and pregnancy outcome groups (~70% vs ~30%). It is, however, important to note that the differences in sample sizes between the subgroups within each population affects the statistical significance of these observations.

Although *LGALS13* seems to have a non-classical, TATA-less promoter, this variant most likely falls within the core promoter of this gene, being located 98 base pairs upstream from the ATG codon. According to the bioinformatic analyses done, this variant does not disrupt any known regulatory elements. A multiple sequence alignment of *LGALS13* and its orthologues from other primates (chimpanzee, rhesus monkey and orang-utan) has revealed many interesting aspects of the putative promoter region of this gene. Predicted *cis*-elements were mapped onto this alignment (Figure 12) which reveals blocks of conservation in these putative regulatory regions. The -98A/C variant occurs in one of these blocks of conservation, between positions 1199 and 1214 in Figure 12). The location of this variant, in a block of sequence conservation, as well as the conservation of the A-allele in the primates, may affect the regulation of *LGALS13.* The A-allele is also probably the ancestral form of this variant.

The other known *cis*-elements identified in this alignment are Sox-5, CdXA, AML-1α, a GC-box, AP-1 as well as several putative E-boxes and a CAP site (Table 4). The AP-1 element is of particular interest since it has been shown to function as a mediator between Estrogen Receptor α (ERα) and its corresponding Estrogen Responsive element (ERE). The tissue specific expression of *LGALS13* (almost exclusively placental expression, Figure 9), makes the prediction of an Estrogen-sensitive regulatory element noteworthy. This observation begs the question whether *LGALS13* is hormonally regulated by Estrogen during pregnancy. Further experimental data is necessary to analyse this hypothesis. Site-directed mutagenesis of this putative element, followed by Luciferase gene expression assays could be conducted in future to bring more clarity to this topic. The tissue and developmental stage specific expression of *LGALS13* emphasises the complexity of its regulation. Further analysis and characterization of the promoter area of this gene is necessary. Functional studies of the promoter region of this gene are underway and will shed more light on the significance of this variant and whether it impacts on the expression of *LGALS13*.

2.2 Variation within the coding region of *LGALS13*

Data generated from genotyping individuals from the Western Cape region (South Africa) has revealed the existence of two novel DNA sequence polymorphisms in exon 3 of *LGALS13* (Bruiners, 2007). These two variants have been named the 229delT variant and the so-called “hotspot mutation”. These variants were genotyped in the study cohort using the multiphor SSCP/HDP system, which was found to be a sensitive detection tool for these polymorphisms.

2.2.1 229delT

The 229delT variant is a deletion mutation at base pair number 229 in exon 3, a Thymine base is deleted. Figure 15 clearly indicates that this polymorphism results in the truncation of PP13 to a protein 111 amino acids in length. The deletion of this nucleotide results in a severe change in the amino acid composition of the mutant protein. Several of the important amino acid residues, such as Tyr80, Cys92, Cys136 and Cys138, are disrupted by this deletion. Some of the amino acids which fold into beta sheets are also disrupted; consequently, 6 of the 11 beta sheets present in secondary and tertiary structural conformations of PP13 are most likely absent in the mutant protein.

From Table 8 it can be seen that the genotype frequencies for this deletion did not differ significantly between the two populations. The 229delT allele has, to date, never been observed in the homozygous state. This is most likely due to the low frequency of this variant observed in the population studied. This might imply that bearing two copies of the *LGALS13* allele with this deletion is not viable and that such an individual might not develop fully after fertilization. It could be possible that the effect of this deletion is so severe that individuals who are homozygous for this variant are never observed in the population. This might be due to the fact that this deletion causes a truncated protein product and disrupts several key features within the DNA and amino acid sequences.

Table 8: Observed genotype and allele frequencies for the 229delT deletion mutation in the total cohort. A "+" sign represents the wild type allele, while a "–" sign represents the mutant allele.

		Total Cohort N=309	**Coloured N=219**	**Black N=90**
Genotype frequency	**+/+**	306 (0.990)	218 (0.995)	88 (0.978)
	+/-	3 (0.010)	1 (0.005)	2 (0.002)
	-/-	0 (0.000)	0 (0.000)	0 (0.000)
Allele frequency	**+**	615 (0.995)	437 (0.998)	178 (0.989)
	-	3 (0.005)	1 (0.002)	2 (0.011)

Association studies between pregnancy outcome and this deletion mutation in *LGALS13* could not be performed due to small sample sizes and incomplete data. One study participant from the Coloured population was found to have this mutation. The pregnancy outcome of this subject was, however, unknown at the time this analysis was done. Of the two individuals bearing this mutation in the Black population, one had a good pregnancy outcome and the other one's outcome was unknown at the time of data analysis. Another study participant who also had the 229delT mutation, developed pre-eclampsia at 35 weeks of gestation. The outcome of this pregnancy was, however, unknown at the time of completion of this study. This individual did not ethnically match the South African population studied and was subsequently left out of the genotype and allele frequency analyses.

The 229delT mutation has previously been found to be associated with adverse pregnancy outcome and preterm labour in Coloured women (Bruiners, 2007). This study also showed that the +/- genotype frequency for the 229delT was higher in women who developed pregnancy related hypertension than in controls (5.7% vs 2.4%). No association between hypertension and the 229delT mutation was, however, statistically demonstrated. It is necessary to study the possible association between this mutation in *LGALS13* in the context of a larger cohort, including both Coloured and Black subjects.

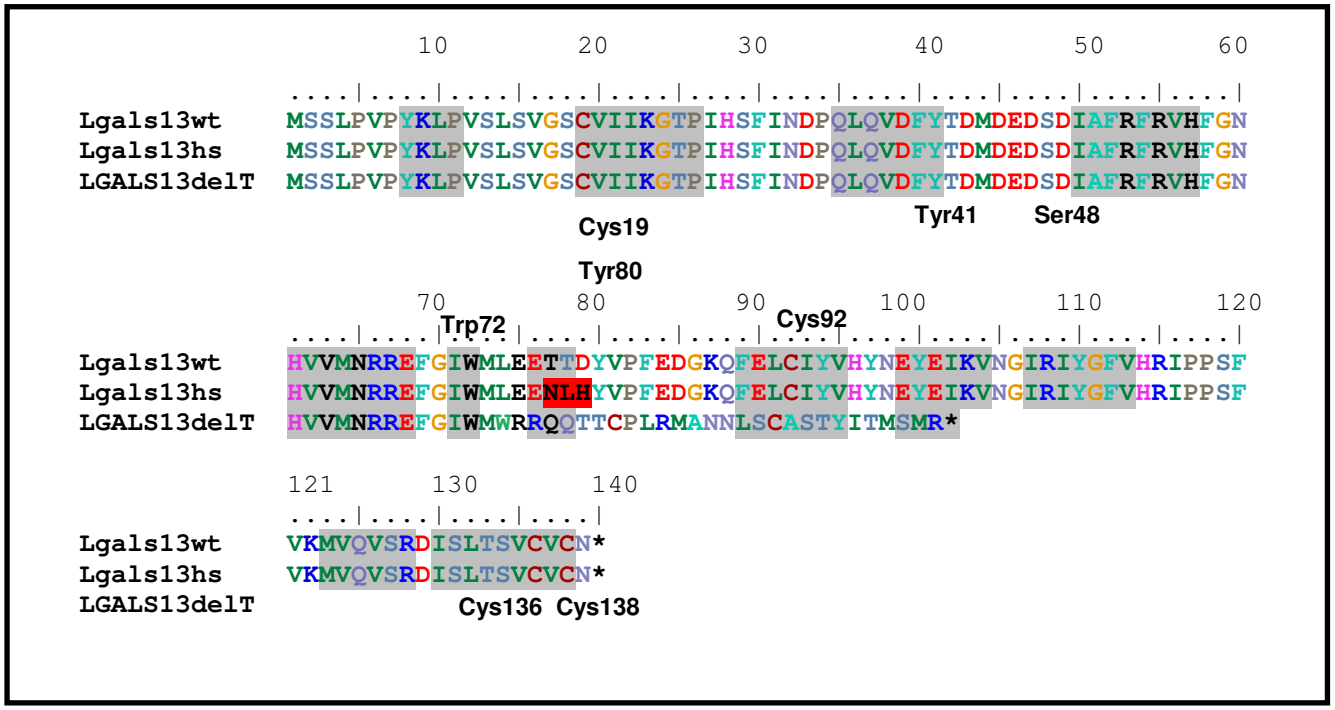


Figure 15: Amino acid sequence alignment of the wild type PP13 with two sequence variants, the translated hotspot and DelT variant sequences. Numbers above the alignment indicate the true amino acid positions in each peptide. Lgals13wt is the wild type protein sequence, Lgals13hs is the translation of the DNA sequence with the hotspot mutation and Lgals13delT is the translation of the DNA sequence with the deletion of a T nucleotide. Amino acid residues which form part of the 11 Beta sheets (S1-6 and F1-5) in the secondary structure of PP13, are shaded in grey. The names and positions of the amino acids which are involved in the establishment of the tertiary protein structure are given above or below each respective residue (Cys: Cysteine, Tyr: Tyrosine, Ser: Serine). The amino acid residues which are predicted to participate and aid in carbohydrate recognition and binding are indicated by black typing (Visegrády *et al*. 2001; Cooper *et al*. 2002). The three amino acid changes caused by the “hotspot” mutation are indicated by red shading and black typing between positions 70 and 80 in the Lgals13hs sequence.

2.2.2 The so-called “Hotspot mutation”

The second DNA sequence polymorphism present in the study population has been named the “hotspot mutation”, and results in a six base pair change within the 11bp region between base pair numbers 137 and 148 in exon 3 (Figure 4). Figure 15 indicates that this polymorphism results in a three amino acid change in the protein product. The mutant protein has the amino acids, Asparagine, Leucine and Histidine (NLH), whereas the wild type protein has two Threonines followed by an Aspartic Acid (TTD) between positions 77 and 79 in the peptide chain (Figure 15). Although these amino acid substitutions seem mild in comparison to the changes caused by the DelT mutation, the functional and structural properties of the protein might be affected. When considering the biochemical properties of the amino acid residues involved, Asparagine, Leucine and Threonine are neutral. However, the third amino acid change, from an Aspartic Acid to a Histidine might have an effect on the protein since Aspartic Acid has an acidic side chain and Histidine a basic one. Another important feature of this mutation is that it is located in one of the regions of the amino acid which folds into a beta sheet. This mutation might thus have an effect on the structure of PP13. If the amino acid changes occur in the region where the PP13-specific antibody used to measure PP13 levels in the diagnostic kit, the mutant protein might also have implications for the specificity of the test.

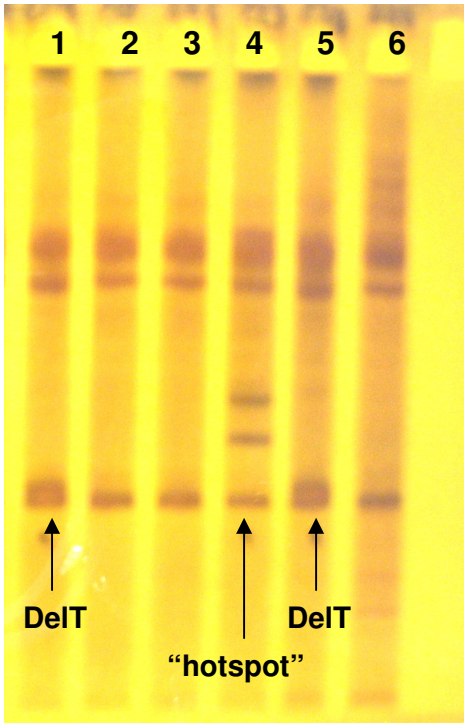


Figure 16: Image of a multiphor SSCP/HDP gel depicting the conformations of the DelT (lanes 1 and 5) and "hotspot" (lane 4) polymorphisms identified in *LGALS13*.

Table 9 shows the genotype frequencies of the so-called "hotspot mutation" in the study cohort. From this table, it can be inferred that there are no significant differences in the frequency of this variant between the two population groups studied. In the Black population, seven individuals had this sequence variant in the *LGALS13* gene. All of the known pregnancy outcomes in this group were good (N=4). The pregnancy outcomes of nine out of the 15 subjects from the Coloured population who had this variant were known. Six subjects had good pregnancy outcomes, two developed pre-eclampsia and one developed ROM. Association studies between the outcome of pregnancy and the frequency of this variant could unfortunately not be performed due to incomplete data and small sample sizes. It would, however, be interesting to study whether this variant is associated with an adverse pregnancy outcome in the Coloured population, since 3 out of the nine subjects who had this variant exhibited adverse pregnancy outcomes.

Table 9: Observed genotype and allele frequencies for the so-called "hotspot" mutation in the total cohort. A "+" sign represents the wild type allele while a "–" sign represents the mutant allele.

		Total Cohort N=309	Coloured N=219	Black N=90
Genotype frequency	**+/+**	287 (0.93)	204 (0.93)	83 (0.92)
	+/-	22 (0.07)	15 (0.07)	7 (0.08)
	-/-	0 (0.00)	0 (0.00)	0 (0.00)
Allele frequency	**+**	596 (0.96)	423 (0.97)	173 (0.96)
	-	22 (0.04)	15 (0.03)	7 (0.04)

This variant has never been observed in the homozygous state. This might be due to the relatively low incidence rate of this mutation. The hotspot allele has been found to be present in 2.9% of the Coloured and 5.6% of the Black women of a Western-Cape based study cohort (Bruiners, 2007). The reason for the relatively low incidence rate of this mutation might be that the postulated gene conversion event which caused this sequence variant occurred relatively recently, in evolutionary terms. The variant has thus not been distributed throughout the population. It would be interesting to perform more statistical analyses on the frequency of this variant, in a range of populations and larger cohort sizes than used in this study. This might indicate whether selective pressures are affecting the frequency of this variant. If this variant is detrimental to the population, it would probably be selected against. However, to date, no data has supported this theory.

2.3 Variation within the non-coding regions of *LGALS13*

The gene encoding PP13, *LGALS13*, spans a genomic region of 4950 base pairs in length and has four exons. Four intronic polymorphisms (IVS+72T/A, IVS-15G/A, IVS-22A/G and IVS-36G/A) have been identified around exon 3 of *LGALS13* (Table 5 and Figure 10). Study participants were genotyped for these polymorphisms.

2.3.1 The IVS3 +72 T/A polymorphism

This sequence variant (dbSNP: rs2233708) is located 72 base pairs downstream from exon 3 and results in the change from a Thymine (T) to an

Adenosine (A) allele. Individuals in the study cohort were genotyped for this variant via Restriction enzyme digestion of PCR products amplified from exon 3. The A-allele creates a restriction site which is recognised by the enzyme, *Stu*I. This yields differentially sized fragments which were visualised via agarose gel electrophoresis in this study. The TT genotype yields no digestion and only one fragment (265bp) is visible on an agarose gel, the TA genotype is distinguished by two fragments visible on an agarose gel while the AA genotype produces one visible fragment (246bp) which is smaller than the TT fragment (Figure 17).

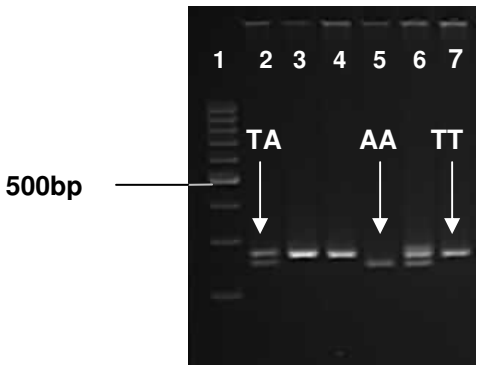


Figure 17: Agarose gel electrophoresis image depicting DNA fragments resulting from differential restriction enzyme digestion. PCR products of *LGALS13* exon 3.2 were digested using *Stu*I according to manufacturers' specifications. The two fragments in lane 2 represent the heterozygous genotype (TA). The single fragment in lane 7 represents the TT genotype while the AA genotype is depicted by the single smaller fragment in lane 5. Lane 1 shows a 100bp DNA ladder (Gene O'ruler 100bp DNA ladder, Fermentas).

Table 10: Observed genotype and allele frequencies at the IVS3 +72 T/A locus in the total cohort.

		Total Cohort N=307	**Coloured N=217**	**Black N=90**
Genotype frequency	**T/T**	235 (0.77)	176 (0.81)	59 (0.66)
	T/A	63 (0.21)	37 (0.17)	26 (0.29)
	A/A	9 (0.03)	4 (0.02)	5 (0.06)
Allele frequency	**T**	533 (0.87)	389 (0.90)	138 (0.80)
	A	81 (0.13)	45 (0.10)	36 (0.20)

(Coloured population vs Black population Genotype frequencies Chi^2 p-value= 0.009, Allele frequencies Mantel-Haenzel p-value=0.0007 and Fisher's exact test two-tailed p-value= 0.0013).

This variant was found to be in Hardy-Weinberg equilibrium in both populations investigated ($p>0.05$, Table 10). A significant difference between the two populations was found with regard to the allele ($p=0.0013$) and genotype ($p=0.009$) frequencies at this locus. Table 10 shows that the heterozygous genotype (T/A) has a slight enrichment in the Black population of the Western Cape, in comparison to the Coloured population (29% versus 17%). The A/A genotype frequencies are similar for the two populations but a difference between the two populations can be observed in the frequency of the A-allele. The A-allele is present in the Black population at a higher frequency than in the Coloured population (20% vs 10%). The A-alleles in the Black population are also distributed to a greater extent into the heterozygous genotype rather than the homozygous A/A genotype, resulting in a higher frequency of the T/A genotype in the Black population.

Table 11 gives the genotype and allele frequencies at the +72 locus for the Adverse pregnancy outcomes and Control groups of each population studied. This variant was found to be in Hardy Weinberg equilibrium in all of the subgroups within the two populations studied. The +72 allele and genotype frequencies were statistically analysed and compared between the healthy (Control) and Adverse groups. A significant difference between the Adverse and Control groups from the Coloured population was found with regard to the frequency of the A-allele ($p= 0.046$). This allele was found to be present at a higher frequency in the Control group than in the Adverse outcome group (16% vs 4%).

From Table 11 it can be seen that the genotype frequencies also differ between these groups. The frequency of homozygous T/T individuals is greater in the Adverse group than in the Control group (91% vs 73%), while the heterozygous genotype is at a lower frequency in the Adverse group in comparison to the Control group (9% vs 23%). This suggests that the T allele is distributed more evenly amongst genotypes of the Control individuals than in the Adverse group. The statistical tests performed did not support the significance of the observed differences in genotype frequency between the two groups ($p>0.05$).

Table 11: Observed genotype and allele frequencies at the IVS3 +72 T/A locus in the study participants with known pregnancy outcomes from the study cohort.

Population group		**Coloured** N=92		**Black** N=42	
Pregnancy outcome		**Adverse** N=23	**Controls** N=70	**Adverse** N=8	**Controls** N=34
Genotype frequency	**T/T**	21 (0.91)	51 (0.73)	6 (0.76)	22 (0.64)
	T/A	2 (0.09)	16 (0.23)	1 (0.12)	10 (0.30)
	A/A	0 (0.00)	3 (0.04)	1 (0.12)	2 (0.06)
Allele frequency	**T**	44 (0.96)	118 (0.84)	13 (0.81)	54 (0.78)
	A	2 (0.04)	22 (0.16)	3 (0.19)	14 (0.20)

(Adverse outcome vs Controls Coloured population Genotype frequencies Chi^2 p-value= 0.170 Allele frequencies Mantel-Haenzel p-value= 0.047 and Fisher's exact test 2-tailed p-value= 0.046; Black population Genotype frequencies Chi^2 p-value= 0.547, Allele frequencies Fisher's exact test 2-tailed p-value=1.00).

From this data, it could be inferred that the A-allele might have a protective effect in the Coloured population, since it seems that the lack thereof is associated with an Adverse pregnancy outcome. However, another study focusing on the allele and genotype frequencies of this variant in pregnant women from a similar study cohort, found no association between the genotype or allele frequencies and the pregnancy outcome in the Black and Coloured populations (Bruiners, 2007). These observations need to be verified before any conclusions can be drawn.

2.3.2 The IVS2 -15 G/A, IVS2 -22 A/G and –IVS2 36 G/A polymorphisms

The IVS2 -15 G/A sequence polymorphism is a novel SNP which was previously identified in the South African population (Bruiners, 2007). This variant is located 15 base pairs upstream from exon 3 and results in the change from a Guanine (G) to an Adenosine (A) base. The genotype results of 312 individuals out of the study cohort could be used for analysis. In this study, no IVS2 -15 G/A polymorphisms were detected. It would be interesting to genotype a larger cohort for this mutation and calculate the mutation rate of this allele. A previous study focusing on the genotype frequency of this variant in the South African population, found this variant to be rare (1.9%). It is thus possible that this variant would be observed in a larger study cohort.

The IVS2 -22 A/G polymorphism (dbSNP: rs2233706) results in the change from an A to a G and is located 22 base pairs upstream from exon 3. This variant was found to be in Hardy-Weinberg equilibrium in both populations under investigation ($p>0.05$, Table 12), which correlates with the findings of another study which reported on the genotype and allele frequencies of this variant in the South African population (Bruiners, 2007). No significant difference was observed between the genotype or allele frequencies of the two populations studied ($p>0.05$). Table 12 shows that the A-allele frequency is higher in the Coloured than in the Black population (87% vs 81%) while the G-allele frequency is higher in the Black population (19% vs 13%).

Table 12: Observed genotype and allele frequencies at the IVS2 -22 A/G locus in the total cohort.

		Total Cohort N=307	**Coloured N=219**	**Black N=88**
Genotype frequency	**A/A**	221 (0.72)	164 (0.75)	57 (0.65)
	A/G	81 (0.26)	53 (0.24)	28 (0.32)
	G/G	5 (0.02)	2 (0.01)	3 (0.03)
Allele frequency	**A**	523 (0.85)	381 (0.87)	142 (0.81)
	G	91 (0.15)	57 (0.13)	34 (0.19)

(Coloured vs Black populations Genotype frequencies Chi^2 p-value= 0.096, Allele frequencies Fisher's exact test two-tailed p-value= 0.0589)

The genotype and allele frequencies at the IVS2 -22 A/G locus of *LGALS13* were determined for two subgroups (Adverse and Good pregnancy outcomes respectively) within each population group and are shown in Table 13. No significant differences were observed in the allele or genotype frequencies of the two subgroups within the population groups ($p>0.05$). This finding correlates with that of Bruiners, 2007.

Table 13: Observed genotype and allele frequencies at the IVS2 -22 A/G locus in the study participants with known pregnancy outcomes from the study cohort.

Population group		**Coloured** N=94		**Black** N=42	
Pregnancy outcome		**Adverse** N=23	**Controls** N=71	**Adverse** N=8	**Controls** N=34
Genotype frequency	**A/A**	16 (0.700)	49 (0.700)	5 (0.625)	21 (0.620)
	A/G	7 (0.300)	19 (0.270)	2 (0.250)	12 (0.350)
	G/G	0 (0.000)	3 (0.030)	1 (0.125)	1 (0.030)
Allele frequency	**A**	39 (0.848)	117 (0.824)	12 (0.750)	54 (0.794)
	G	7 (0.152)	25 (0.176)	4 (0.250)	14 (0.206)

(Adverse outcome vs Controls Coloured population Genotype frequencies Chi^2 p-value= 0.589, Allele frequencies Fisher's exact test two-tailed p-value= 0.823; Black population Genotype frequencies Chi^2 p-value= 0.485, Allele frequencies Fisher's exact test 2-tailed p-value=0.738).

The IVS2 -36 G/A variant (Bruiners, 2007) is located 36 base pairs upstream of exon 3, and results in a G to A change. As evident from Table 14, this mutation was only found in the heterozygous genotype (G/A), in this study cohort. This mutation is most likely rare in the South African population studied. It would be interesting to measure the frequency of this variant in other population groups. No significant differences in genotype of allele frequencies were observed between the two population groups in this study (p.> 0.05).

Table 14: Observed genotype and allele frequencies at the IVS2 -36 G/A locus in the total cohort.

		Total Cohort N=309	**Coloured** N=219	**Black** N=90
Genotype frequency	**G/G**	297 (0.96)	210 (0.96)	87 (0.97)
	G/A	12 (0.04)	9 (0.04)	3 (0.03)
	A/A	0 (0.00)	0 (0.00)	0 (0.00)
Allele frequency	**G**	606 (0.98)	429 (0.98)	177 (0.98)
	A	12 (0.02)	9 (0.02)	3 (0.02)

(Coloured vs Black population Genotype frequencies Chi^2 p-value= 0.748, Allele frequencies Fisher's exact test two-tailed p-value=1.00).

Table 15 gives the genotype and allele frequencies of this variant, measured in each subgroup within the two population groups investigated. No association was found between pregnancy outcome and genotype/allele

frequency at this locus in either of the populations (p>0.05). These findings are in agreement with those from a previous study of the genotype and allele frequencies at this locus in pregnant women from the Western Cape of South Africa (Bruiners, 2007).

Table 15: Observed genotype and allele frequencies at the -36 G/A locus in the study participants with known pregnancy outcomes from the study cohort.

Population group		**Coloured** N=93		**Black** N=42	
Pregnancy outcome		**Adverse** N=23	**Controls** N=70	**Adverse** N=8	**Controls** N=34
Genotype frequency	**G/G**	21(0.91)	67 (0.96)	8 (1.00)	33 (0.97)
	G/A	2 (0.09)	3 (0.04)	0 (0.00)	1 (0.03)
	A/A	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Allele frequency	**G**	44 (0.96)	137 (0.98)	16 (1.00)	67 (0.99)
	A	2 (0.04)	3 (0.02)	0 (0.00)	1 (0.01)

(Adverse outcome vs Controls Coloured population Genotype frequencies Chi^2 p-value=0.416, Allele frequencies Fisher's exact test two-tailed p-value= 0.598; Black population Genotype frequencies Chi^2 p-value= 0623, Allele frequencies Fisher's exact test 2-tailed p-value=1.00).

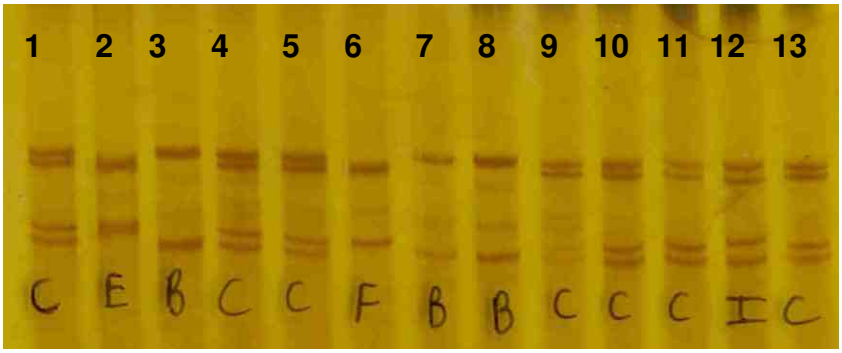


Figure 18: Image of a multiphor gel depicting the conformations of the intronic variants found in exon 3 of *LGALS13* . The genotypes of the different conformations (C, E, B, F and I) are given in Table 16.

Table 16: The different conformations and corresponding genotypes identified using the Multiphor SSCP/HDP system to resolve PCR products from *LGALS13* exon 3.1 (accompanying Figure 18).

Conformation	**Genotype**		
	IVS2 -15 G/A	IVS2 -22 A/G	IVS2 -36 G/A
B	G/G	A/A	G/G
C	G/G	A/G	G/G
E	G/G	G/G	G/G
F	G/G	A/G	G/A
I	G/G	A/A	G/A

These intronic variants might disrupt intronic *cis*-acting elements involved in the splicing regulation of *LGALS13*. This might lead to the generation of defective proteins. Such proteins might lose their functional and structural properties and would most likely also not be detected by the PP13-specific antibody used in the diagnostic kit (Diagnostic Technologies Ltd.). Two alternatively spliced isoforms of PP13 have been documented and published on the Ensembl database, but has up to date not been isolated or further characterized (http://www.ensembl.org/Homo_sapiens/Transcript/SupportingEvidence?db=core;g=ENSG00000105198;r=19:44784556-44789953;t=ENST00000221797). These isoforms are, however, not documented on the NCBI web site (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=retrieve&dopt=full_report&list_uids=29124&log$=databasead&logdbfrom=nucest#tranprod). This warrants a thorough experimental investigation into the existence of these isoforms.

Bioinformatic analyses revealed the presence of splice factor recognition sites within exon 3 (Table 3). The program used for this analysis, ESEfinder (Cartegni *et al*. 2003), searches for splice elements within the exons of a gene and could thus not be used to search for putative intronic splice elements in *LGALS13*. The *in silico* prediction of alternative exons and existence of splice elements is complicated by the degeneracy of splice signals (Thanaraj and Stamm, 2003). An *in silico* analysis of the intronic regions of *LGALS13* would have been useful for the prediction of possible effects the intronic variants

might have on the splice mechanism of this gene. An experimental approach to elucidate the effects of these variants would be the development of a minigene assay. This functional splicing assay has been used to illustrate the effect of seemingly neutral polymorphisms on the splice mechanism of several genes (Baralle and Baralle, 2005; Cooper, 2005).

2.4 Summary of the main findings from Section 2

This study focused on seven sequence variants identified in *LGALS13*. The -98 A/C polymorphism was found to be in Hardy Weinberg equilibrium in both ethnic groups studied ($p>0.05$). This polymorphism might influence the regulation of expression of *LGALS13* due to its location within a region of high sequence conservation between different primates in the putative promoter region of the gene. No associations between genotype/allele frequencies and population group or pregnancy outcome were observed for this locus (Tables 6 and 7).

Two sequence variants located in the coding region of the gene were analysed. The first is the DelT, which results in a deletion of a thymine base at position 229 in exon 3. This mutation leads to a truncated PP13 peptide, 111 amino acids in length (Sammar *et al*. 2006), which might have major implications for the function of this protein. (This finding has been experimentally confirmed by Diagnostic Technologies Ltd., data available on request). The second variant in the coding region is named the “hotspot” mutation and results in a six base pair change within the 11bp region between base pair numbers 137 and 148 in exon 3. It is believed that this mutation might be the result of a non-reciprocal gene conversion event. This will be discussed in section 3.

Four sequence polymorphisms located in the non-coding regions around exon 3 were also analysed. These are: IVS3 +72T/A, IVS2 -15G/A, IVS2 -22A/G and IVS2 -36G/A. All of the intronic polymorphisms, except IVS2 -15 G/A were found to be in Hardy Weinberg equilibrium in both ethnic groups

investigated. The IVS-15G/A mutation was never identified in the study cohort, which indicates that it is a very rare polymorphism. The IVS-36G/A was only found in the heterozygous genotype, no individuals had the A/A genotype, which might indicate that this mutation is very rare in the study cohort (Tables 14 and 15). No associations between pregnancy outcome and the genotype/allele frequencies at these loci were observed in this study ($p>0.05$).

Association analyses between the genotype and allele frequencies of the IVS3 +72T/A variant and the population groups was demonstrated in this study ($p=0.009$, $p=0.0013$). The T/A genotype frequency was higher in the Black population in comparison to the Coloured population.The A-allele frequency was also found to be increased in the Black population (Table 10).

A significant difference between the Adverse and Control groups from the Coloured population was found with regard to the frequency of the A-allele ($p= 0.046$) at this locus. This allele was found to be present at a higher frequency in the Control group than in the Adverse outcome group (16% vs 4%). This might indicate that the A-allele has a protective effect against adverse pregnancy outcome in this population (Table 11).

The above-mentioned intronic sequence polymorphisms might influence the splicing mechanism of *LGALS13* by abolishing or creating novel splice elements. This needs to be further investigated using a minigene system (Cooper, 2005).

3 Clinical data

Part of the aim of this study is to establish a well-characterised study cohort for the collection of research material to be used for the comprehensive assessment of the association between *LGALS13*, PP13 and pre-eclampsia. Research material collection was initiated in this study and included samples for DNA, RNA and protein measurement. All clinical information on the study participants were also gathered and entered into a database for later use. The current study thus forms part of a larger-scale analysis which is ongoing. The serum samples collected are to be sent to Israel, where DTL Ltd. will measure the PP13 levels for this patient cohort. These results would subsequently be used for comparative analyses of different variables (birth weight, BMI, genotypes etc) in the "Good" and "Adverse" pregnancy outcome groups of each population group. The profile of the study cohort, as well as preliminary clinical data and outcomes of pregnancies from this study cohort will be discussed in this section.

3.1 Demographic profile of the study cohort

For this study, sampling was mostly performed in the Paarl region of the Western Cape of South Africa. The majority of the pregnant women participating in this study were sampled at one occasion, during the third trimester of pregnancy. A number of individuals participated in a longitudinal study in addition to this study. These subjects were sampled on four occasions throughout pregnancy. The study cohort consisted of two major population groups, namely the Coloured (70%) and Black (28.8%) populations of the Western Cape. Non-South Africans (0.95%) were omitted from the study (Figure 19). Calculations of mean values of clinical variables as well as genotype and allele frequencies were performed separately for each population group.

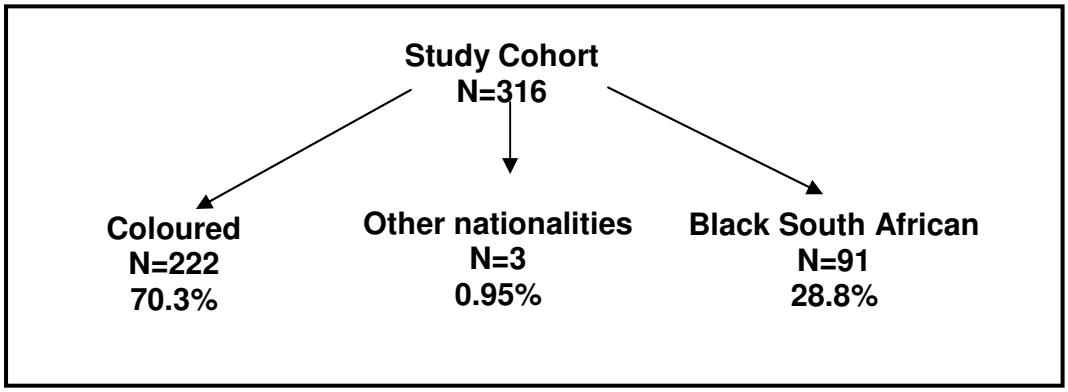


Figure 19: Schematic representation of the ethnicities within the study cohort

The mean values of clinical variables measured during pregnancy, such as BMI of the patient, gestation at delivery and birth weight of the neonate did not differ significantly between the two population groups (Figure 21). However, when looking at the ranges of values obtained for each population group, it can be seen that the coloured population has a wider range of values than the Black population. This observation might, however, be due to the difference in sample size between the two groups. The BMI values ranged between 15 and 45.6 in the Coloured population and between 20 and 40 in the Black population. The gestation at delivery ranged between 27 and 42 weeks in the Coloured, and 36 and 40 weeks in the Black population. The average birth weight was found to be 2.7 kg in the coloured, and 3.4 kg in the Black population. Birth weight ranged between 800g and 4.6 kg in the Coloured and between 2.1 and 4.5kg in the Black population.

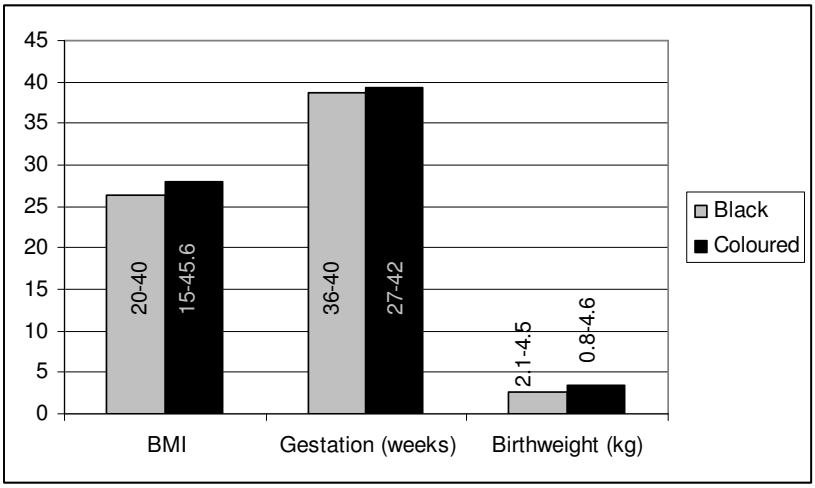


Figure 20: Comparison between the two major population groups in this study cohort. The mean values of certain measurements (BMI, Gestation and Birth weight) taken during pregnancy is shown for the Black (grey bars) and Coloured (black bars) individuals.

A summary of the demographic information available for the study cohort is given in Table 17. A similar trend can be observed in all of the mean values representing demographic and clinical data. The Adverse outcome groups have lower mean values than the Control groups in all of the measurements taken. This study shows, in accordance with Bruiners (2007), that the mean gestational birth weight differs significantly between complicated (adverse outcome) and healthy (good outcome) pregnancies in the population of the Western Cape ($p= 5.81 \times 10^{-7}$). This emphasises the necessity of proper diagnosis and clinical management of complications during pregnancy.

Table 17: Demographic data of the Adverse and Good pregnancy outcome groups in each population.

	Black		**Coloured**	
	Adverse pregnancy outcome N=7	**Good pregnancy outcome N=24**	**Adverse pregnancy outcome N=19**	**Good pregnancy outcome N=67**
BMI (mean value)	27.857 (range: 22-37)	29.586 (range: 20-40)	24.358 (range: 17.6-34.3)	26.809 (range: 17.6-45.6)
Gestation (mean value weeks)	37.375 (range: 36-40)	40.125 (range: 38-42)	35.957 (range:27-39)	39.536 (range: 35-42)
Birth weight (mean value kg)	2.862 (range: 2060-4460)	3.628 (2700-4540)	2.600 (range: 800-3240)	2.680 (range: 1880-4560)

Note: Sample sizes in this Table differ from those in Figure 21 and Table 18 due to data availability.

3.2 Pregnancy outcomes of the study cohort

Out of the 316 individuals recruited, 139 had known pregnancy outcomes at the time of the completion of this study. These subjects were divided into the two major population groups for further analyses (Figure 21). Possible reasons for the 177 subjects with unknown pregnancy outcomes might be that these individuals are still pregnant or lost to follow-up. Some of these individuals might have given birth at another location. Accessibility to hospitals and clinics are limited to some individuals, especially those living in rural areas, due to long distances and transport. These issues pose a health risk to South Africans and should be addressed. These factors also complicate medical research due to the limitation in patient feedback and follow-up consultations.

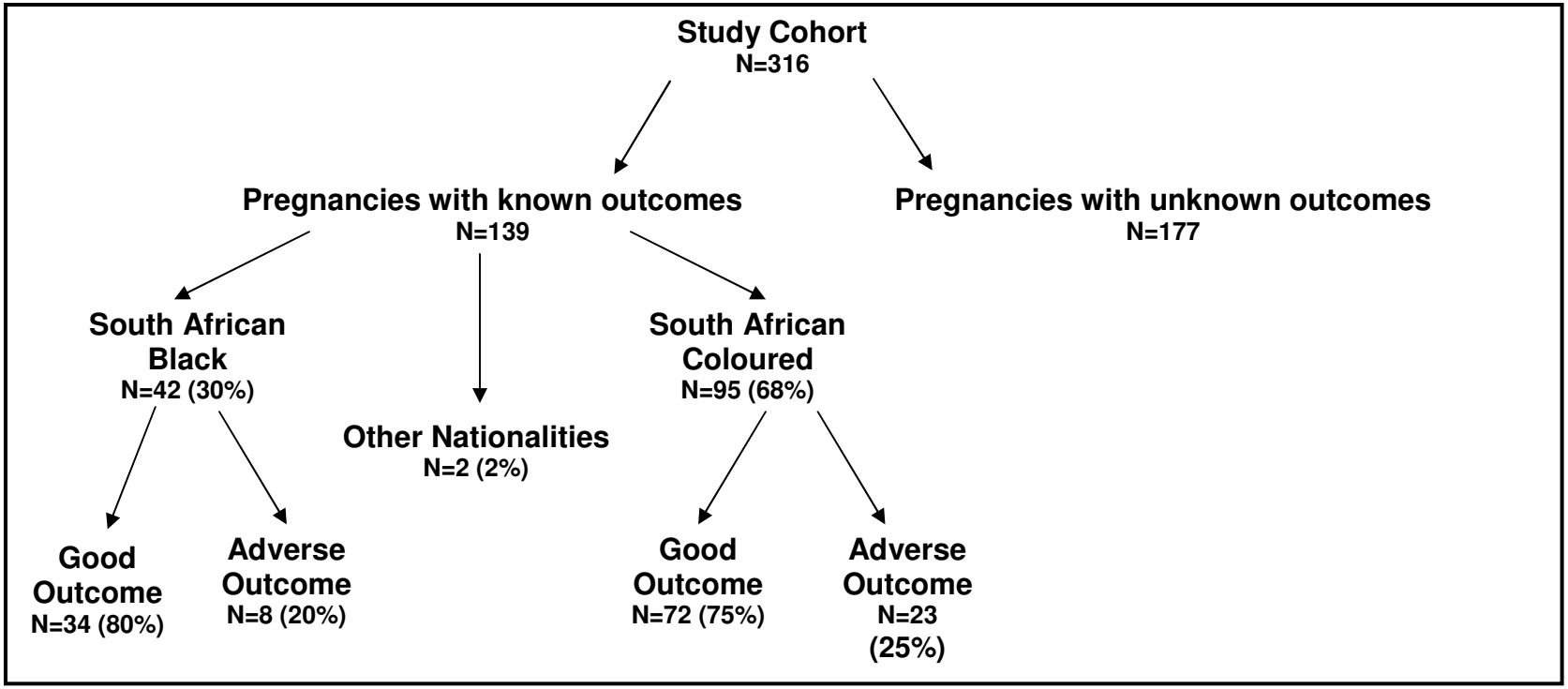


Figure 21: Schematic overview of the outcomes of the pregnancies in the study cohort investigated.

The frequency of pregnancies with adverse outcomes in this study did not differ significantly ($p>0.05$) between the Coloured (24.2%) and Black (19%) populations investigated (Table 16). The incidence of adverse pregnancy outcome observed in the Black population during this study is higher than has been reported by a similar study (Bruiners, 2007). This study found that 22.5% of the Coloured population and 12.2% of the Black population had an adverse pregnancy outcome.

As shown in Table 18, the complication with the highest incidence in the Coloured and Black populations was preterm labour. Hypertension was the second most frequent cause for adverse pregnancy outcome in the Black population. Pre-eclampsia and Rupture of membranes (ROM) had the second highest incidence and hypertension the third highest in the Coloured population.

Other pregnancy complications documented in the Coloured population included isolated cases of placenta previa, intrauterine growth restriction (IUGR), chorioamniotis, vaginal bleeding and low Amniotic Fluid Index (AFI) measurements.

In this study, it was found that preterm labour occurred at a rate of 15% in the total Coloured population. In 2007, the incidence of preterm labour in the Coloured population was reported to be 8.2%. Worldwide, the incidence of preterm delivery is estimated to be between 5 and 11%. It thus seems that this frequency is increasing. This is a serious problem which the South African health system is faced with and should be urgently addressed.

The incidence of pre-eclampsia was found to be 5.3% in the Coloured population. This is lower than was reported by Bruiners, 2007, who found the frequency of PE to be 7.8% in this population. However, this study also found that the incidence of ruptured membranes was 0.7% in the Coloured population (Bruiners, 2007) while the current study shows an incidence rate of 5.3%. Differences between the findings of Bruiners and this study are also

evident in the Black population. The frequency of preterm labour was found to be 10% in this study, while Bruiners found it to be 4.4%. It is, however, important to keep in mind that the sample sizes of this study are smaller that that of Bruiners, 2007 (Coloured N=306 vs N=95, Black N=115 vs N=42). The small sample sizes used for the analyses of pregnancy outcomes therefore limits the statistical value of any observations made.

Table 18: A summary of the pregnancy-related complications and their relative frequencies documented in this study.

Complication	**Number of cases (N)**	
	Black N= 8 (19.0%)	**Coloured N= 23 (24.2%)**
Hypertension	3 (38%)	3 (13.0%)
Pre-eclampsia	1 (12%)	5 (21.7%)
Preterm labour	4 (50%)	14 (60.9%)
Intrauterine death	0	1 (4.3%)
Early neonatal death	0	1(4.3%)
Other	2 (25%)	9 (38%)

Note: Some of the complications listed above occurred simultaneously in individual cases.

In South Africa, complications due to hypertension in pregnancy remain the greatest direct cause of maternal death, and the second most common primary cause of death (National Committee on Confidential Enquiries into Maternal Deaths in the office of the Minister of Health, 2004). According to the Confidential Enquiries into maternal deaths, hypertensive disorders of pregnancy are avoidable causes of death with known methods of prevention. Close antenatal surveillance is very important for the proper management of such pregnancy-related disorders. Many risk factors for the development of pregnancy-induced hypertension have been identified. These include raised body mass index, smoking and a history of pre-eclampsia. Lifestyle-associated risk factors such as smoking and obesity could also be addressed by educating and informing patients of these risk factors.

In order to provide women at high risk of developing pre-eclampsia with the necessary antenatal care and treatment strategies before the onset of the disease, it is thus crucial to diagnose them during the first trimester (Papageorghiou and Campbell, 2006). Improvement of clinical trial design could aid in the identification of biomarkers to predict maternal risk of

developing complications during pregnancy (Chafetz *et al*., 2007). PP13 could serve as an early biomarker for the detection of various pregnancy complications (Burger *et al*., 2004). This diagnostic test could be routinely integrated into the clinical setting by measuring circulating PP13 levels during the first trimester and subsequently predict the risk of developing adverse pregnancy outcomes (Chafetz *et al*., 2007). Patients identified in this way could thereby benefit from close antenatal surveillance and care.

3.3 Summary of findings from Section 3

The study cohort consisted of two major ethnic groups, the Coloured (70%) and Black (28.8%) populations of the Western Cape. The mean values of clinical variables such as BMI of the patient, gestation at delivery and birth weight of the neonate did not differ significantly between these two population groups or between the Adverse and Good pregnancy outcome groups. Of the 316 individuals recruited, 139 had known pregnancy outcomes at the time of the completion of this study. Possible reasons for the 177 subjects with unknown pregnancy outcomes might be that these individuals are still pregnant or lost to follow-up.

Some of the common pregnancy complications identified in this study cohort were preterm labour, pre-eclampsia, hypertension and rupture of membranes. One intrauterine death and one early neonatal death occurred. The frequency of pregnancies with adverse outcomes did not differ significantly between the Coloured (24.2%) and Black (19.0%) populations investigated in this study. The complication with the highest incidence rate in this study population was preterm labour, occurring in 60% of the complicated pregnancies in the Coloured population, and in 50% of the complicated pregnancies in the Black population. It is thus evident from this investigation that preterm birth remains a major health problem in South Africa and needs to be addressed. The early prediction of adverse pregnancy outcome can assist in the effective management thereof. PP13 could aid in first trimester predictive diagnostic testing for the risk of adverse pregnancy outcome.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

PP13 has been shown to be decreased during the first trimester of pregnancy, in the serum of women who later develop pre-eclampsia. This protein is currently being used as a predictive biomarker for the early diagnosis of pre-eclampsia. The function of this protein has, to date, not been completely elucidated. Many questions regarding the role PP13 plays during pregnancy and in the development of pre-eclampsia remain unanswered. Is the reduction in PP13 levels during the first trimester of pre-eclamptic pregnancies a cause or effect of the disease? Do DNA polymorphisms in *LGALS13* affect the expression or function of PP13? Could these polymorphisms possibly cause the observed reduction in PP13 levels in pre-eclampsia? To further analyse and attempt answering some of these questions, the gene encoding PP13, *LGALS13*, was characterised in this study. This was achieved using a comprehensive range of *in silico* tools for molecular analyses.

PP13 was first analysed within the context of the galectin protein family to which it belongs. From the data gathered, it can be concluded that PP13 shares great sequence homology with other members (such as Galectin 10 and -14) of the subfamily of galectin proteins into which it has been classified. However, this study shows that PP13 is most homologous to the predicted protein product of the putative gene, LOC148003, located upstream from the *LGALS13* gene on chromosome 19. This has implications for the specificity of the PP13-based predictive ELISA kit used for the early diagnosis of pre-eclampsia. It is crucial to establish whether this putative gene is expressed in humans and whether the protein product it encodes is recognised by the PP13-specific antibody in the diagnostic ELISA test.

Secondly, the architecture of the gene, *LGALS13*, was analysed. This study has shown that *LGALS13* has a non-classical promoter structure. A number of putative *cis*-elements were identified in the predicted promoter region of this gene, using *in silico* tools and phylogenetic footprinting. One of these regulatory elements (AP-1) predicted to be present in the putative promoter region of *LGALS13*, is of particular interest due to its possible function as a mediator between Estrogen Receptor α (ERα) and its corresponding Estrogen Responsive element (ERE). Several studies have reported associations with

polymorphisms in Estrogen Receptor genes and pre-eclampsia (Molvarec *et al*. 2007; Maruyama *et al*. 2004). It has been postulated that low Estrogen levels are necessary for normal trophoblast invasion of the maternal spiral arteries early during primate pregnancy, and that the increase in Estrogen levels later during pregnancy might act as a suppressive force against further trophoblast invasion (Albrecht *et al*. 2006). Pre-eclampsia has been associated with a reduction in Estrogen levels (Innes and Byers, 1999; Zeisler *et al*. 2002). Another study has shown that Estrogen levels are reduced between 30 and 36 weeks of gestation in women living at high altitudes, who developed pre-eclampsia as opposed to those who did not have the disorder (Zamudio *et al*. 1994). If Estrogen levels are reduced during pre-eclampsia, the putative Estrogen Responsive element in the promoter region of *LGALS13* would not be activated. If this gene is dependent on this element for transcriptional activity, this would lead to a reduction in *LGALS13* and PP13 expression. The existence and activity of this element has, however, not been confirmed. Functional promoter studies are underway and would, hopefully, aid in the understanding of *LGALS13* regulation.

Thirdly, the regulation of *LGALS13* was investigated. According to the *in silico* analyses performed in this study, the putative promoter region of *LGALS13* (defined by the genomic region ~1000bp upstream of the start codon of PP13) lacks CpG islands which indicates that this gene is most likely not regulated by methylation. This does, however, not rule out the possibility of *LGALS13* being imprinted. Genomic imprinting involves the parent-of-origin-specific silencing of a gene. Several genes involved in extra-embryonic tissue development have been shown to be genomically imprinted. It has also been postulated that pre-eclampsia is caused by a mutation in the active copy of an imprinted autosomal gene (Graves, 1998; Oudejans *et al*. 2004). Furthermore, it has been found that placental genes which are genomically imprinted are often not regulated via methylation but by histone modifications and non-coding RNA species (Wagschal and Feil, 2006). In future studies, it could be investigated whether *LGALS13* is genetically imprinted. This could be done by establishing whether paternally and maternally derived alleles are expressed

in heterozygous placental tissue samples. This could be performed using tissue collected in this study.

The transcriptome provides another level of investigation of the molecular differences between normal and pre-eclamptic pregnancies. Possible biomarkers for the early detection of pre-eclampsia might be present in the RNA of pregnant individuals. Such biomarkers could be non-coding RNA species, transcription factors or epigenetic features which cannot be measured during proteome analysis (Smets *et al*. 2006). The transcriptome also gives a more thorough view of the expression profiles of genes which are possibly associated with multifactorial pregnancy-related disorders such as pre-eclampsia. Some studies have been aimed at a comparison between the global gene expression profiles of pre-eclamptic and healthy pregnancies (Nishizawa *et al*. 2006; Zhou *et al*. 2006). In these studies, the gene expression profiles of human placentae were studied using genome-wide oligonucleotide microarrays. Several genes, with a range of metabolic and cellular functions, were subsequently shown to be differentially expressed in pre-eclampsia. Many of the genes, shown to be differentially expressed, have previously been associated with the pathologies of pre-eclampsia. Such comparative genome-wide gene expression studies are made possible by advances in microarray and quantitative real-time PCR technologies. Although these studies provide a more time-efficient measurement of differential gene expression in pre-eclampsia, its multifactorial nature warrants a broader investigation into the underlying genomic, transcriptomic, proteomic and metabolimic changes which occur in this disorder. Some studies have aimed at establishing comparative gene expression profiles of metabolism-related genes between pre-eclamptic and healthy individuals (Pang and Xing, 2004). For future studies, it might be useful to target certain biological processes associated with the development of pre-eclampsia, and to investigate alterations on the DNA, RNA, protein and metabolite levels between affected and healthy individuals. Biological processes which could be studied include immunological processes, lipid, carbohydrate and oxygen metabolism (Nishizawa *et al*. 2006).

The sequence polymorphisms which have been identified in *LGALS13* might influence the expression of the gene and the structure and function of the resulting peptide. The majority of the sequence polymorphisms are located in highly conserved, functionally and structurally important regions of the gene. Whether these polymorphisms do, in fact, influence the expression and function of PP13, remains to be elucidated. This would be ascertained once the PP13 measurements have been performed by DTL.

Statistically significant differences were observed in the genotype and allele frequencies of the IVS3 +72 T/A variant between the Coloured and Black populations of the Western Cape. The allele frequency of the mutated allele (A) was also found to be higher in the subgroup with good pregnancy outcomes (Controls) in this study, in comparison to the Adverse outcome group. It is possible that this allele has a protective effect on the bearer, meaning individuals who carry this allele are perhaps less likely to develop adverse pregnancy outcomes than those who do not have the allele.

The variant located in the putative promoter region of *LGALS13* occurs in a region of high sequence conservation, suggesting that this region has an important regulatory function. This observation needs to be further analysed by experimental procedures such as site-directed mutagenesis and Luciferase gene-reporter assays.

Furthermore, it needs to be established whether the two exonic sequence variants, identified in *LGALS13*, are functionally relevant. This could be done by protein expression studies. It would also be interesting to investigate whether these variants are expressed in individuals who bear these mutations and whether functional mRNA and protein molecules are produced. The bioinformatic analyses performed in this study suggested that the exonic variants do not abolish any splice enhancer sites predicted to be present in the coding regions of *LGALS13*.

It is possible that the intronic sequence polymorphisms of *LGALS13* affects the gene's splicing mechanisms by creating or abolishing splice elements in

the genomic DNA sequence. This has to date not been proven by *in silico* analyses or the presence of alternatively spliced isoforms of PP13. The variable nature of intronic sequence elements, as well as the vast range of motifs characterised as such elements, complicate the use of *in silico* based approaches to search for such elements in a DNA sequence. Functional splice mechanism assays, such as the minigene system, are currently the most effective way to study alternative and defective splicing in genes (Cooper, 2005). It is recommended that future studies use this approach to elucidate whether the intronic polymorphisms found in *LGALS13* affect the splicing mechanism of this gene.

The use of PP13 as a predictive biomarker for pre-eclampsia warrants a comprehensive study of this peptide. In this study, ~312 pregnant subjects were genotyped for the known variants in *LGALS13*. Plasma and serum samples were also obtained from study participants. A comparison of the genotypes and PP13 concentrations in the study cohort could give an indication of whether the sequence polymorphisms in *LGALS13* influence the expression of PP13. This would provide information on the i) the efficacy of the ELISA kit (produced by DTL) for the early diagnosis of pre-eclampsia in the South African population and ii) the possible role of PP13 during pregnancy and the development of pre-eclampsia. Due to time constraints, serum samples collected during this study were not analysed for PP13 concentrations. This is an urgent priority and is to be completed in the near future.

One of the major obstacles encountered in this study was that PP13 is almost exclusively expressed in placental tissue. *LGALS13* RNA species can only be obtained from placental tissue samples, which restricts sample collection to a narrow window period directly after delivery of the placenta. Since it is not possible to collect placental samples throughout pregnancy, gestational differences in *LGALS13* gene expression cannot be investigated. For gene expression studies focusing on placental samples, a well-coordinated infrastructure is necessary. Placental samples have to be snapfrozen and stored at -80ºC to ensure that placental RNA species are not degraded before

total RNA extraction and downstream applications are performed. Logistically, this was problematic since the hospital where placental samples were collected is situated approximately 30 km away from the laboratories where experimental procedures were carried out. For future studies, it is recommended that a laboratory, equipped with the necessary tools, should be set up in close proximity to the sampling station.

It is of great importance to document all clinical and demographic data of the entire study cohort. Information on certain risk factors for the development of pre-eclampsia, such as smoking, age, parity and gravidity, would also be of value if obtained for each study subject.

This study has clearly demonstrated the use of an *in silico* based approach to molecularly characterise *LGALS13*, thereby directing the focus of future experimental work. These suggested experiments are to be performed using samples obtained from the newly established study cohort. This would aid in gaining a better knowledge of the role *LGALS13* and PP13 play in the early prediction of pre-eclampsia.

REFERENCES

Albrecht, E., Bonagura, T., Burleigh, D., Enders, A., Aberdeen, G., Pepe, G. (2006) Suppression of extravillous trophoblast invasion of uterine spiral arteries by Estrogen during early baboon pregnancy. *Placenta* **27** (4-5): 483-490.

Aldrich, C., Verp, M.S., Walker, M.A., Ober, C. (2000) A null mutation in *HLA-G* is not associated with pre-eclampsia or intrauterine growth restardation. *Journal of Reproductive Immunology* **47**: 41-48.

Alsat, E., Guibourdenche, J., Couturier, A., Evain-Brion, D. (1998) Physiological role of human growth hormone. *Molecular and Cellular Endocrinology* **140** (1-2) 121-127.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yeh, L.L. (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research* **32**: 115-119.

Arias, F., Rodriquez, L., Rayne, S.C., Kraus, F.T. (1993) Maternal Placental Vasculopathy and Infection: two distinct subgroups among patients with preterm labour and preterm ruptured membranes. *American Journal of Obstetrics and Gynecology* **168** (2): 585-591.

Ars, E., Serra, E., García, J., Kruyer, H., Gaona, A., Lázaro, C., Estivill, X. (2000) Mutations affecting mRNA splicing are the most common molecular defects in patients with Neurofibromatosis type 1. *Human Molecular Genetics* **9** (2): 237-247.

Austgulen, R., Arntzen, K.J., Vatten, L.J., Kahn, J., Sunde, A. (1995) Detection of cytokines (interleukin-1, interleukin-6, transforming growth factor-β) and soluble tumour necrosis factor receptors in embryo culture fluids during in-vitro fertilization. *Human Reproduction* **10** (1): 171-176.

Baralle, D., Baralle, M. (2005) Splicing in action: assessing disease causing sequence changes. *Journal of Medical Genetics* **42**: 737-748.

Barondes, S. H., Cooper D. N. W., Gitt, M.A., Leffler, H. (1994) Galectins structure and function of a large family of animal lectins . *The Journal of Biological Chemistry* **269** (33): 20807-20810.

Baumann, M.U., Bersinger, N.A., Surbek, D.V. (2007) Serum markers for predicting pre-eclampsia. *Molecular Aspects of Medicine* **28** (2): 227-244.

Benvenuto, G., M. L. Carpentieri, Salvatore, P., Cindolo, L., Bruni, C.B., Chiarotti, L. (1996) Cell-Specific Transcriptional Regulation and Reactivation of Galectin-1 Gene Expression Are Controlled by DNA Methylation of the Promoter Region. *Molecular and Cellular Biology* **16**(6): 2736–2743.

Benyo, D.F., Miles, T.M., Conrad, K.P. (1997) Hypoxia Stimulates Cytokine Production by Villous Explants from the Human Placenta. *Journal of Clinical Endocrinology and Metabolism* **82** (5): 1582-1588

Bermingham, J., Jenkins, D., McCarthy, T., O'Brien, M. (2000) Genetic analysis of insulin-like growth factor II and HLA-G in pre-eclampsia. *Biochemistry Society* **28** (2): 215-219.

Blom, N., Gammeltoft, S., Brunak, S (1999) Sequence and structure based prediction of eukaryotic protein phosphorylation. *Journal of Molecular Biology* **294**(5): 1351-1362.

Bruiners, N. (2007) Molecular genetic analysis of preterm labour. Unpublished MSc Thesis, University of Stellenbosch, Department of Genetics.

Burger, O., E. Pick, Zwickel, J., Klayman, M., Meiri, H., Slotky, R., Mandel, S., Rabinovitch, L., Paltieli, Y., Admon, A., Gonen, R. (2004) Placental Protein 13 (PP-13): Effects on Cultured Trophoblasts, and Its Detection in Human Body Fluids in Normal and Pathological Pregnancies. *Placenta* **25**: 608-622.

Caniggia, I. (2000) Oxygen and placental development during the first trimester: implications for the pathophysiology of pre-eclampsia. *Placenta* **21**: 25-30.

Cartegni, L., Chew, S.L., Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics* **3** 285-298.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., Krainer, A.R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Research* **31**: 3568-3571.

Chafetz I, Kuhnreich I, Sammar M, Tal Y, Gibor Y, Meiri H, Cuckle H, Wolf M. (2007) First-trimester placental protein 13 screening for preeclampsia and intrauterine growth restriction. *American Journal of Obstetrics and Gynecology* **197** (1): 1-7.

Chiariotti, L., Salvatore, P., Frunzio, R., Bruni C.B. (2004) Galectin genes: Regulation of expression. *Glycoconjugate Journal* **19**: 441–449.

Chomczynski, P., Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162** (1): 156-159.

Cooper, D. N. W., Barondes, S.H. (1999) God must love galectins; He made so many of them. *Glycobiology* **9**(10): 979-984.

Cooper, D. N. W. (2002) Galectinomics: finding themes in complexity. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1572**: 209– 231.

Cooper, T.A. (2005) Use of minigene systems to dissect alternative splicing elements. *Post-transcriptional Regulation of Gene Expression* **37** (4): 331-340.

Crocker, I. (2004) Placentation. In: Baker, P.N.K., Kingdom, J.C.P. *Pre-eclampsia-Current Perspectives on Management*. New York, Parthenon Publishing**:** 81-88.

Davey, D.A., MacGillivray, I. (1988) The classification and definition of the hypertensive disorders of pregnancy. *American Journal of Obstetrics and Gynecology* **158**: 892-898.

Dekker, A.D., Sibai, B.M. (1999) The immunology of pre-eclampsia. *Seminars in Perinatology* **23** (1): 24-33.

Dekker, G., Robillard, P. (2007) Pre-eclampsia: Is the immune maladaption hypothesis still standing? An epidemiological update. *Journal of Reproductive Immunology* **76** (1-2): 8-16, Proceedings of the 5th International Workshop on Reproductive Immunology, Immunological Tolerance and Immunology of Preeclampsia.

Dickerson, R.E.; Drew, H.R.; Conner, B.N.; Wing, R.M.; Fratini, A.V. and Kopka, M.L. (1982) The anatomy of A-,B-, and Z-DNA. *Science* **216** (4545): 475-485.

Duley, L. (2003) Pre-eclampsia and the hypertensive disorders of pregnancy. *British Medical Bulletin* **67**: 161-176.

Duley, L., Henderson-Smart, D.J., Meher, S., King, J.F. (2007) Antiplatelet agents for preventing pre-eclampsia and its complications (Review) *The Cochrane Library* **4**: 1-117.

Ebrashy, A., Ibrahim, M., Marzook, A., Yousef, D. (2005) Usefulness of Aspirin Therapy in High-risk Pregnant Women with Abnormal Uterine Artery Doppler Ultrasound at 14-16 Weeks Pregnancy: Randomized Controlled Clinical Trial. *Croation Medical Journal* **46**(5): 826-831.

El Hassan, M.A.; Calladine, C.R. (1996) Propeller-twisting of base-pairs and the confomration mobility of dinucleotide steps in DNA. *Journal of Molecular Biology* **259** (1): 95-103.

Esplin, M.S., Fausett, M.B., Fraser, A., Kerber, R., Mineau, G., Carrillo, J., Varner, M.J. (2001) Paternal and maternal components of the predisposition to preeclampsia. *The New England Journal of Medicine* **344** (12): 867-872.

Fairbanks, D.J., Andersen, W.R. (eds) (1999) *Genetics: the continuity of life*. U.S.A., Brooks/Cole Publishing Company: 20-100.

Gaudin, J., Arar, C., Monsigny, M., Legrand, A. (1997) Modulation of the expression of the rabbit galectin-3 gene by p53 and c-Ha-ras proteins and PMA. *Glycobioloby* **7**(8): 1089-1098.

Genbacev, O., Zhou, Y., Ludlow, J.W., Fisher, S.J. (1997) Regulation of human placental development by oxygen tension. *Science* **277** (5332): 1669-1672.

Goldman-Wohl, D., Yagel, S. (2000) Examination of distinct fetal and maternal molecular pathways suggests a mechanism for the development of preeclampsia. *Journal of Reproductive Immunology* **76** (1-2): 54-60.

Gonen, R.; Shahar, R.; Grimpel, Y.; Chefetz, I.; Sammar, M.; Meiri, H.; Gibor, Y. (2006) Placental protein 13 (PP13) as an early marker for pre-eclampsia: a prospective longitudinal study. Presented at: *The 26th*

Annual Meeting of the Society of Maternal Fetal Medicine, Miami Beach, Florida.

Graham, C., Keth, R. (1996) Fetus-Placenta-Newborn: Altered expression of gelatinase and surface-associated plasminogen activator activity by trophoblast cells isolated from placentas of pre-eclamptic patients. *American Journal of Obstetrics and Gyneacology* **175** (3): 555-562.

Gratacós, E. (1998) Lipid-mediated endothelial dysfunction: a common factor to preeclampsia and chronic vascular disease. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **92** (1): 63-66.

Graves, J.A. (1998) Genomic imprinting, development and disease-is pre-eclampsia caused by a maternally imprinted gene? *Reproduction, Fertility and Development* **10**(1): 23-29.

Hayman, R., Myers, J. (2004) Definition and Classification. In: Baker, P.N.K., Kingdom, J.C.P *Pre-eclampsia: Current Perspectives on Management.* New York, Parthenon Publishing: 11-13.

Hendy, M.D., Penny, D., Steel, M.A. (1994) A discrete Fourier analysis for evolutionary trees. *Proclamations of the National Academy of Science U.S.A.* **91**: 3339-3343.

Hiby, S.E., King, A., Loke, Y.W., Sharkey, A. (1999) Molecular studies of trophoblast HLA-G: polymorphism, isoforms, imprinting and expression in preimplantation embryo. *Tissue Antigens* **53**: 1-13.

Hiby, S.E., Walker, J.J., O'Shaughnessy, K.M., Redman, C.W.G., Carrington, M., Trowsdale, J., Moffet, A. (2004) Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive stress. *The Journal of Experimental Medicine* **200** (8): 957-965.

Hims, M.M., Ibrahim, E.C., Leyne, M., Mull, J., Liu, L., Lazaro, C., Shetty, R., Gill, S.,Gusella, J.F., Reed, R., Slaugenhaupt, S.A. (2007) Therapeutic potential and mechanism of kinetin as a treatment for the human splicing disease familial dysautonomia. *Journal of Molecular Medicine* **85** (2): 149-161.

Hubel, C.A. (1999) Oxidative Stress in the Pathogenesis of Preeclampsia. *Proceedings of the Society for Experimental Biology and Medicine* **222**: 222-235.

Huelsenbeck, J.P., Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17 (8): 754-755.

Huppertz, B., Kingdom, J., Caniggia, I., Desoye, G., Black, S., Korr, H., Kaufmann, P. (2003) Hypoxia favours apoptotic shedding of placental syncytiotrophoblast into the maternal circulation. *Placenta* **24**: 181-190.

Huppertz, B., Meiri, H., Bartz, C.I., Marei, S. (2005) Dramatic increase in villous explants of preeclamptic placentas *in vitro*. *Placenta* **26**: A53-A53.

Innes, K.E., Byers, T.E. (1999) Pre-eclampsia and breast cancer risk. *Epidemiology* **10**: 722-732.

Jacquemin, P., Martial, J.A., Davidson, I. (1997) Human TEF-5 is preferentially expressed in placenta and binds to multiple functional elements of the human chorionic somatomammotropin-B gene enhancer. *The Journal of Biological Chemistry* **272** (20): 12928-12937.

Kankainen, M., Pehkonen, P., Rosenstöm, P., Törönen, P., Wong, G., Holm, L. (2006) POXO: a web-enabled tool series to discover transcription factor binding sites. *Nucleic Acids Research* **34**: W534-W540.

Kenny, L. (2004) Trophoblast invasion. In: Baker, P.N.K., Kingdom, J.C.P. *Preeclampsia- current perspectives on management.* New York, Parthenon Publishing: 63-70.

Kozak, M. (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Research* **9** (20): 5233-5252.

Leffler, H., Carlsson, S., Hedlund, M., Qian, Y., Poirier, F. (2004) Introduction to galectins. *Glycoconjugate Journal* **19**: 433-440.

Le Gal, F.A., Riteau, B., Sedlik, C., Khalil-Daher, I., Menier, C., Dausset, J., Gullet, J., Carosella, E.D., Rouas-Freiss, N. (1999) HLA-G mediated inhibition of antigen-specific cytotoxic T lymphocytes. *International Immunology* **11** (8): 1351-1356.

Lietchi-Gallati, S., Schneider, V., Neeser, D., Kraemer, R. (1999) Two buffer PAGE system-based SSCP/HD analysis: a general protocol for rapid and sensitive mutation screening in cystic fibrosis and any other human genetics disease. *European Journal of Human Genetic Disease* **7**: 590-598.

Loubser O, Marais AD, Kotze MJ, Godenir N, Thiart R, Scholtz CL, de Villiers JN, Hillermann R, Firth JC, Weich HF, Maritz F, Jones S, van der Westhuyzen DR (1999) Founder mutations in the LDL receptor gene contribute significantly to the familial hypercholesterolemia phenotype in the indigenous South African population of mixed ancestry. *Clinical Genetics*. **55**: 340-345.

Lyall, F. (2007) Development of the utero-placental circulation: purported mechanisms for cytotrophoblast invasion in normal pregnancy and pre-eclampsia. In: Lyall, F. and Bellfort, M. *Pre-eclampsia: Etiology and Clinical practice*. Cambridge University Press, New York: 20-31.

Martin, J.A., Hamilton, B.E., Sutton, P.D., Ventura, S.J., Menacker, M.A.F., Kirmeyer, S. (2006) Births: final data for 2004. *National Vital Statistics Reports* **55** (1):1-102.

Martin, U., Davies, C., Hayavi, S., Hartland, A., Dunne, F. (1999) Is normal pregnancy atherogenic? *Clinical Science* **96**: 421-425.

Maruyama, A., Nakayama, T., Sato, N., Mizutani, Y., Furuva, K., Yamamoto, T. (2004) Association study using single nucleotide polymorphisms in the estrogen receptor-β (ESR2) gene for pre-eclampsia. *Hypertension Research* **27** (12): 903-909.

Maynard, S.E., Min, J., Merchan, J., Lim, K., Li, J., Mondal, S., Libermann, T.A., Morgan, J.P., Selke, F.W., Stillman, I.E., Epstein, F.H., Sukhatme, V.P., Karumanchi, S.A. (2003) Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension and proteinuria in preeclampsia. *The Journal of Clinical Investigation* **111** (5): 649-658.

Melczer, Z., Bánhidy, F., Csömör, S. Jr, Tóth, P., Kovács, M., Winkler, G., Cseh, K. (2003) Influence of leptin and the TNF system on insulin resistance in pregnancy and their effect on anthropometric parameters of newborns. *Acta Obstetrica et Gynecologica Scandinavica* **82**: 432-438.

Miller, S.A., Dykes, D.D., Polesky, H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* **16**(3): 1215.

Moffet, A., Hiby, S. (2007) Immunological factors and placentation: implications for pre-eclampsia. In: Lyall, F. and Bellfort, M. *Pre-eclampsia: Etiology and Clinical practice*. Cambridge University Press, New York: 92-99.

Molina, C., Grotewold, E. (2005). Genome wide analysis of Arabidopsis core promoters. *BMC Genomics* **6** (25): 1-12.

Molvarec, A., Ver, A., Fekete, A., Rosta, K., Derzbach, L., Derzsy, Z., Karadi, I., Rigo, J. (2007) Association between estrogen receptor α (ESR1) gene polymorphisms and severe pre-eclampsia. *Hypertension Research* **30** (3): 205-211.

Myers, J., Brockelsby, J. (2004) The epidemiology of pre-eclampsia. In: Baker, P. N. K., Kingdom, J.C.P. *Pre-eclampsia- Current Perspectives on Management*. New York, Parthenon Publishing**:** 25.

Need J.A., Bell B., Meffin E., Jones W.R. (1983) Pre-eclampsia in pregnancies from donor inseminations. *Journal of Reproductive Immunology* **5**: 329-338

Nicolaides, K. H., R. Bindra, Turan, O.M., Chefetz, I., Sammar, M., Meiri, H., Tal, J., Cuckle, H.S. (2006) A novel approach to first-trimester screening for early pre-eclampsia combining serum PP-13 and Doppler ultrasound. *Ultrasound in Obstetrics and Gynecology* **27**(1): 13-17.

Nishizawa, H., Pryor-Koishi, K., Kato, T., Kowa, H., Kurahashi, H., Udagawa, Y. (2006) Microarray analysis of differentially expressed fetal genes in placental tissue derived from early and late onset severe pre-eclampsia. *Placenta* **28** (5-6): 487-497.

Norwitz, E. R. (2006) Defective implantation and placentation: laying the blueprint for pregnanoy complications. *Reproductive BioMedicine Online* **13**(4): 591-599

Ohler, U. (2006) Identification of core promoter modules in Drosophila and their application in accurate transcription start site prediction. Nucleic *Acids Research* **34** (20): 5943-5950.

Oudejans, C.B.M., Mulders, J., Lachmeijer, A.M.A., van Dijk, M., Könst, A.A.M., Westerman, B.A., van Wijk, I.J., Leegwater, P.A.J., Kato, H.D., Matsuda, T., Wake, N., Dekker, G.A., Pals, G., Kate, L.P., Blankenstein, M.A. (2004) The parent-of-origin effect of 10q22 in pre-

eclamptic females coincides with two regions clustered for genes with down-regulated expression in androgenetic placentas. *Molecular Human Reproduction* **10** (8): 589-598.

Pang, Z.J., Xing, F.Q. (2004) Comparative profiling of metabolism-related gene expression in pre-eclamptic and normal pregnancies. *Archives of Gynecology and Obstetrics* **269** (2): 91-95.

Papadakis, M.N., Patrinos, G.P. (1999) Contribution of gene conversion in the evolution of the human β-like globin gene family. *Human Genetics* **104:** 117-125.

Papageorghiou, A. T., Campbell, S. (2006) First trimester screening for preeclampsia. *Current Opinion in Obstetrics and Gynecology* **18**: 594-600.

Pavesi, G., Mereghetti, P., Mauri, G., Pesole, G. (2007) Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Research* **32** Web Server Issue: 199-203.

Pavesi, G., Mereghetti, P., Zambelli, F., Stefani, M., Mauri, G., Pesole, G. (2006) MoD Tools: regulatory motif discovery in nucleotide sequences from co-regulated or homologous genes. *Nucleic Acids Research* **34**: 566-570.

Pavesi, G., Zambelli, F., Pesole, G. (2007) WeederH: an algorithm for finding conserved regulatory motifs and regions in homologous sequences. *BMC Bioinformatics* **8**: 46

Pennacchio, L.A.; Rubin, E.M. (2001) Genomic strategies to identify mammalian regulatory sequences. *Nature Reviews Genetics* **2**: 100-109.

Perillo, N. L., Marcus, M.E., Baum, L.G. (1998) Galectins: versatile modulators of cell adhesion, cell proliferation, and cell death. *Journal of Molecular Medicine* **76**: 402–412.

Pijenborg, R., Vercruysse, L., Hanssens, M., Van Assche, F.A. (2007) Trophoblast invasion in pre-eclampsia and other pregnancy disorders. In: Lyall, F. and Bellfort, M. *Pre-eclampsia: Etiology and Clinical practice*. Cambridge University Press, New York: 3-13.

Plasencia, W., Maiz, N., Bonino, S., Kaihura, C. (2007) Uterine artery Doppler at 11+0 to 13+6 weeks in the prediction of pre-eclampsia. *Ultrasound in Obstetrics and Gynecology* **30** (5): 742-749.

Posada, D., Crandall, K.A., Holmes, E.C. (2002) Recombination in evolutionary genomics. *Annual Review of Genetics* **36**: 75-97.

Rabinovich, G. A., Baum, L.G., Tinari, N., Paganelli, R., Natoli, C., Liu, F., Iacobelli, S. (2002) Galectins and their ligands: amplifiers, silencers or tuners of the inflammatory response? *Trends in Immunology* **23**(6): 313-320.

Rajakumar, A., Whitlock, K.A., Weissfeld, L.A., Daftary, A.R., Markovic, N., Conrad, K.P. (2001) Selective overexpression of the hypoxia-inducible transcription factor, *HIF-2α*, in placentas from women with preeclampsia. *Biology of Reproduction* **63**: 499-506.

Rebelo, I., Carvalho-Guerra, F., Pereira-Leite, L., Quintanilha, A. (1995) Lactoferrin as a sensitive blood marker of neutrophils activation in normal pregnancies. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **62:**189-194.

Redman, C.W.G. (1991) Current topic: Pre-eclampsia and the placenta. *Placenta* **12**: 301-308.

Redman, C.W.G., Sacks, G.P., Sargent, I.L. (1999) Preeclampsia: an excessive maternal inflammatory response to pregnancy. *American Journal of Obstetrics and Gynecology* **180**(2): 499-506.

Redman, C.W.G., Sargent, I.L. (2000) Placental debris, oxidative stress and pre-eclampsia. *Placenta* **21:** 597–602.

Redman, C.W.G., Sargent, I.L. (2007) Immunological factors and placentation: implications for pre-eclampsia. In: Lyall, F. and Bellfort, M. *Pre-eclampsia: Etiology and Clinical practice*. Cambridge University Press, New York: 103-115.

Reik, W., Walter, J. (2001) Genomic imprinting: parental influence on the genome. *Nature Reviews Genetics* **2**: 21-32.

Reister, F., Kingdom, J. C. P (2004) Screening for pre-eclampsia. *Pre-eclampsia: Current Perspectives on Management*. P. N. Baker and J. C. P. Kingdom. New York, Parthenon Publishing**:** 119-132.

Roberts, J.M., Cooper, D.N.W. (2001) Pathogenesis and genetics of pre-eclampsia. *The Lancet* **357** (9249): 53-56.

Roberts, J.M., Hubel, C.A. (1999) Is oxidative stress the link in the two-stage model of pre-eclampsia? *The Lancet* **354** (9181): 788-789.

Roberts, J.M., Lain, K.Y. (2002) Recent insights into the pathogenesis of pre-eclampsia. *Placenta* **23**: 359-372.

Robillard, P., Dekker, G., Chaouat, G., Hulsey, T. (1998) Etiology of preeclampsia: maternal vascular predisposition and couple disease-mutual exclusion or complementarity? *Journal of Reproductive Immunology* **76** (1-2): 1-7.

Robillard, P., Dekker, G., Chaout, G., Hulsey, C. (2007) Etiology of preeclampsia: maternal vascular predisposition and couple disease-mutual exclusion or complementarity? *Journal of Reproductive Immunology* **76**: 1-7.

Robillard, P., Hulsey, T.C., Perianin, J., Janky, E., Miri, E.H., Papiernik, E. (1995) Association of pregnancy-induced hypertension with duration of sexual cohabitation before conception. *Obstetrical and Gynecological Survey* **50** (4): 256-257.

Rombauts, S., Florquin, K., Lescot, M., Marchal, K., Rouze, P., van de Peer, Y. (2003) Computational approaches to identify promoters and *cis*-regulatory elements in plant genomes. *Plant Physiology* **132**: 1162-1176.

Sacks, G.P.; Studena, K.; Sargent, I.L.; Redman, C.W.G. (1998) Normal pregnancy and preeclampsia both produce inflammatory changes in peripheral blood leukocytes akin to those of sepsis. *American Journal of Obstetrics and Gynecology* **179** (1): 80-86.

Saha, A., Wittmeyer, J., Cairns, B.R. (2006) Chromatin remodeling: the industrial revolution of DNA around histones. *Nature Reviews Molecular Cell Biology* **7:** 437-447.

Sammar, M., Stalk, M., Gebhardt, S., Golan, E., Meiri, H., Huppertz, B., Hillermann, R. (2006). RNA splicing and DNA polymorphism leading to two shorter sub-forms of placenta protein 13 in pre-eclampsia. *American Journal of Obstetrics and Gynecology* **195** (6) Supplement 1: S141-S141.

Sattar, N.; Greer, I.A. (2002) Pregnancy complications and maternal cardiovascular risk: opportunities for intervention and screening? *British Medical Journal* **235**: 157-160.

Schäffer, L. (2003) Oxygen-regulated expression of TGF-β3, a growth factor involved in trophoblast differentiation. *Placenta* **24** (10): 941-950.

Scherf, M., Klingenhoff, A., Werner, T. (2000) Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: A novel context analysis approach. *Journal of Molecular Biology* **297**(3): 599-606.

Semenza. G.L. (1998) Hypoxia-inducible factor 1: master regulator of O_2 homeostasis. *Current Opinion in Genetics and Development* **8** (5): 588-594.

Smets, E.M.L., Visser, A., Go, A.T.J.I., van Vugt, J.M.G., Oudejans, C.B.M. (2006) Novel biomarkers in pre-eclampsia. *Clinica Chimica Acta* **364** (1-2): 22-32.

Solovyev, V.V., Shahmuradov, I.A. (2003) PromH: promoters identification using orthologous genomic sequences. *Nucleic Acids Research* **31** (13): 3540-3545.

Spencer K, Cowans NJ, Chefetz I, Tal J, Meiri H (2007) First-trimester maternal serum PP-13, PAPP-A and second-trimester uterine artery Doppler pulsatility index as markers of pre-eclampsia. *Ultrasound in Obstetrics and Gynecology*. **29**: 128-134.

Spencer K, Cowans NJ, Chefetz I, Tal J, Kuhnreich I, Meiri H (2007) Second-trimester uterine artery Doppler pulsatility index and maternal serum PP13 as markers of pre-eclampsia. *Prenatal Diagnosis*. **27**: 258-263.

Strachan, T., Read, A.P. (eds) (2004) *Human Molecular Genetics, Third edition.* London and New York, Garland Publishing.

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., Patapoutian, A., Hampton, G.M., Schultz, P.G., Hogenesch, J.B. (2002) Large-scale analysis of the human and mouse transcriptomes. *Proceedings of the National Academy of Sciences* **99** (7): 4465-4470.

Takai, D.; Jones, P.A. (2003) The CpG Island Searcher: A New WWW Resource. *In Silico Biology* **3**: 235-240.

Teppa, R. (2000) Free leptin is increased in normal pregnancy and further increased in pre-eclampsia. *Metabolism* **49** (8): 1043-1048.

Teraoka, S.N., Telatar, M., Becker-Catania, S., Liang, T., Önengüt, S., Tolun, A., Chessa, L., Sanal, O., Bernatowska, E., Gatti, R.A., Concannon, P. (1999) Splicing defects in the Ataxia-Telangiectasia gene ATM: underlying mutations and consequences. *American Journal of Human Genetics* **64**: 1617-1631.

Than, N.G., Abdul Rahman, O., Magenheim, R., Nagy, B., Fule, T., Hargitai, B., Sammar, M., Hupuczi, P., Tarca, A.L., Szabo, G., Kovalszky, I., Meiri, H., Sziller, I., Rigo, J. Jr, Romero, R., Papp, Z. (2008) Placental protein 13 (galectin-13) has decreased placental expression but increased shedding and maternal serum concentrations in patients presenting with preterm pre-eclampsia and HELLP syndrome. *Virchows Archives* **453**(4):387-400.

Than, N.G.; Paidas, M.J.; Mizutani, S.; Sharma, S.; Padbury, J.; Barnea, E.R. (2007) Embryo-Placento-Maternal Interaction and Biomarkers: from diagnosis to therapy- a workshop report. *Placenta* **28** Supplement A Trophoblast Research (21): 107-110.

Than, N.G., Pick, E., Bellyei, S., Szigeti, A., Burger, O., Berente, Z., Janaky, T., Boronkai, A., Kliman, H., Meiri, H., Bohn, H., Than, G.N., Sumegi, B. (2004) Functional analyses of placental protein 13/galectin-13. *European Journal of Biochemistry* **271**(6): 1065-1078.

Thanaraj, T.A., Stamm, S. (2003) Prediction and statistical analysis of alternatively spliced exons. *Progress in Molecular and Subcellular Biology* 31: 1-5.

Thijs, G., Lescot, M., Marchal, K., Rombauts, S., De Moor, B., Rouzé, P., Moreau, Y. (2001) A higher order background model improves the detection of regulatory elements by Gibbs Sampling. *Bioinformatics* **17**(12): 1113-1122.

Thijs, G., Marchal, K., Lescot, M., Rombauts, S., De Moor, B., Rouzé, P., Moreau, Y. (2002) A Gibbs Sampling method to detect over-represented motifs in upstream regions of co-expressed genes. *Journal of Computational Biology* (special issue Recomb'2001) **9**(2): 447-464.

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22** (22): 4673-4680.

Tower, C. (2004) Pre-eclampsia Current Perspectives on Management, p41, Editors: Baker, P.N.; Kingdom, J.C.P. Parthenon Publishing, New York.

Vanio, M.; Kujansuu, E.; Iso-Mustajarvi, M.; Maenpaa, J. (2002) Low dose acetylsalicylic acid in prevention of pregnancy-induced hypertension and intrauterine growth retardation in women with bilateral uterine artery notches. *BJOG International Journal of Obstetrics and Gynecology* **109**: 161-167.

Verwoerd, G.R., Hall, D.R., Grové, D., Maritz, J.S., Odendaal, H.J. (2002) Primipaternity and duration of exposure to sperm antigens as risk factors for pre-eclampsia. *International Journal of Gynecology and Obstetrics* **78**: 121-126.

Visegrády, B., Than, N. G., Kilár, F., Sümegi, B., Than, G.N., Bohn, H. (2001) Homology modelling and molecular dynamics studies of human placental tissue protein 13 (galectin-13) *Protein Eng.* **14**(11): 875-880.

Wagschal, A., Feil, R. (2006) Genomic imprinting in the placenta. *Cytogenetic Genome Research* **113**: 90-98.

Wang, G.S., Cooper, T.A. (2007) Splicing in disease: disruption of the splicing code and the machinery. *Nature Reviews Genetics* **8** (10): 749-761.

Wang, Y., Su, B. (2004) Molecular evolution of microcephalin, a gene determining human brain size. *Human Molecular Genetics* **13** (11): 1131-1137.

Wareing, M. (2004) Endothelium In: Baker, P. N. K., Kingdom, J.C.P. *Pre-eclampsia Current Perspectives on Management*. New York, Parthenon Publishing**:** 109.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Prüß, M., Reuter, I., Schacherer, F. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Research* **28** (1): 316-319.

Wolffe, A.P. (1998) Packaging principle: how DNA methylation and histone acetylation control the transcriptional activity of chromatin. *Journal of*

Experimental Zoology Part A: Comparative Experimental Biology. **282** (1-2): 239-244.

Wu, H., Su, B. (2008) Adaptive evolution of *SCML1* in primates, a gene involved in male reproduction. *BMC Evolutionary Biology* **8** (192): 1-12.

Yamazaki, K., Kawai, A., Kawaguchi, M., Hibino, Y., Li, F., Sasahara, M., Tsukada, K., Hiraga, K. (2001) Simultaneous induction of Galectin-3 phosphorylated on Tyrosine Residue, $p21^{WAF1/Cip1/Sdi1}$, and the proliferating cell nuclear antigen at a distinctive period of repair of hepatocytes injured by CCI4. *Biochemical and Biophysical Research Communications* **280** (4): 1077-1084.

Yang, Q.S., Ying, K., Yuan, H.L., Chen, J.Z., Meng, X.F., Wang, Z., Xie, Y., Mao, Y. (2002) Cloning and expression of a novel human galectin cDNA, predominantly expressed in placenta. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1574**(3): 407-411.

Zamudio, S., Leslie, K.K., White, M., Hagerman, D.D., Moore, L.G. (1994). Low serum estradiol and high serum progesterone concentrations characterize hypertensive pregnancies at high altitude. *Journal of the Society for Gynecologic Investigation* **1** (3): 197-205.

Zeisler, H. Jirecek, S., Hohlagschwandtner, M., Knöfler, M., Tempfer, C., Livingston, J.C. (2002) Concentrations of estrogens in patients with pre-eclampsia. *Wiener Klinische Wochenschrift* **114**(12):458-461.

Zhou, Y., Damsky, C.H., Fisher, S.J. (1997) Preeclampsia is associated with failure of human cytotrophoblasts to mimic a vascular adhesion phenotype. *Journal of Clinical Investigation* **99** (9): 2152-2164.

Zhou, R., Zhu, Q., Wang, Y., Ren, Y., Zhang, L., Zhou, Y. (2006) Genomewide oligonucleotide microarray analysis on placentae of pre-eclamptic pregnancies. *Gynecologic and Obstetric Investigation* **62** (2): 108-114.

Databases and *in silico* tools:

- Basic Local Alignment Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi)
- Bioedit software package (Ibis Biosciences, Carlsbad, CA)
- ClustalW (http://www.ebi.ac.uk/Tools/clustalw/)
- CpGplot from the European Bioinformatics Institute 2006-2008, EMBL(http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html)
- Ensembl Genome Browser (http://www.ensembl.org/index.html)
- EpiInfo 6 statistical software package (http://www.cdc.gov/epiinfo/Epi6/ei6.htm)
- ESEfinder v 3.0, 2001-2006, Cold Spring Harbor Laboratory (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home)
- European Bioinformatics Institute, ©EBI, 2006-2008, European Molecular Biology Laboratory (http://www.ebi.ac.uk)
- Genomics Institute of the Novartis Research Foundation (GNF) (http://www.gnf.org/)
- Institute for Bioinformatics (SIB) (http://au.expasy.org/uniprot)
- MotifSampler (http://bayesweb.wadsworth.org/gibbs/gibbs.html)
- National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov/www.ncbi)
- UniprotKB/Swiss-Prot, European Bioinformatics Institute (EBI) and SwissProt Protein Knowledgebase (http://www.ebi.ac.uk/swissprot/)
- WeederH (http://159.149.109.9/modtools/)

Other sources:

MRC Research Unit for Maternal and Infant Health Care Strategies, PPIP Users and the Saving Babies Technical Task Team (2006) Saving Babies 2003-2005: Fifth perinatal care survey of South Africa. MRC Research Unit for Maternal and Infant Health Care Strategies, Pretoria, South Africa www.ppip.co.za.

National Committee on Confidential Enquiries into Maternal Deaths in the office of the Minister of Health (2004). Confidential Enquiry into Maternal Deaths in South Africa (NCCEMD) www.doh.gov.za/docs/reports/2004/savings.pdf

Statistics South Africa (2001). South African National Census of 2001: http://www.statssa.gov.za/census01/html/default.asp.

APPENDICES

Appendix 1: Patient consent form and informed consent for DNA analysis and storage

FACULTY OF HEALTH SCIENCES
STELLENBOSCH UNIVERSITY

INFORMATION AND INFORMED CONSENT DOCUMENT FOR DNA ANALYSIS AND STORAGE

TITLE OF THE RESEARCH PROJECT: ***Profiling placental protein 13 / galectin-13 in pre-eclampsia.***

REFERENCE NUMBER: ……..……..…….

PRINCIPAL INVESTIGATOR: Drs R Hillermann & GS Gebhardt

Address: Departments of Obstetrics and Gynaecology and Genetics, University of Stellenbosch, Tygerberg and Paarl Hospital

DECLARATION BY OR ON BEHALF OF PARTICIPANT:

I, THE UNDERSIGNED,

……..…….. *(name)*

[ID No: ……..……] of ……..…….

…….……...……

(address).

A. HEREBY CONFIRM AS FOLLOWS:

1. I was invited to participate in the above-mentioned research project of the Departments of Obstetrics and Gynaecology and Genetics, Faculty of Health Sciences, University of Stellenbosch.
2. The following aspects have been explained to me:

   Aim: Pre-eclampsia (high blood pressure in pregnancy) is a severe complication of pregnancy and it affects about 5% of all pregnancies. Currently there is no way to predict who will be affected. We are investigating a protein normally secreted by the placenta (the afterbirth) as it seems that this protein may predict the development of later problems.

2.1 **Procedures:** I will be requested to provide information about my medical history. Blood (10*ml*, about 2 teaspoonfuls) will be collected from me at delivery of my baby together with the routine blood samples taken at birth. In addition 6 small biopsies (pea-size) will be taken from the afterbirth (the placenta) after delivery and before the placenta is routinely destroyed.

2.3 **Genetic considerations**

- The DNA may be stored for several years until the technology for meaningful analysis becomes available;
- The clinically relevant results of the (possible) analyses carried out on this material in the current study can be made known to me at my request and in accordance with the relevant protocol, if and when it becomes available. In addition, I authorise(s) the investigator(s) to make the information available to ……*(doctor's name),* the doctor involved in my care, as well as to the following family members ……..…….. ..…… *(names);*
- The DNA will be maintained indefinitely, unless I request to have it and/*or the stored clinical data destroyed by contacting the investigator conducting the present study, dr GS Gebhardt at 938 9131 or the Chairperson of the Research Subcommittee C/Ethics Committee at 9389111 if the former cannot be located;
- The analyses in the current study are specific to the condition or disease mentioned above and cannot determine the entire genetic make-up of an individual;
- Genetic analyses may not be successful in revealing additional information regarding some families or some family members;
- Even under the best conditions, current technology of this type is not perfect and could lead to unreliable results.

2.4 **Confidentiality:** My identity will be kept confidential throughout. Information will not be associated with my name. The research staff

will use only a coded number, access will be limited to authorized scientists and any scientific publications, lectures or reports resulting from the study will not identify me by name.

2.5 **Voluntary participation: Participation** is voluntary and I may decline participation, or withdraw from the study at any time without any loss of benefits to which I am otherwise entitled. Future management at this or any other institution will not be compromised by refusal or withdrawal.

2.6 **Risks:** There are no more than minimal medical or psychological risks associated with this study:

- I may feel some pain associated with having blood withdrawn from a vein and may experience discomfort, bruising and/or slight bleeding at the site;
- Biopsies of the placenta is not painful as it is taken after delivery before routine destruction of the placenta
- As some insurance companies may mistakenly assume that my' participation is an indication of a higher risk of a genetic disease which could hurt my access to health or other insurance, no information about me or my family will be shared with such companies as this investigation cannot be regarded as formal genetic testing for the presence or absence of certain genes.

2.7 **Benefits:**

- Although there may not be any direct benefits to me by participating at this stage, family members and future generations may benefit if the researchers succeed in scientifically delineating certain disorders further. Thereby the rational approach to the clinical diagnosis and therapy of its manifestations may be facilitated. The identification and location of the genes involved in such disorders, could in the end lead to the development of methods for prevention and to forms of new treatment aimed at curing or alleviating these conditions;
- In the unlikely event that the research may lead to the development of commercial applications, I or my heirs will not

receive any compensation, but profits will be reinvested into supporting the cause of further research which may bring benefits to my family and to the community, such as health screening, medical treatment, educational promotions, etc;

2.8 **Permission for further studies:** Before my material is used in further projects in the future, the written approval of the Research Subcommittee C/Ethics Committee, Faculty of Health Sciences, will be obtained.

3. The information conveyed above was explained to me by ……..……..…….…….…….. *(name)* in English and I am fluent in this language

4. I was afforded adequate time to pose any questions and all questions were answered to my full satisfaction.

5. I was not pressurized to participate.

6. I will not be paid for participation, but reimbursement of travel costs will be considered (if applicable).

7. I will not incur any additional costs through participation.

8. I have received a copy of this document for my records.

9. The Research Subcommittee C/Ethics Committee, Faculty of Health Sciences, Stellenbosch University, has approved recruitment and participation of individuals in this study on the basis of:
   - Guidelines on Ethics for Medical Research of the SA Medical Research Council;
   - Declaration of Helsinki;
   - International Guidelines : Council for International Organisations of Medical Sciences (CIOMS);

- Applicable RSA legislation.

I. I HEREBY CONSENT VOLUNTARILY TO PARTICIPATE/*ALLOW THE POTENTIAL PARTICIPANT TO PARTICIPATE IN THIS STUDY:

Signed/*Confirmed at ……..……..…… *(place)*

On---------- *(date)* 20……

……..

Signature or right thumb print of witness/participant/*representative of participant

DECLARATION BY OR ON BEHALF OF INVESTIGATOR(S):

I, ……..…*(name)*

declare that

- I explained the information in this document to ……. *(name of the patient/*participant)* and/or his/her representative ……..…… *(name of the representative);*
- she/*he was encouraged and afforded adequate time to ask me any questions;
- this conversation was conducted in Afrikaans/*English/*Xhosa/*Other ……..…….. and no translator was used/*was translated into ……*(language)* by ……..……...……..… *(name).*

Signed at ……..…….……. on ……..……20 …...

(*place)* *(date)*

……..……

……..……

Signature of investigator/representative of investigator *Signature of witness*

……

……..……

…….

Signature *of translator* *Signature of witness*

IMPORTANT MESSAGE TO PARTICIPANT/*REPRESENTATIVE OF PARTICIPANT:

Dear participant/*representative of participant,

Thank you very much for your participation in this study. Should, at any time during the study,

- an emergency arise as a result of the research, or
- you require any further information with regard to the study, kindly contact Dr GS Gebhardt at 938-0131, 938-4707 (after hours) or 87 21 711 (all hours) or come to the labour ward on the second floor, Tygerberg Hospital.

Appendix 2: DNA extraction protocol

Extraction buffers and chemicals used:

Nuclear Lysis Buffer (1L, pH 8.2 set with 1M NaOH, store at 4°C)

1.211g	Tris-Cl	0.01M
23.4g	NaCl	0.4M
0.6g	EDTA	0.002M

Cell Lysis Buffer (1L, pH 7.4 set with concentrated HCl, store at 4°C)

8.3g	NH_4Cl	0.155M
1.1g	$KHCO_3$	0.01M
0.03g	EDTA	0.0001M

Phosphate Buffered Saline (PBS) (1L, store at 4°C)

2g	KCl	0.027M
8g	NaCl	0.137M
1.14g	Na_2HPO_4	0.008M
0.2g	KH_2PO_4	0.0015M

10% SDS (100mL, store at room temperature to prevent precipitation)

10g	SDS

Work in vapour hood

Proteinase K (store at -20°C)

10mg/mL dissolved in dH_2O, aliquot in 1.5mL Eppendorf tubes

NaCl (500mL, saturated solution)

175.32g	NaCl	6M

Genomic DNA extractions from blood:

Day 1:

Place ~10mL blood in a 50mL Falcon tube

Add 30mL Lysis Buffer

Place on ice for 15 min, shaking each 5 min interval

Centrifuge @ 1500rpm for 10 min

Poor off supernatant, keeping pellet

Add 10mL PBS and mix

Centrifuge @ 1500rpm for 10 min

Pour off supernatant, keeping pellet

Dissolve pellet in: 30mL Cell Lysis Buffer

30µL Proteinase K (10mg/mL)

300µL 10% SDS

Mix well and incubate overnight in 55°C water bath

Day 2:

Add 1mL NaCl and shake for 1 min

Centrifuge @ 3500rpm for 30 min

Transfer supernatant to new tube and shake for 15sec

Centrifuge @ 2500rpm for 15 min

Transfer supernatant to new tube without foam or pellet

Add volumes ice cold 100% ethanol to precipitate DNA

Scoop DNA and place in Eppendorf tube with 500µL ethanol

Centrifuge @ 14 000rpm for 10 min @ 4°C

Dissolve pellet in 200-800µL ddH_2O/TE buffer depending on pellet size

Appendix 3: Multiphor SSCP/HD protocol

Multiphor Solutions:

40% acrylamide PDA solution: 1L

396g acrylamide
4g Piperazine diacrylamide
bring to volume (1L) with dH_2O (keep in foil in fridge)

0,75M TRIS- Formate Buffer: 1L

90.8g TRIS
600ml dH_2O (dissolve)
pH = 9.0 with formic acid
bring to volume (1L) with dH_2O

41% Glycerol: 100mL

41mL glycerol
59mL dH_2O

TRIS- Borate Buffer: 1L

125.9g TRIS
17.3g Boric acid
700mL dH_2O (dissolve)
pH = 9.0
50µL Bromophenol Blue (4% solution)
bring to volume (1L) with dH_2O

SSCP loading Dye: 50mL

47.5mL 95% Formamide
0.16g 100mM NaOH
0.125g 0.25% Bromophenol Blue
0.125g 0.25% Xylene cyanol
bring to volume (50mL) with dH_2O

10% APS: 1.5mL **10% acetic acid (plate glue):**
0.15g APS 200µL Trimethoxysily
bring to volume (1.5mL) with dH_2O 50 mL 100% EtOH
Gel mix: X10 **Per 1 Gel:**
53mL 40% acrylamide-PDA 15ml Gel Mix
85mL Tris-Formate Buffer 150µL 10% APS
30mL 41% Glycerol 15µL TEMED keep in foil in fridge

Silver Staining solutions:

Solution I: **1L**

0.1% silver nitrate:1g silver nitrate
1L dH_2O
Max. 2 days in the dark

15% Formaldehyde: 10mL (must be fresh each time!)

3.75mL 40% Formaldehyde stock
6.25mL dH_2O

Solution II: 1L

15g NaOH
1L dH_2O
10mL 15% Formaldehyde (FRESH!)

Multiphor Protocol:

Gel Plates preparation: do everything in quick succession

- Wash plates 4X with ethanol
- Every 4 runs clean plate with blocks with gelslick (80µL) especially over blocks
- Put 80µL plateglue and 8µL 10% acetic acid on plate without blocks
- Wipe well until resistance is felt and wash a few times with ethanol
- Clean spacers and place them on plates
- Camp well on both short sides and long area close to blocks and then place upside down on bench
- Prepare gel mix and pour quickly using a syringe
- Allow to set at room temp for 30min (can stand for 3 days -don't wrap! It shrinks)

- Separate plates (gel should be on one without blocks)
- Put water on surface of multiphor
- Don't get water on plates when placing on multiphor apparatus
- Slide plate facing upwards over water (no bubbles)
- Use 2 buffer strips/gel-side/2h run soaked in TRIS-borate buffer
- Place strips on both sides of gel close to the wells
- Wash electrodes of multiphor gently each time before use

PCR Products:

- Take 3-5µL PCR product and 3-5µL SSCP loading dye (all on ice)
- Denature @ 95℃ for 5min in PCR thermocycler

- Place on ice immediately for 3 min
- Load 3µL (skipping ends of gel) onto gel and run @ 9℃ first (if resolution not good, run at 11℃)
- 2h at 355V

Visualization of bands:

- remove gel from apparatus
- rinse X2 in dH_2O (remove buffer because it interferes with the stain)
- incubate for 10min @room temp in solution I (silver nitrate intercalates with DNA)
- rinse X2 with dH_2O
- incubate for 10min @room temp in solution II
- rinse X2 with dH_2O
- blot dry with paper towel
- cover with filter paper
- leave overnight, then put paper off glass and label paper with dried gel

Appendix 4: DNA purification protocol

Bioline SureClean ® (Bioline, UK)

Column-free protocol:

- Add an equal volume of Bioline SureClean to the amplified nucleic acid sample and incubate at room temperature for 10 min
- Centrifuge at 13 000rpm for 10 min and remove supernatant
- Add 100µL of 70% ethanol and vortex for 30sec
- Centrifuge at 13 000rpm for 10 min, remove supernatant and air-dry pellet
- Resuspend in Nuclease-free water (starting volume)

Appendix 5: *LGALS13* gene annotation

```
Summary of Exons in sequence:

LGALS13-Exon1 (5001-5060 -> 60bp)
LGALS13-Exon2 (7079-7155 -> 77bp)
LGALS13-Exon3 (7655-7865 -> 211bp)
LGALS13-Exon4 (9700-9950 -> 251bp)
Total size for LGALS13 = 599 bp

OTTHUMG00000071603-Exon1 (5006-5060 -> 55bp)
OTTHUMG00000071603-Exon2 (7079-7155 -> 77bp)
OTTHUMG00000071603-Exon3 (7655-7865 -> 211bp)
OTTHUMG00000071603-Exon4 (9700-9953 -> 254bp)

Summary of SNPs in sequence:

1 - (269)(c/t)(dbSNP:rs6508879)
2 - (318)(c/a)(dbSNP:rs8106020)
3 - (434)(t/a)(dbSNP:rs8107128)
4 - (2355)(a/g)(dbSNP:rs10401220)
5 - (2448)(g/a)(dbSNP:rs28839483)
6 - (3065)(a/g)(dbSNP:rs12610294)
7 - (3197)(g/c)(dbSNP:rs11881603)
8 - (3538)(a/g)(dbSNP:rs10409099)
9 - (4294)(c/t)(dbSNP:rs2233702)
10 - (4444)(c/t)(dbSNP:rs2233703)
11 - (4585)(g/c)(dbSNP:rs3764842)
12 - (4661)(-/g)(dbSNP:rs35552633)
13 - (4694)(t/g)(dbSNP:rs2233704)
14 - (4740)(a/g)(dbSNP:rs2233705)
15 - (4948)(c/a)(dbSNP:rs3764843)
16 - (5164)(c/t)(dbSNP:rs3764844)
17 - (5758)(t/c)(dbSNP:rs1986156)
18 - (6080)(c/t)(dbSNP:rs2158964)
19 - (6365)(a/g)(dbSNP:rs1014206)
20 - (6686)(t/c)(dbSNP:rs28445525)
21 - (7308)(t/c)(dbSNP:rs10426654)
22 - (7633)(a/g)(dbSNP:rs2233706)
23 - (7926)(t/c)(dbSNP:rs2233707)
24 - (7937)(t/a)(dbSNP:rs2233708)
25 - (7964)(c/t)(dbSNP:rs2233709)
26 - (8555)(c/a)(dbSNP:rs12978134)
27 - (9458)(g/c)(dbSNP:rs9917050)
28 - (9527)(c/t)(dbSNP:rs17795657)
29 - (9589)(a/g)(dbSNP:rs2233710)
30 - (9740)(g/a)(dbSNP:rs2233711)
31 - (9829)(c/g)(dbSNP:rs1801654)
32 - (11227)(c/t)(dbSNP:rs9916956)
33 - (11495)(a/g)(dbSNP:rs7246392)
34 - (11500)(-/aaaaa)(dbSNP:rs10653764)
35 - (11640)(t/c)(dbSNP:rs11881743)
36 - (11967)(c/t)(dbSNP:rs12980535)
37 - (12285)(a/g)(dbSNP:rs16973535)
38 - (12289)(a/g)(dbSNP:rs12981089)
39 - (13940)(-/c)(dbSNP:rs35368312)
40 - (13989)(g/a)(dbSNP:rs10411383)
41 - (14023)(t/a)(dbSNP:rs10412744)
42 - (14371)(g/a)(dbSNP:rs10412234)
43 - (14381)(a/g)(dbSNP:rs10411717)
```

```
Total of 43 SNPs in annotation

Summary of STSs in sequence:

1 - (9715-9904)(STS:STS-R31571)

Total of 1 STSs in annotation


RNA splicing branch sites are searched for in the 60 bases before the
splice acceptor site (3' of intron), only 4 patterns are searched for:
    "TCCTRAY(found in U12 introns)", "YNYYRAY", "YRYRAY", and "CTRAY"

Note! IUB codes used for DNA in Branch site descriptors:
   A = Adenosine, C = Cytosine, G = Guanosine, T - Thymidine,
   R = A or G (puRine), Y = C or T (pYrimidine), K = G or T (Keto),
   M = A or C (aMino), S = G or C (Strong), W = A or T (Weak),
   N = A or C or G or T (aNy base).



Annotated sequence file:

LOCUS       19 14950 bp DNA HTG 19-SEP-2006
DEFINITION  Homo sapiens chromosome 19 NCBI36 partial sequence
            44780004..44794953 reannotated via EnsEMBL
ACCESSION   chromosome:NCBI36:19:44780004:44794953:1
VERSION     chromosome:NCBI36:19:44780004:44794953:1
KEYWORDS    .
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates;
Haplorrhini;
            Catarrhini; Hominidae; Homo.
COMMENT     This sequence was annotated by the Ensembl system. Please
visit the
            Ensembl web site, http://www.ensembl.org/ for more
information.
COMMENT     All feature locations are relative to the first (5') base
of the
            sequence in this file.  The sequence presented is always
the
            forward strand of the assembly. Features that lie outside
of the
            sequence contained in this file have clonal location
coordinates in
            the format: .:..
COMMENT     The /gene indicates a unique id for a gene,
            /note="transcript_id=..." a unique id for a transcript,
/protein_id
            a unique id for a peptide and note="exon_id=..." a unique
id for an
            exon. These ids are maintained wherever possible between
versions.
COMMENT     All the exons and transcripts in Ensembl are confirmed by
            similarity to either protein or cDNA sequences.
FEATURES             Location/Qualifiers
     source          1..14950
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
     gene            5001..9950
                     /gene=ENSG00000105198
     mRNA            join(5001..5060,7079..7155,7655..7865,9700..9950)
                     /gene="ENSG00000105198"
                     /note="transcript_id=ENST00000221797"
     CDS             join(5046..5060,7079..7155,7655..7865,9700..9816)


/translation="MSSLPVPYKLPVSLSVGSCVIIKGTPIHSFINDPQLQVDFYTDM

DEDSDIAFRFRVHFGNHVVMNRREFGIWMLEETTDYVPFEDGKQFELCIYVHYNEYEI
                     KVNGIRIYGFVHRIPPSFVKMVQVSRDISLTSVCVCN"
     exon            7655..7865
                     /note="exon_id=ENSE00000706133"
     exon            5001..5060
                     /note="exon_id=ENSE00001058171"
     exon            9700..9950
                     /note="exon_id=ENSE00000877684"
     exon            7079..7155
                     /note="exon_id=ENSE00000877685"
     gene            5006..9953
                     /gene=OTTHUMG00000071603
     mRNA            join(5006..5060,7079..7155,7655..7865,9700..9953)
                     /gene="OTTHUMG00000071603"
                     /note="transcript_id=OTTHUMT00000152233"
     CDS             join(5046..5060,7079..7155,7655..7865,9700..9816)
                     /gene="OTTHUMG00000071603"
                     /protein_id="OTTHUMP00000077843"
                     /note="transcript_id=OTTHUMT00000152233"
                     /db_xref="Vega_transcript:LGALS13-001"
                     /db_xref="Vega_translation:OTTHUMP00000077843"

/translation="MSSLPVPYKLPVSLSVGSCVIIKGTPIHSFINDPQLQVDFYTDM

DEDSDIAFRFRVHFGNHVVMNRREFGIWMLEETTDYVPFEDGKQFELCIYVHYNEYEI
                     KVNGIRIYGFVHRIPPSFVKMVQVSRDISLTSVCVCN"



BASE COUNT  4404 a 3198 c 3070 g 4278 t
ORIGIN
     1 cacacttccc acatactaaa aactctaacc ctgtaaggaa ctactgaaag
caatacaatt
    61 acaggaggca gctcagattg aagtccagca tcttatcatc caattactct
cagatgtcca
   121 tgaggagact acaaactgtc aggtgcactt tctctcaaaa tatgcaccaa
agggatacat
   181 atttgcccta tacaaatcaa tcacacaaga ctgaaacaga caaaggaaaa
tgcagttaac
   241 acaactattt tagaggggca aaagataa^ca aatggcacct tttgcctcat
agcactctga
       (269)(c/t)(dbSNP:rs6508879)
   301 gaagcaagag gggtcac^cag cagctctact agagccactt atgaaaattc
agcctgggtt
       (318)(c/a)(dbSNP:rs8106020)
   361 ctcactcttg tcgcatctgc tcaaaccagc atagtctgtt ccggaaatac
tacctcaatg
   421 tgtgccccca gtg^tttccat aagtacttta aaaacataca tttcattaag
ttggactaag
```

(434)(t/a)(dbSNP:rs8107128)
481 aggtcttctt tgaatggata accaaggcat ccacacaacg aaagaaaccg
tgttagctct
541 ttgtacataa atttaaaatt caaacaaaca aacaaacaag tacagaaaag
aaaatagacc
601 acagtcacgg tccatatcac tgaggaaacc aacaagatct ctttttcat
tttctctggg
661 aacgtttccc tgcagcaatt ggcttcatcc tccattctta gcttcttgga
actcctacca
721 agatggcaga atgacagtcc tttttccact acaagagttt tgattatcta
cttcatagtg
781 catatagaga accatttagg agaagataaa caacatattt cctaaaagtc
aacgaaggat
841 gattgaagag agtctcatca aatagccttt tttctgctgg gccaatctca
ggcatcattt
901 tgatcatcgt catcctctct ccccaggttc attgccacag cttgttcaac
tccaacctta
961 ctcatgggta tgcctttctc tgtactgctt ctttaaccta acatctggat
accttgaggt
1021 atcaatgcac agtgaattat gctcttcatg cctttcttcc ttgtttcttt
ccagtataaa
1081 atgaaatttc ctcccctctc attggaaact ctcaggttct gattaggtac
actctcctcc
1141 ttacaccatc cctgtcacag tcctggtcaa ttctgtattt tgatagtgga
ttataggctc
1201 catcaccagc atggaattca tttgttgggg atttaggcaa ttctttctct
gagaggatga
1261 agactatcct ggaatctatg acttcctgag cagaacatga gtttgatttc
atatctgctt
1321 gcctctggat gacacccatt ttatttgagt ggcctacaga gtcctacatg
ctggccttgc
1381 aacacacaca gtgactccca tttcttggct gacattcacc agctcatttc
tgttcacatg
1441 ggccatttgt tacactcatt tttatctcag gtaatacgtt ttgctcatgg
tggccaggat
1501 gacatctcag acacaaagag atgacataga attcaggagc caatgtttaa
caaagaattt
1561 ttgaattgcc tctggggacc tactcaacat cattttcata ctcctcaacc
aaaattccct
1621 atcttctccc agcaaatacg gcatgctcaa aacacagaaa gtcgtacttg
gactctaaag
1681 ctttcattct caaaggtatt taggcttttt agtttttgtt gaagcctccc
aaggaataaa
1741 tctatttggt gataatattg tccttgtccc tcaaagctgt ttcacttgtg
gtgctgtccg
1801 tggtgctgag cattggtgca ttaggccctg tatttgaaag tgaggagcca
atacgacaga
1861 gtaagagagg aattttgttt tgttttcttt tgtatttgct agaccataga
caaaacctcc
1921 atgatgacta aaacagaaat tcaacttgga ttttcaaacc acaatttctt
cagtgccctc
1981 tctcctcact tctatttaac ataatattgg aaggcctgac cagagcaatc
acgcaagaga
2041 aaagaagaaa gggtatctaa atacaaagag aggtcaaacc atccctgttt
gcagatgtca
2101 tgattctata tctagaaaac tccatagtct tcacccgaaa gatccttaag
ctggtaaacc
2161 actttagcaa agtttcagga tacaaaatca atatgcaaaa atcactagaa
ctcctatatg
2221 ccaacaacag cgacagggac aaatcaagaa tgcaattcca ttcaaatttg
ccataaaaag

2281 aataaaacac ctaaaaatac agctaatcag ggaggtaaaa gatcgttaca
atgagaatta
2341 caaaacacta ctca^aggaaa ccagagatga cacaaagaaa tagaaaaaca
ttccatgctc
(2355)(a/g)(dbSNP:rs10401220)
2401 atagataaga aaaaccaatt atcattaaga cagccatact gcccaaa^gaa
atttagatgc
(2448)(g/a)(dbSNP:rs28839483)
2461 aaggctattt gtctcaaact atgaatgaca ttcttcacag aattagaaga
aatattattt
2521 taaaatttaa atggaaccaa aaaaatgcca aataaccggg gcaattctaa
gtgaaaagaa
2581 caaaactgga ggcatcaaat tacccaactt caaaatatac tacaaggcta
cagtaaccaa
2641 aacagcatgg tatcggtatg aaaagagaca cacagaccaa tggaacagaa
tggaaagccc
2701 taaataaggc cacaaaccta caaccatctg gtcttcaaca aagctgacca
aacaagcaat
2761 ggaaaaagga ctccctattc aagaaatggt gctgggagaa ctggctagcc
atatgcagaa
2821 gattgaaact ggacccctc tttacaccat atacaaaaga tagattaaag
acttaaaaga
2881 ttaaacccaa aactataaac atcctggaag acaacctaga caataccatg
tgggacatag
2941 gcatgagaaa aatttcatga tgaagatgtc aaaggcaatt aggagaaaag
caaaaattga
3001 caaacagaat ctaattaaac taaagagctt ctgtacagtg aaggaaacta
tcaagagagt
3061 aaac^agacag cctaaggaat gggagaaaat gttctcaaac tatgtatctg
aagaaggtct
(3065)(a/g)(dbSNP:rs12610294)
3121 aatatccgag catttataag aaatttcaac aacttcacga gaaagaaaca
aacaaaccca
3181 taaaaaagtg ggtaaa^gagc ttgacacagt ggctcatgcc tgtaatccca
gcgctttggg
(3197)(g/c)(dbSNP:rs11881603)
3241 caaccaaggc tggtggatca cttgagccca ggaatttgag aacaggctgg
gtaacatggc
3301 aaaaacccat ctctactaaa aatagaaaaa aaaaatagcc tagcatggag
acatgcacct
3361 gtagtcccag ctactcagga gactgagctt agaaaaatca cctgagcccc
ggagccagag
3421 gttacagtga gccaagatgg tgccacatca ctccagcctg aatgtcagag
tgagacctgt
3481 tttaaaacaa caacaaaaaa gttgagcaaa ggacacgaac agacactttt
caaaagt^aag
(3538)(a/g)(dbSNP:rs10409099)
3541 gatacaggtg gccaaaaaac atataaaaaa gctcaatatc actgaacatt
agagaaatgc
3601 aaatcaaaac cacagtgaga taccatctca caccagtcag aatggctgtt
attaaaaagt
3661 caaaaaagga cagatgcttc aaggttgtgg agaaaaggga gaactcacac
attgttggtg
3721 acagtgtaaa ttagttcacc cattgtggaa agcagtgtgg tgattcctca
aagagctaac
3781 aatagatcta ccattcaacc cagcaatccc atcactgggt gtatacccag
aagaaaagaa
3841 atcattctgc cataaagaca tatgcacatg aatgttcatt gcaacactat
gcacaatagc
3901 gaagacacag aatgaaccca aatgccaatc aatgacagat cggttacaga
aaatgtggta

```
  3961 catgtacaca atgaaatact atgcagtcct aaaaaagaat gatattatgt
cttttgcagg
  4021 aacatggatg gaactggagg ccattattct tagcaaacta atgcaggaac
agaaaaccaa
  4081 atactgcatg ttctcactta tcagcgtgaa ctaaatgagg aggctcatgg
acatcaagag
  4141 ggaaacaaca cacactgggg cctccttgag ggtggaggat gaaaggtgga
agaggagaaa
  4201 aaaaataact tttgagtact acacttagta accaggtgaa gaaatcatat
acaccacaaa
  4261 tcccatgaca cgactttatc tgcatatgta ccc^cgaacct aaaataaaat
gaaagaaaaa
       (4294)(c/t)(dbSNP:rs2233702)
  4321 tcaaataaaa acaacaatct caccttcctc ccagaatctg caggagaaac
agtgtctctg
  4381 aaacttttat ttacatggac accaagagaa aggagagggt gtagctagag
cccaggtgtg
  4441 tga^cagtcag gtgagagtgg cccttacctg gcagagccaa gcagcctagt
gtggaatagg
       (4444)(c/t)(dbSNP:rs2233703)
  4501 ttcatgcaag aaatttctct gcttctctcc acagatcctc ctgaggatca
aggtcagtgc
  4561 tctcccagaa cgttcaaaag caga^gagaat cattaccttc tagggctaat
gtatttcccc
       (4585)(g/c)(dbSNP:rs3764842)
  4621 aaaatctcca actttttttct tttcagccgt tgtacaaatt ^gggaaaatgt
acactttctg
       (4661)(-/g)(dbSNP:rs35552633)
  4681 ggtatattct gca^tagatga gaaaaggctt gagaggcata aaggctgggc
tttacaacc^a
       (4694)(t/g)(dbSNP:rs2233704), (4740)(a/g)(dbSNP:rs2233705)
  4741 ctaccgctca ctccatcatc tttatctatg catttccctg gtaacccaat
ccacagcctg
  4801 gtaactcccc cactgtcact tctgctgcat tacagataca gacctgcaaa
tagctatggt
  4861 tgtgactgag tttcttccta acaactaaac ctgtctaaag ctgcaaggaa
atcttgctgg
  4921 gaggagcttg gaatctggaa tgaagcc^cga gggcaaggct gaagtgggtc
atttaaatgc
       (4948)(c/a)(dbSNP:rs3764843)
OTTHUMG00000071603-Exon1 (5006-5060 -> 55bp)
TOP
  4981 ^tgcaactcag agattcactc AGAAGACTGG ACTCAATTCT GAAGGTCGCC
AAGAAGGAGA
       (4981)(branch site - YRYRAY)
  5041 GAACA^ATGTC TTCTTTACCC gtgagttgaa aaggcacagc cttcaaaaat
ttcgtgtcac
       start
  5101 acaaaccaag aaagaaatgg gagattttat gagatgaaaa tatgagcatt
tttgctgtga
  5161 atg^ctttact tagagctatt gaggtgtgga atagaaaccc tgaggctatg
gtatctgaga
       (5164)(c/t)(dbSNP:rs3764844)
  5221 tgcttgtggg gattgcggtg tggaccctgg accagtgtaa ccttcggtga
gtgtgaggtg
  5281 gtgtctgatg agcataaact ctgcggagat aatgtgactt taagtgggag
gtggattacc
  5341 caccaggacg agcaggtgtg tgaatggaga gtaaaggggt gagtcggctt
tgtttcctgt
  5401 ctatgatgtg agtatgtgtg tgagcacagc cagaccagtg catgagcatc
gttatgtgca
```

```
  5461 cagtgagcag gtgtgtgtga cgcagtgagt ggtgtggctc tgtgtgactg
tgagtgtgtg
  5521 tgctgggatg tgactgtatg ttgaagtggg atcatgcatg tggctttctt
cacatgtgaa
  5581 tgtttctccc cacacagcag cacctctgat gtgaatccca tgtggcttaa
ctggtgggga
  5641 tcttggggat tgtgagttac caaggactta tttcagatca cttgtgagag
cagaatgtat
  5701 gatcagtaga tgtctcttcc cagtaagagt ggggaagagg aatgcagact
atggggg^tga
       (5758)(t/c)(dbSNP:rs1986156)
  5761 tggttcctca ggtggggtct gcacagagtg accactgtct cttgcctccc
aggctcagct
  5821 tgcctgtccc tagcccatcc agttcttccc actgtgtgac aggagggtgt
gcagatactg
  5881 gtttgttagg gaaggacatt agaggccgtg atttcagaga ctgaatcaca
caaaatatcc
  5941 actgtaagct ctgtctgaac tgagactaaa cttgttctac acccatgaaa
gaacaagtta
  6001 agtctcagtt cagaaggagc ttacagtgga tattaatcaa ccaagcagct
gggttcactg
  6061 gcttatggct ataatccca^c caattttgga ggcaaagatg agaggataac
gaggccagaa
       (6080)(c/t)(dbSNP:rs2158964)
  6121 gtttgagacc agtctgggta tcagagtgag acccttttctt taaaaaactt
taaatatttg
  6181 actggcagag aggcaggtac ctttcttccc agctattcaa gagggagcag
atgtgtgaga
  6241 ttcactgcag cccaggagtc tgaggtttca gtgaaccatg attgcatccc
tgcatgtcag
  6301 cctgggtgac agagcaagaa cctatctcaa aagtacagaa aaatcatctc
atctacttgt
  6361 agtc^atcata gaaatcaatc attccctcca gttatgtccc tgacccacaa
gcttcatttg
       (6365)(a/g)(dbSNP:rs1014206)
  6421 tgcaagtact ggggctgtgc tctcagtagt gtgtgcccct tcttggaagg
atgtccatgg
  6481 cccttgatga tagtgatgca tgtgcatccc acacacaggg gtttcctgtt
ctttcatcat
  6541 ttccctctcc ctttcaacaa gtgttgtttc tcactggagt gaatgtttaa
attcttacat
  6601 ttacaagtct tgttttaatt ttcagagttg aaaatgaagg caggtcagtc
tttcaagagg
  6661 actctcctgt cctgttaggt caccc^taaga ggccctacgc atgacaggga
ttagggtaac
       (6686)(t/c)(dbSNP:rs28445525)
  6721 acttttacat gtattaactc cctattacgg agaacatcct acaaagtaga
aattcctgat
  6781 aatcctcgtt gtcctaggga aaatgaggga gagttcaagt agcaaatcta
agactgcact
  6841 gctagtaaaa ggcagaatga ggactggaat ccaggcatcc tggctcctga
acccttgctc
  6901 taacttatgg gtccgccata tcttcaggaa tatggggccc tgaatgcggt
agggttaaag
  6961 aggagagtcc acagagtctg ccctttcatc tccaacctcc tgcaccatga
gaatatgtta
OTTHUMG00000071603-Exon2 (7079-7155 -> 77bp)
TOP
  7021 caggagggga gactgcacct gaccctgcac ctctcactta c^tctcaatac
tctggcagGT
       (7062)(branch site - YNYYRAY)
```

```
  7081 GCCATACAAA CTGCCTGTGT CTTTGTCTGT TGGTTCCTGC GTGATAATCA
AAGGGACACC
  7141 AATCCACTCT TTTATgtgag tactccatgg tccaatggag gggttggaga
agaagggaga
  7201 atatttgcga agatttgacc ttacatgtgg gtgatgtgga aatgtctagt
tggcagaatg
  7261 gaggccccat gcaggtgcag gtcttggaga ccctccagca ccaggca^tga
gacccacagc
       (7308)(t/c)(dbSNP:rs10426654)
  7321 aactgcatgt gggccagcca tggggcctac aggaggagaa cactcagtgg
gttggggctg
  7381 tggctcttta tcagggggca tagaattttc gggaatgaac aagtcatagg
cccaggtcag
  7441 tgactgtggc tcagttttca tctggggatg aggagcacag aatctccctg
cctgggggca
  7501 tgaggagctg aagcatcccc acagggacct ggccatcagt attatctggg
agacttttc
  7561 cctaggtaaa tgggggaagg gatttgtgtg tgtgtcgagt gtgtgtctgc
gcaagggagg
OTTHUMG00000071603-Exon3 (7655-7865 -> 211bp)
TOP
  7621 gacctgccca ac^attctgcg tgcttcaccc tcagCAATGA CCCACAGCTG
CAGGTGGATT
       (7633)(a/g)(dbSNP:rs2233706)
  7681 TCTACACTGA CATGGATGAG GATTCAGATA TTGCCTTCCG TTTCCGAGTG
CACTTTGGCA
  7741 ATCATGTGGT CATGAACAGG CGTGAGTTTG GGATATGGAT GTTGGAGGAG
ACAACAGACT
  7801 ACGTGCCCTT TGAGGATGGC AAACAATTTG AGCTGTGCAT CTACGTACAT
TACAATGAGT
  7861 ATGAGgtgag catcccagga gctcccagca cccaggctct gtgggctccc
aaaacaggag
  7921 gcagc^tctca ttgatc^tggc ctcagtagtc cgtcagggtc cat^ctcccat
aactactcct
       (7926)(t/c)(dbSNP:rs2233707), (7937)(t/a)(dbSNP:rs2233708),
(7964)(c/t)(dbSNP:rs2233709)
  7981 gcccctggtt ttcatcacag agcacccccc gcttgcactg ccatcctcag
ctctttccca
  8041 aatctgacca atgtcaaggt cagctcacct gacatttccc ccaaagtaga
taatctcctc
  8101 tgtcttttca catactgctc atttctactt atgccattat ttaaattttc
atttagctga
  8161 gtgaatacaa tatacttcag gaaaaaaaag tttaaaaaaa aaaggttgtt
ttaatcagtc
  8221 acaaattaag tctttgtctc attttggggc ccgaattcta ccataggaaa
aacattacca
  8281 gagctctgct ttcattcaaa gaggttttgt ttgtttgttt gtttttgttt
gtttgttttc
  8341 ttgagataga gtctcatttt tcatccaggt tggagtgcag tggcatgata
tcagctcact
  8401 gcaacctcca cctcctggat tcaaacaatt ctctcgcctc agcctcccaa
gtagctggga
  8461 ctacaggtgt gtgccagcac acctggctaa tttttgtatt tttaggggag
acgggggttc
  8521 accactttgg ccaggctggt cttgaactcc ccac^ctcaag ttatctgccc
aatttggcct
       (8555)(c/a)(dbSNP:rs12978134)
  8581 ctcaatataa tgggattaca ggcttaagcc accatgccca gcccttcaga
ggaaatttt
  8641 gcttatggaa gcatatagat atgtctatga tttcatttgt ttttaagcaa
atcctggtat
```

```
  8701 aagttgctct acacattacg cttcaagaac taagttgatt tgttgccaaa
gaatatatat
  8761 tgaagacatt atcaatctgt cgtgctgctt cagtttttaa agattgcata
atattgtgtt
  8821 aagaagttta cccaaagttg tggaacttgg ccctttgggg ggcactaaat
ttccttccag
  8881 tattttgctc tcacaaatga tgtcgtatta agttttctaa tcaatccttt
tgtcattacc
  8941 aacctcccaa tacttgtact agaagataat tacaaaaaaa tggaaaaact
ggctcaaaaa
  9001 tgtgtgcatt ttaaattcga aaaattaata ctaataaggt tactttctct
atagcctcaa
  9061 tataaagaat acatccatac tttctggctt tctctaacga gtaggccaaa
acattcacat
  9121 cttattcaga ttcaagcgag ggatcttttc ctatgcttag aagtttttga
aggtaatatc
  9181 ctgtttcaca agaacagttc tcctttcatt ctataagaac cctgggtttc
cttctttgcc
  9241 actaacacct cagatgtagg ttactccaaa gagaaactgt gccatcagta
gtgaaaaaca
  9301 atcacagttc atggaactga aaagtatgca ttcaacgaac atggtctagc
actaaccctg
  9361 tggcaggtcc tgtgcgagat gcagggtctc aagttcctga gacacagtcc
ctggcgatgg
  9421 ggatcttcca ggttagaagg gaggctgagt aagcaaa^gac tttgtgacac
agagtataga
       (9458)(g/c)(dbSNP:rs9917050)
  9481 acttcgctag aggaatgagt ggaaacatta gaaataaagt cgggta^ctgt
ctcaaatagg
       (9527)(c/t)(dbSNP:rs17795657)
  9541 cacaaaacgt catctgtaaa cataagtgta tctaatacgt taacttgt^at
aactaggaat
       (9589)(a/g)(dbSNP:rs2233710)
  9601 tttcttgggg aatgttattt gtaccaggac agagtggaga ggaggccgaa
aacttgtttg
OTTHUMG00000071603-Exon4 (9700-9953 -> 254bp)
TOP
  9661 gtggcatgct ttct^ttctga tgcatttttc ctcttgtagA TAAAGGTCAA
TGGC^ATACGC
       (9675)(branch site - YNYYRAY), (9715-9904)(STS:STS-R31571)
  9721 ATTTACGGCT TTGTCCATC^G AATCCCGCCA TCATTTGTGA AGATGGTGCA
AGTGTCGAGA
       (9740)(g/a)(dbSNP:rs2233711)
  9781 GATATCTCCC TGACCTCAGT GTGTGTCTGC AAT^TGAGGGA GATGATCA^CA
CTCCTCATTG
       stop, (9829)(c/g)(dbSNP:rs1801654)
  9841 TTGAGGAATC CCTCTTTCTA CCTGACCATG GGATTCCCAG AACCTGCTAA
CAGAATAATC
  9901 CCTGCTCACA TTTTCCCCTA CACTTTGTCA TTAAAACAGC ACGAAAACTC
ACAtgatttg
  9961 gttcttgctt tcagagggga aaagaggaag ttgtcatccc caaggggggc
cagggcattc
 10021 tatgggaggc atcaggaaat caaaggggat aaaccttcct gtgacaaagg
gagtgagtga
 10081 caaggtccgt ggaatgtctg agaagacatt agaaacagca tccttctata
gcacgtagtt
 10141 gacctcagac agtctgagct aaaatactta cccaatctca aatgactcaa
ggcacttaaa
 10201 tgttaggtag ctgtttacag ccacgcaagt gtacatggtg gctttgggga
aaatctgtaa
 10261 ctcgaggaaa attatgtgga atttcttaat ttttttccat tgttctcagg
```

```
atgcttactg
 10321 tgtatgaaca ctgtgcccgg ctttgtgcgg gctgctgtga aggaactgaa
gaaaaagatg
 10381 acaagtttgc atgtgaagtt ttcaatccaa gtaatagttt agagaatcac
tggagcccag
 10441 tgatttccct taaagaatcc caattattct tattttcaaa tatttgaaaa
ttgcttcatt
 10501 ttatgtgggg agaatccagg ggtttctctt tggaaatgaa tggtcatgat
ttgagatcag
 10561 agaacatata gattattatt ggaatagata caagtcacgt acattgacac
agtgagcaga
 10621 cccttatttg ttgcttttcc agtgatgaat tgcagcacag caaacctcac
aacacacatc
 10681 ggctaaaagc aatcatttat tttgatctct ccccgcttca caggtttact
gaacttaagt
 10741 gagtggtaga ttatgtgacg ggagtcgcgt tttctgagag atcttttagg
ctgggatttc
 10801 agaggatgct ttactcatgt ccaactcaac agcaggactg ggtgaacaca
tagaaactgg
 10861 gaagctgtct ccttcaatac ctcaccttcc catgcatgac attaatttcc
taaaatcagg
 10921 agagtctcat ggttctggga cttaataaat ggtgattggc tttcaccaaa
gaaagccttt
 10981 caaaaggctc atggagaagg tgcacagcat cacaggactt acctttgaat
ggcagacagt
 11041 gttacttcta aggttttaca ttgatcaagg tgagtcatac agacggccct
aatcaaaagg
 11101 agaggactcc acaagctgtg gttattgtga ctgtagtcat gaggggccat
ccttggcaac
 11161 tagttccact cagaactagt tcccacatac cgaaaacact aaccatgtaa
ggaatctcag
 11221 gctggg^cgcg gtggctcacg cctgtaatcc cagcactttg ggaggccgag
gcgggcggat
       (11227)(c/t)(dbSNP:rs9916956)
 11281 cacgaggtca ggagatcgag accatcccgg ctaaaacggt gaaaccccgt
ctctactaaa
 11341 actacaaaaa atagccgggc gtagtggcgg gcgcctgtag tcctagctac
ttgggaggct
 11401 gaggcaggag aatggcgtga acccgggagg cggagcttgc agtgagccga
gatcccgcca
 11461 ctgcactcca gcctgggcga cagagcgaga ctcc^atctc^a aaaaaaaaaa
aaaaaaaagg
       (11495)(a/g)(dbSNP:rs7246392), (11500)(-
/aaaaa)(dbSNP:rs10653764)
 11521 aatctctgaa ggcattacaa ttgcaaaaac aggctcagag taaagtgcaa
tatcttatct
 11581 gatcaggaca ccgtgtctat gatgtgacta caaattatca cttgcaatcc
cctgcaaca^t
       (11640)(t/c)(dbSNP:rs11881743)
 11641 gtgcaccaag ggggcacatt atctgcttta tactaatcaa agaagactga
acgagacagg
 11701 aaaagtgcaa tgaacacata tattttagat gggcaaaaaa aaaaaaaaaa
aaaaatagac
 11761 agcagtcact aatccatatt tttgctgaaa ccaaccagaa aactatttac
attttctctg
 11821 ggaatgtttt tctgtagcac ttgtccccac cctccattct cagtttcttg
gagttcctac
 11881 caagacggca gaatgacagc ccttttccca ctataagagc tgtgattatc
tacttcatag
 11941 catgtatgga aaaccactta ggagaa^caga aacaacatat tacctaaaaa
tccatgaagc
```

(11967)(c/t)(dbSNP:rs12980535)
12001 tagaatgaag agagttccag aaaatagcgt ctttgtttgc aaggaccatc
ttaggtgtca
12061 ttttgatcat tattttctct ccccatgttc attgcaacag tttctgcaac
tccaacctca
12121 ctcatgggta tgcctctctc cttacacttc ctttaaccta acttccagac
actttgaggt
12181 gccaacgcac agtgaatgaa actcttcatg cctctcttcc ttgtttcatt
ccaggatgga
12241 ataaaattcc ctgccctctc actggaaact cttgcgttct gatt^aggt^ac
actcttctcc
(12285)(a/g)(dbSNP:rs16973535), (12289)(a/g)(dbSNP:rs12981089)
12301 atacacttat acagtcttga tcacttctca ttttgataat ggatcataga
ccctatcacc
12361 agcatggaat tcatttcttt gggaactatg cagtttgtgc tctgagtgga
tgaaacctat
12421 cctggaaccc ctggcttcct gaacagaaga caacttggag tttatatcta
cttgctgctt
12481 gcctggcgcc cattttatgt gaataaccaa caaagtccta aatttcagcc
ctgctaggca
12541 aacacacagt gactgccctt tcttgaatga tattcaccag ctcagtgcac
ttcctatggg
12601 ccattcgtta cattcatctt tattcagcta ctactctgag ctcatagtgg
ccaggaatga
12661 catctcagat acaaagagat gacatagagt ccagggccaa tgttcaacta
aaaatttttg
12721 aattgcctct ggaaacctat tcaacatctt tcttcatcct gctcaatcca
aattcatgat
12781 cttctcccag caaatagggc attgctcaag aagcagaaag ccatttctga
actctaaggc
12841 ctcctcatct taggtctttt ggtcttccca gttattgctg aagcccccaa
agaataagta
12901 tactctgtga taatattgcc cttgaagccc agttccctca gcattgctat
tcctttggga
12961 tgccatctgt gatgctacta aacactggtg tatgttggcc tctatttgca
agtgaggggt
13021 caagaagatg gaataaaaga gaaatttaaa aaattttccc caatttattt
atgtttacca
13081 tttttatttg aagttcaagt gtacatgtgc aggtttgtta tataggttaa
tcctttgtca
13141 tgcgtgttag ttgtacagat tatttcatca cccaggtaac taaacaaagt
acccaatggt
13201 tatttttttct gctcttctct cttttcctac ccctccactc tcaggtaggc
cccagtttgt
13261 gttgttccct tctttgtgtt tatgagttct catcatttag ctcccaccta
taagtgagaa
13321 catgcagtat ttttttctgg tcctgcatta ctttgctaag gataatggcc
tctagctcta
13381 cctatgttcc cgcaaaagac atgaccttgt tctttttatg gctgcatagt
attccatggt
13441 gtatatgtac cacattttct gtattcagtc agccattaat gaacatttag
gttgattcca
13501 tgtctttgct attgtgaaca gtgctgcaat gaacacttgt gtgcatgtgt
ctttatggta
13561 gaatgattta tattctaata tttatatgta tttttatttc aataatttttt
gggacccagg
13621 aggtttttgg ttgcatggtt acattctttg gtggtgattt ttgagacttt
gctgtaccca
13681 taacctaagc agtgtacact gtacccaata tgtagcattt tatctgtcac
ccctcccacc
13741 cttccttcga gtctccaaag ttcattatat catcttaatg cctttgcatc

```
ctcatggttt
 13801 agctcccact tatgagtgag aacatacagc atctggtttt ttattcctga
gttgcttcac
 13861 ttagaacaat gacctccagt tgcatccaag ttgctgcaat ggccattatt
tctttctgct
 13921 tttatggctg agtagtatt^c catggtgcat atttaccaca tgttctttct
ccatttattg
       (13940)(-/c)(dbSNP:rs35368312)
 13981 atccatgg^gc acttaggtag attccatgcc actctaaaca tg^tgtgtgta
tgtctctttt
       (13989)(g/a)(dbSNP:rs10411383), (14023)(t/a)(dbSNP:rs10412744)
 14041 ctatataatg acttcttttc ttctggatag gtacccagca gtggaattac
tggatcaaat
 14101 ggttgttctg ctttcagttc ttgaaagaat ctccatattc ttttccacag
tggttgtact
 14161 agtttacagt cccacaagca gtgtaaaagc attccatttt caccacatcc
acaccaacat
 14221 atattctttt ttaatgttgt atttatgatc attcttgcag gactaatttg
gtatctcgtt
 14281 gtgggtttaa tttgcatttc actgatgatt agagatgttg aacatttttt
aatatgttta
 14341 ttggctgctt gtatatcctt ttttgtctcc ^gtactaaaac ^attattttat
ttttatgttg
       (14371)(g/a)(dbSNP:rs10412234), (14381)(a/g)(dbSNP:rs10411717)
 14401 tttttgaacc tattaattca gtttatttt ttccataaat tattggggta
caggcggtat
 14461 ttggttacat gaataagtta tttattggtg atttgtgaga ttttggtgca
cccatcaacc
 14521 aagcagtata cactgcatta tatttgcagt ctcttatccc tcacccttcc
ccaacacttc
 14581 ccttcaactc cccatagttc attatatcat ttttatgcct tggtgtcctc
atagcttagc
 14641 tcccacatat caatgagaac atatgatgtt tggtttttca ttcttgagtt
acttcactta
 14701 gaatgatagc cttcaaattc atccagatca ctgcaaatgc tgtgaattta
ttccttttta
 14761 tggctgcata gtattctgtt atgtgtgtgt gtatatatat ataccacagt
ttctttatcc
 14821 actcgttgaa tgatgggcat ttgtgttggt tccatgattt tgctactata
aattgtgtca
 14881 cagtaaacat gggtgtgcaa ggatcctttt tgaatattga cttcttttcc
tctgggtaga
 14941 tacccagtag
//
```

SUPPLEMENTARY DATA

Additional Statistical data

Statistical analyses: Patients vs Controls (genotype and allele frequencies)

-98 A/C

Coloured vs Black

Table 1

		Total Cohort N=302	Coloured N=214	Black N=88
Genotype frequency	**A/A**	153 (0.51)	103 (0.48)	50 (0.57)
	A/C	127 (0.42)	97 (0.45)	30 (0.34)
	C/C	22 (0.07)	14 (0.07)	8 (0.09)
Allele frequency	**A**	433 (0.72)	303 (0.71)	130 (0.74)
	C	171 (0.28)	125 (0.29)	46 (0.26)

Allele
TABLE = [303 , 130 , 125 , 46]
Left : p-value = 0.25530070484288925
Right : p-value = 0.80470850518645
2-Tail : p-value = 0.4872034916292145

Genotype
Chi^2: 3.36
p= 0.186

Coloured

Adverse outcomes

Table 2

PP13

	Num	freq	p^2	2pq	q^2
A / A	9	0.391	0.484	0.423	0.093
A / C	14	0.609			
C / C	0	0.000			
No of Individuals	23				
A	32	0.696			
C	14	0.304			
No of alleles	46				

HWE

4.4023438	= Chi-Squared
2DF, p =	0.1107

Controls

Table 3

	Num	freq	p^2	2pq	q^2
A / A	38	0.551	0.536	0.392	0.072
A / C	25	0.362			
C / C	6	0.087			
No of Individuals	69				
A	101	0.732			
C	37	0.268			
No of alleles	138				

HWE

0.406974	= Chi-Squared
2DF, p =	0.8159

Adverse vs Controls

Allele

	Adverse outcomes	Controls
C	14	37
A	32	101

Table 4

Odds ratio <0.54/>2.63 Cornfield 95% confidence limits	1.19
Relative Risk <1.14/>1.96 Taylor Series 95% confidence limits	1.14
Uncorrected	0.634
Mantel-Haenzel	0.635
Yates Corrected	0.775

TABLE = [14 , 37 , 32 , 101]
Left : p-value = 0.7495642333128372
Right : p-value = 0.3828515082058961
2-Tail : p-value = 0.7043058806446271

Genotype

	Adverse outcomes	Controls
AA	9	38
AC	14	25
CC	0	6

Chi^2: 5.33 (not valid)
p= 0.07

Black

Adverse
Table 5

	Num	freq	p^2	2pq	q^2
A / A	5	0.625	0.473	0.430	0.098
A / C	1	0.125			
C / C	2	0.250			
No of Individuals	8				

A	11	0.688			
C	5	0.313			
No of alleles	16				

HWE

4.0224793	= Chi-Squared
2DF, p =	0.1338

Controls
Table 6

	Num	**freq**	**p^2**	**2pq**	**q^2**
A / A	19	0.559	0.541	0.389	0.070
A / C	12	0.353			
C / C	3	0.088			
No of Individuals	34				
A	50	0.735			
C	18	0.265			
No of alleles	68				

HWE

0.2961778	= Chi-Squared
2DF, p =	0.8624

Adverse vs Controls

Allele

	Adverse outcomes	Controls
C	5	18
A	11	50

Table 7

Odds ratio <0.33/>4.71 Cornfield 95% confidence limits		1.26
Relative Risk <0.47/>3.09 Taylor Series 95% confidence limits		1.21
Uncorrected		0.700
P-values	Mantel-Haenzel	0.701
	Yates Corrected	0.941
	Fisher exact 1-tailed	0.459
	Fisher exact 2-tailed	0.758

TABLE = [5 , 18 , 11 , 50]
Left : p-value = 0.761594527771471
Right : p-value = 0.45866757579496736
2-Tail : p-value = 0.7584845355630441

Genotype

	Adverse outcomes	Controls
AA	5	19
AC	1	12
CC	2	3

Chi^2: 2.56 (not valid)
p= 0.278

-22 A/G
Black vs Coloured
Table 8

		Total Cohort N=307	Coloured N=219	Black N=88
Genotype frequency	**A/A**	221 (0.72)	164 (0.75)	57 (0.65)
	A/G	81 (0.26)	53 (0.24)	28 (0.32)
	G/G	5 (0.02)	2 (0.01)	3 (0.03)
Allele frequency	**A**	523 (0.85)	381 (0.87)	142 (0.81)
	G	91 (0.15)	57 (0.13)	34 (0.19)

Genotype
Chi^2: 4.67
p-value= 0.09665

Allele
Table 9

Odds ratio <0.38/>1.02 Cornfield 95% confidence limits	0.62
Relative Risk <0.73/>1.02 Taylor Series 95% confidence limits	0.86
Uncorrected	0.047
Mantel-Haenzel	0.047
Yates Corrected	0.63

TABLE = [381 , 142 , 57 , 34]
Left : p-value = 0.981150082205703
Right : p-value = 0.0331258254261578
2-Tail : p-value = 0.0589732050528019

Coloured

Adverse
Table 10

	Num	freq	p^2	2pq	q^2
A / A	16	0.696	0.719	0.258	0.023
A / G	7	0.304			
G / G	0	0.000			
No of Individuals	23				
A	39	0.848			
G	7	0.152			
No of alleles	46				

HWE

0.7409599	= Chi-Squared
2DF, p =	0.6904

Controls
Table 11

	Num	freq	p^2	2pq	q^2
A / A	49	0.690	0.679	0.290	0.031

A / G	19	0.268			
G / G	3	0.042			
No of Individuals	71				
A	117	0.824			
G	25	0.176			
No of alleles	142				

HWE

0.4276203	= Chi-Squared
2DF, p =	0.8075

Adverse vs Controls

Allele

	Adverse outcomes	Controls
A	39	117
G	7	25

Table 12

Odds ratio <0.30/>2.25 Cornfield 95% confidence limits	0.84
Relative Risk <0.43/>1.78 Taylor Series 95% confidence limits	0.88
Uncorrected	0.708
Mantel-Haenzel	0.709
Yates Corrected	0.882

TABLE = [39 , 117 , 7 , 25]
Left : p-value = 0.7198191638972884
Right : p-value = 0.45103356764607616
2-Tail : p-value = 0.8235366385013259

Genotype

	Adverse outcomes	Controls
AA	16	49
AG	7	19
GG	0	3

Chi^2: 1.06
p-value= 0.589

Black

Adverse
Table 13

	Num	Freq	p^2	2pq	q^2
A / A	5	0.625	0.563	0.375	0.063
A / G	2	0.250			
G / G	1	0.125			
No of Individuals	8				
A	12	0.750			
G	4	0.250			

No of alleles	16				

HWE

0.8888889	= Chi-Squared
2DF, p =	0.6412

Controls
Table 14

	Num	**Freq**	**p^2**	**2pq**	**q^2**
A / A	21	0.618	0.631	0.327	0.042
A / G	12	0.353			
G / G	1	0.029			
No of Individuals	34				
A	54	0.794			
G	14	0.206			
No of alleles	68				

HWE

0.2141597	= Chi-Squared
2DF, p =	0.8985

Adverse vs Controls

Allele

	Adverse outcomes	Controls
A	12	54
G	4	14

Table 15

Odds ratio <0.30/>5.28 Cornfield 95% confidence limits		1.29
Relative Risk <0.45/>3.34 Taylor Series 95% confidence limits		1.22
P values	Uncorrected	0.699
	Mantel-Haenzel	0.701
	Yates Corrected	0.961
	Fisher's exact 1 tailed	0.464
	Fisher's exact 2 tailed	0.739

TABLE = [12 , 54 , 4 , 14]
Left : p-value = 0.4642891280204133
Right : p-value = 0.7715599358562616
2-Tail : p-value = 0.7387518061672984

Genotype

	Adverse outcomes	Controls
AA	5	21
AG	2	12
GG	1	1

Chi^2: 1.45
p-value= 0.485

-36 G/A

Black vs Coloured
Table 16

		Total Cohort N=309	Coloured N=219	Black N=90
Genotype frequency	**G/G**	297 (0.96)	210 (0.96)	87 (0.97)
	G/A	12 (0.04)	9 (0.04)	3 (0.03)
	A/A	0	0	0
Allele frequency	**G**	0.98	0.98	0.97
	A	0.02	0.02	0.03

Genotype:
Chi^2: 0.10
p-value= 0.748

Allele:
Table 17

Odds Ratio <0.30/>5.83		1.24
Relative Risk <0.76/>1.47 Taylor Series 95% confidence limits		1.06
P values	Uncorrected	0.75
	Mantel-Haenzel	0.75
	Yates Corrected	0.99
	Fisher's exact 1 tailed	0.52
	Fisher's exact 2 tailed	1.00

TABLE = [429 , 177 , 9 , 3]
Left : p-value = 0.5190781120215455
Right : p-value = 0.7281604837188726
2-Tail : p-value = 1

<u>Coloured</u>
Adverse
Table 18

	Num	Freq	p^2	2pq	q^2
G / g	21	0.913	0.915	0.083	0.002
G / a	2	0.087			
A / a	0	0.000			
No of Individuals	23				
G	44	0.957			
A	2	0.043			
No of alleles	46				

HWE

0.0475207	= Chi-Squared
2DF, p =	0.9765

Controls
Table 19

	Num	Freq	p^2	2pq	q^2
G / g	67	0.957	0.958	0.042	0.000
G / a	3	0.043			
A / a	0	0.000			
No of Individuals	70				
G	137	0.979			
A	3	0.021			
No of alleles	140				

HWE

0.033566	= Chi-Squared
2DF, p =	0.9834

Adverse vs Controls

Allele

	Adverse outcomes	Controls
G	44	137
A	2	3

Table 20

Odds ratio <0.23/>15.94 Cornfield 95% confidence limits		2.08
Relative Risk <0.55/>4.96 Taylor Series 95% confidence limits		1.65
P values	Uncorrected	0.422
	Mantel-Haenzel	0.424
	Yates Corrected	0.782
	Fisher's exact 1 tailed	0.361
	Fisher's exact 2 tailed	0.599

TABLE = [44 , 137 , 2 , 3]
Left : p-value = 0.36144457112195383
Right : p-value = 0.9021687147017676
2-Tail : p-value = 0.5987219982459879

Genotype

	Adverse outcomes	Controls
GG	21	67
GA	2	3
AA	0	0

Chi^2: 0.66 (not valid)
p-value= 0.416

Fisher's exact test recommended 1-tailed: 0.361
2-tailed: 0.594

Black

Adverse
Table 21

	Num	Freq	p^2	2pq	Q^2
G/G	8	1.000	1.000	0.000	0.000
G/A	0	0.000			
A/A	0	0.000			
No of Individuals	8				
G	16	1.000			
A	0	0.000			
No of alleles	16				

HWE
Chi^2: not computable
p-value= not computalbe

Controls
Table 22

	Num	freq	p^2	2pq	Q^2
G / g	33	0.971	0.971	0.029	0.000
G / a	1	0.029			
A / a	0	0.000			
No of Individuals	34				
G	67	0.985			
A	1	0.015			
No of alleles	68				

HWE

0.0075741	= Chi-Squared
2DF, p =	0.9962

Adverse vs Controls

Allele

	Adverse outcomes	Controls
G	16	67
A	0	1

Table 23

Odds ratio <0.00/>77.80 Cornfield 95% confidence limits		0.00
Relative Risk		Not valid
P values	Uncorrected	0.626
	Mantel-Haenzel	0.628
	Yates Corrected	0.428
	Fisher's exact 1 tailed	0.810
	Fisher's exact 2 tailed	1.00

TABLE = [16 , 67 , 0 , 1]
Left : p-value = 1
Right : p-value = 0.8095238095238063

2-Tail : p-value = 1

Genotype

	Adverse outcomes	Controls
GG	8	33
GA	0	1
AA	0	0

Chi^2: 0.24 (not valid)
p-value= 0.623

Fisher's exact test recommended: 1-tailed: 0.810
2-tailed: 1.000

-+72 T/A

Black vs Coloured

Genotype
Table 24

		Total Cohort N=307	Coloured N=217	Black N=90
Genotype frequency	**T/T**	235 (0.77)	176 (0.81)	59 (0.66)
	T/A	63 (0.21)	37 (0.17)	26 (0.29)
	A/A	9 (0.03)	4 (0.02)	5 (0.06)
Allele frequency	**T**	533 (0.87)	389 (0.90)	138 (0.80)
	A	81 (0.13)	45 (0.10)	36 (0.20)

Chi^2: 9.34 (not valid)
p-value= 0.009

Allele
Table 25

Odds ratio <0.27/>0.74 Cornfield 95% confidence limits		0.44
Relative Risk <0.75/>0.92 Taylor Series 95% confidence limits		0.75
P values	Uncorrected	0.0007
	Mantel-Haenzel	0.0007
	Yates Corrected	0.001

TABLE = [389 , 138 , 45 , 36]
Left : p-value = 0.9996681283754402
Right : p-value = 0.0008007828469592797
2-Tail : p-value = 0.0013712615372336056

Coloured

Adverse
Table 26

	Num	freq	p^2	2pq	q^2
T / T	21	0.913	0.915	0.083	0.002
T / A	2	0.087			
A / A	0	0.000			
No of Individuals	23				
T	44	0.957			
A	2	0.043			
No of alleles	46				

0.0475207 = Chi-Squared
2DF, p = 0.9765

Controls
Table 27

	Num	Freq	p^2	2pq	q^2
T / T	51	0.729	0.710	0.265	0.025
T / A	16	0.229			
A / A	3	0.043			
No of Individuals	70				
T	118	0.843			
A	22	0.157			
No of alleles	140				

1.3164024 = Chi-Squared
2DF, p = 0.5178

Adverse vs Controls
Table 28: Observed genotype and allele frequencies at the +72 T/A locus in the study participants with known pregnancy outcomes from the study cohort.

Population group		**Coloured N=92**		**Black N=42**	
Pregnancy outcome		**Adverse N=23**	**Controls N=70**	**Adverse N=8**	**Controls N=34**
Genotype frequency	**T/T**	21 (0.91)	51 (0.73)	6 (0.76)	22 (0.64)
	T/A	2 (0.09)	16 (0.23)	1 (0.12)	10 (0.30)
	A/A	0 (0.00)	3 (0.04)	1 (0.12)	2 (0.06)
Allele frequency	**T**	44 (0.96)	118 (0.84)	13 (0.81)	54 (0.78)
	A	2 (0.04)	22 (0.16)	3 (0.19)	14 (0.20)

Allele
Table 29

	Adverse outcomes	Controls
T	44	118
A	2	22

Table 30

Odds ratio <0.04/>1.14 Cornfield 95% confidence limits		0.24
Relative Risk <0.08/>1.18 Taylor Series 95% confidence limits		0.31
P values	Uncorrected	0.046
	Mantel-Haenzel	0.047
	Yates Corrected	0.082

Fisher's Exact Test
http://www.langsrud.com/fisher.htm

TABLE = [381 , 142 , 57 , 34]
Left : p-value = 0.981150082205703
Right : p-value = 0.0331258254261578
2-Tail : p-value = 0.0589732050528019

TABLE = [44 , 118 , 2 , 22]
Left : p-value = 0.993600596420201
Right : p-value = 0.032969641806632914

2-Tail : p-value = 0.04616584747980334

Genotype

	Adverse outcomes	Controls
TT	21	51
TA	2	16
AA	0	3

Chi^2: 3.54 (not valid)
p-value= 0.170

Black

Adverse
Table 31

	Num	Freq	p^2	2pq	q^2
T / T	6	0.750	0.660	0.305	0.035
T / A	1	0.125			
A / A	1	0.125			
No of Individuals	8				
T	13	0.813			
A	3	0.188			
No of alleles	16				

2.78238 = Chi-Squared
2DF, p = 0.2488

Controls
Table 32

	Num	freq	p^2	2pq	q^2
T / T	22	0.647	0.631	0.327	0.042
T / A	10	0.294			
A / A	2	0.059			
No of Individuals	34				
T	54	0.794			
A	14	0.206			
No of alleles	68				

0.3436074 = Chi-Squared
2DF, p = 0.8421

Adverse vs Controls

Allele

	Adverse outcomes	Controls
T	13	54
A	3	14

Table 33

Odds ratio <0.17/>4.06 Cornfield 95% confidence limits		0.89
Relative Risk <0.29/>2.84 Taylor Series 95% confidence limits		0.91
P values	Uncorrected	0.869
	Mantel-Haenzel	0.880
	Yates Corrected	0.856
	Fisher's exact 1 tailed	0.588
	Fisher's exact 2 tailed	1.00

TABLE = [13 , 54 , 3 , 14]
Left : p-value = 0.6819959562143499
Right : p-value = 0.5881217750632177
2-Tail : p-value = 1

Genotype

	Adverse outcomes	Controls
TT	6	22
TA	1	10
AA	1	2

Chi^2: 1.21 (not valid)
p-value= 0.547

229delT

Coloured
Table 34

	Num	Freq	p^2	2pq	q^2
Wildtype/Wildtype	218	0.995	0.995	0.005	0.000
Wildtype/DelT	1	0.005			
DelT/DelT	0	0.000			
No of Individuals	219				
Wildtype	437	0.998			
DelT	1	0.002			
No of alleles	438				

0.0011468 = Chi-Squared
2DF, p = 0.9994

Black

Table 35	Num	freq	p^2	2pq	q^2
Wildtype/Wildtype	88	0.978	0.978	0.022	0.000
Wildtype/DelT	2	0.022			
DelT/DelT	0	0.000			
No of Individuals	90				
Wildtype	178	0.989			
DelT	2	0.011			
No of alleles	180				

0.0113622	= Chi-Squared
2DF, p =	0.9943

"hotspot"

Coloured

Table 36	Num	freq	p^2	2pq	q^2
Wildtype / wildtype	204	0.932	0.933	0.066	0.001
Wildtype / hotspot	15	0.068			
hotspot / hotspot	0	0.000			
No of Individuals	219				
Wildtype	423	0.966			
Hotspot	15	0.034			
No of alleles	438				

0.2753886	= Chi-Squared
2DF, p =	0.8714

Black

Table 37	Num	freq	p^2	2pq	q^2
wildtype / wildtype	83	0.922	0.924	0.075	0.002
wildtype / hotspot	7	0.078			
hotspot / hotspot	0	0.000			
No of Individuals	90				
Wildtype	173	0.961			
Hotspot	7	0.039			
No of alleles	180				

0.1473487	= Chi-Squared
2DF, p =	0.9290

Statistical analyses: Demographic data

Adverse outcomes vs Controls, Black vs Coloured

TABLE = [8 , 23 , 34 , 72]
Left : p-value = 0.33334687402200535
Right : p-value = 0.8114725536028717
2-Tail : p-value = 0.6585159207242672

Table 38

	Black		**Coloured**	
	Adverse pregnancy outcome N=7	**Good pregnancy outcome N=24**	**Adverse pregnancy outcome N=19**	**Good pregnancy outcome N=67**
BMI (mean value)	**27.857 (range: 22-37)**	**29.586 (range: 20-40)**	**26.809 (range: 17.6-34.3)**	**24.358 (range: 17.6-45.6)**
Gestation (mean value weeks)	**37.375 (range: 36-40)**	**40.125 (range: 38-42)**	**39.536 (range:27-39)**	**35.957 (range: 35-42)**
Birth weight (mean value kg)	**2.862 (range: 2060-4460)**	**3.628 (2700-4540)**	**2.680 (range: 800-3240)**	**2.600 (range: 1880-4560)**

RNA extraction from placental tissue

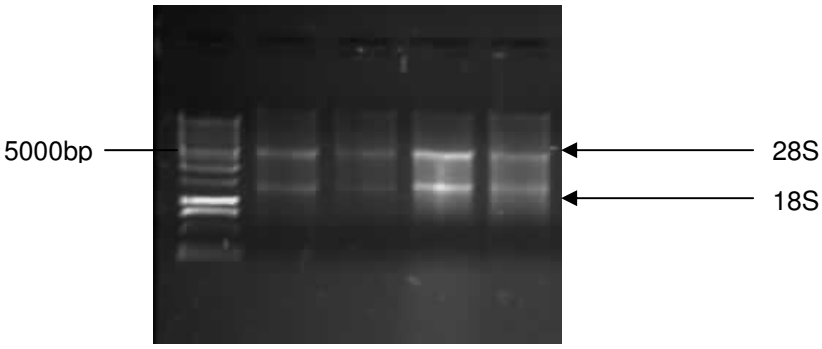


Figure 1: Total RNA extracted from placental tissue. Visualised on 1% agarose gel containing guanidium thiocyanate and stained with EtBr. The 28S and 18S rRNA fragments are visible.

Gene-specific amplification of LGALS13 and LOC148003 from genomic DNA

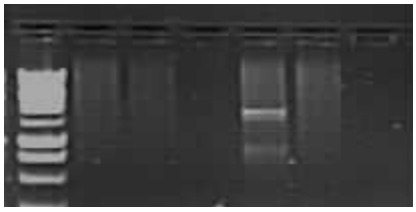


Figure 2: 2000bp LOC148003 fragment amplified from genomic DNA is visible on this 1% agarose gel.

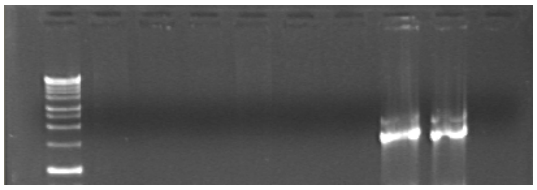


Figure 3: 2000bp *LGALS13* fragment amplified from genomic DNA is visible on this 1% agarose gel.

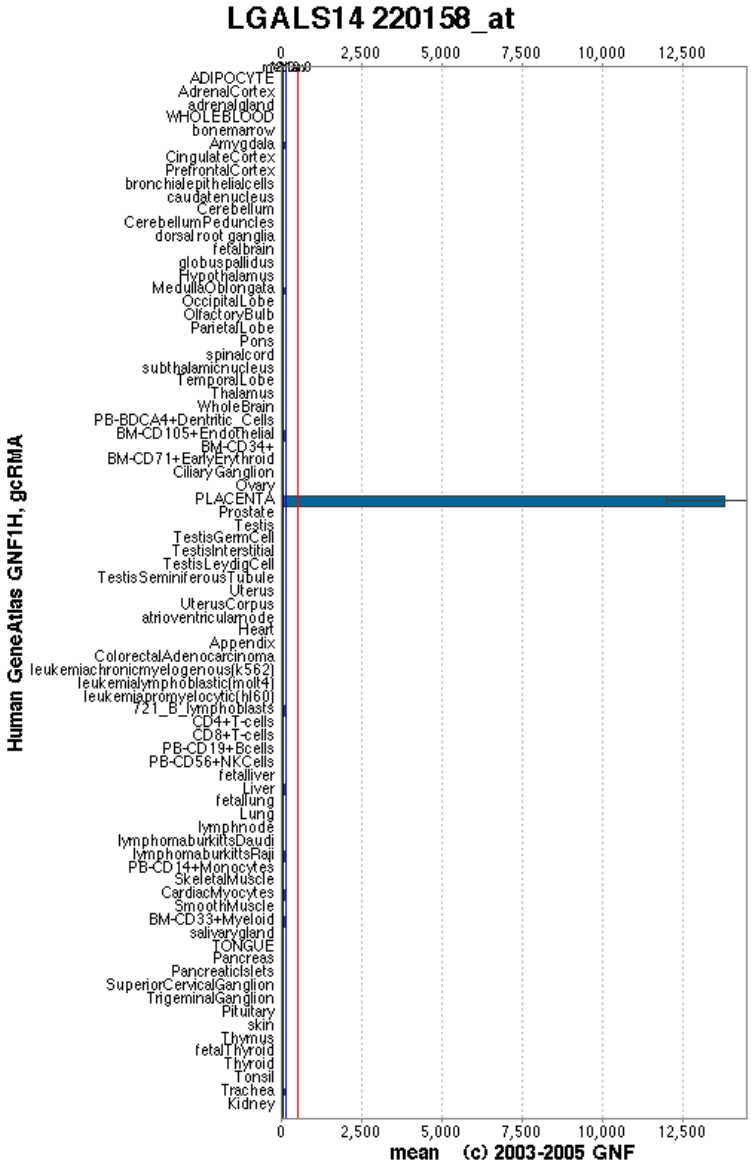


Figure 4: Tissue –specific expression pattern of *LGALS14.* Exclusively expressed in the placenta.

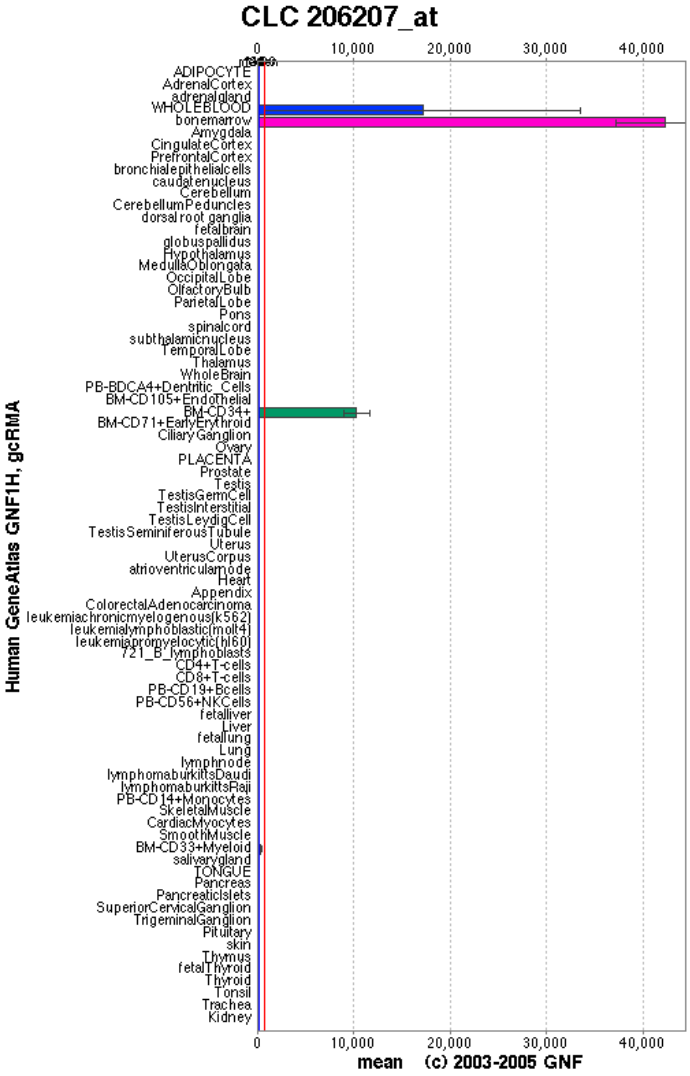
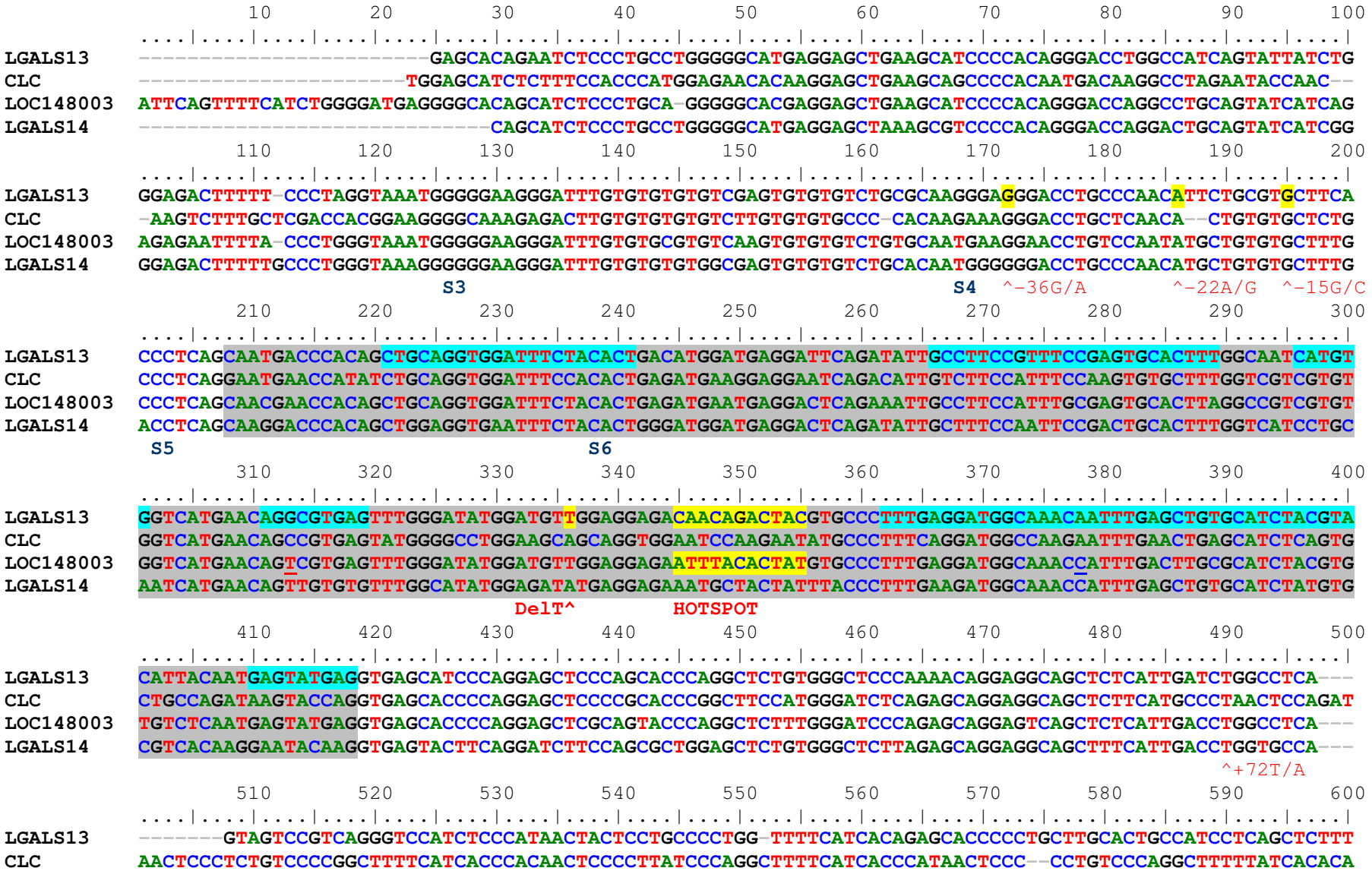


Figure 5: Expression profile of Galectin 10 (CLC). Greatest levels of expression in the bonemarrow, wholeblood and CD34+ cells.


LGALS13
CLC
LOC148003
LGALS14
S3
S4
^-36G/A
^-22A/G
^-15G/C
S5
S6
DelT^
HOTSPOT
^+72T/A

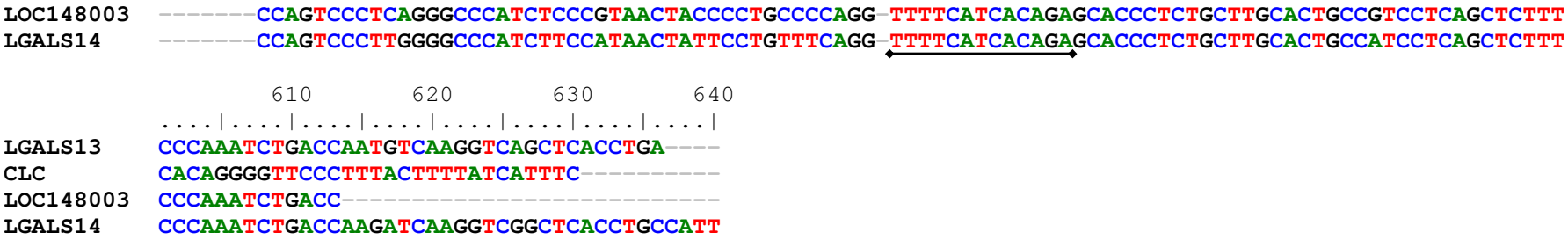


Figure 6: Multiple sequence alignment of exon 3 in different members of the galectin multigene family.Section of alignment which is shaded in grey represents the exon, surrounding sequence is intronic. Yellow blocks indicate areas of high sequence conservation and DNA sequence regions responsible for encoding functionally important β-sheets.