Πανεπιστήμιο Κρήτης

Τμήμα Ιατρικής

Μεταπτυχιακό Πρόγραμμα: "Κυτταρική και Γενετική Αιτιολογία, Διάγνωστική και Θεραπευτική των Ασθενειών του Ανθρώπου"

Μεταπτυχιακή Διατριβή

ΜΗΧΑΝΙΣΜΟΙ ΡΥΘΜΙΣΗΣ ΤΩΝ ΓΟΝΙΔΙΩΝ ΠΟΥ ΕΜΠΛΕΚΟΝΤΑΙ ΣΤΟ ΜΟΝΟΠΑΤΙ ΤΗΣ ΛΙΠΟΠΡΩΤΕΪΝΗΣ ΥΨΗΛΗΣ ΠΥΚΝΟΤΗΤΑΣ (HDL) ΣΕ ΗΠΑΤΟΚΥΤΤΑΡΑ

Τρακάκη Αθηνά

Επιβλέπων Καθηγητής: Καρδάσης Δημήτριος

Επιτροπή: Ηλιόπουλος Αριστείδης, Τσατσάνης Χρήστος

Εργαστήριο Βιοχημείας
Τομέας Βασικών Επιστημών

Νοέμβριος 2015

University of Crete

Medical School

Graduate program: "Cellular and Genetic Etiology, Diagnosis and Treatment of Human Disease"

Master Thesis

REGULATORY MECHANISMS OF THE GENES INVOLVED IN THE HIGH DENSITY LIPOPROTEIN (HDL) PATHWAY IN HEPATOCYTES

Trakaki Athina

Supervisor: Professor Kardassis Dimitrios

Committee: Eliopoulos Aristides, Tsatsanis Christos

Laboratory of Biochemistry
Department of Basic Sciences

November 2015

Table of Contents

Περίληψη

Οι καρδιαγγειακές παθήσεις (ΚΑΠ) αποτελούν τη νούμερο ένα αιτία θανάτου παγκοσμίως. Υπάρχουν αδιαμφισβήτητες αποδείξεις, ότι τα επίπεδα της HDL (υψηλής πυκνότητας λιποπρωτεΐνη) χοληστερόλης στο πλάσμα συνδέονται αντίστροφα με τον κίνδυνο για ΚΑΠ. Η απολιποπρωτεΐνη Ε (apoE) φαίνεται, να κατέχει μία διπλή λειτουργία: εκτός από το ρόλο της στην εκκαθάριση των πλούσιων σε τριακυλογλυκερόλη λιποπρωτεϊνών, συμμετέχει στη βιογένεση σωματιδίων HDL που περιέχουν apoE. Η πρωτεΐνη μεταφοράς χοληστερυλεστέρα (CETP) καταλύει την ανταλλαγή τριγλυκεριδίων από τα σωματίδια που περιέχουν απολιποπρωτεΐνη Β (apoB), όπως η LDL και η VLDL, με εστέρες χοληστερόλης από τον πυρήνα των HDL σωματιδίων. Η δράση της καθορίζει τα επίπεδα της HDL χοληστερόλης στο πλάσμα. Πολυμορφισμοί παρατηρούνται συχνά στα γονίδια της apoE και του CETP και μπορούν να επηρεάσουν σημαντικά τα επίπεδα των λιποπρωτεϊνών στο πλάσμα, ενώ το τελευταίο γονίδιο θεωρείται υποψήφιο για στεφανιαία νόσο. Έτσι, στο πρώτο κεφάλαιο της παρούσας διατριβής, είχαμε ως στόχο την αναγνώριση γενετικών αλλαγών στα γονίδια της apoE και του CETP στον άνθρωπο, που αναγνωρίστηκαν από την ομάδα της Δρ. AnneTybjærg Hansen (Πανεπιστημιακό Νοσοκομείο της Κοπεγχάγης) στην καρδιακή μελέτη της πόλης της Κοπεγχάγης (CCHS), οι οποίες μπορεί να συνεισφέρουν στην ενεργοποίηση ή στην καταστολή των αντίστοιχων γονιδίων, καθώς και η αναγνώριση των πιθανών μηχανισμών που είναι υπεύθυνοι για αυτό το μοτίβο ρύθμισης. Για αυτό το σκοπό κατασκευάσαμε και χρησιμοποιήσαμε φορείς έκφρασης του γονιδίου της λουσιφεράσης, οι οποίοι έφεραν την περιοχή του υποκινητή του CETP γονιδίου ή/και το πρώτο ιντρόνιο του CETP γονιδίου, καθώς και φορείς έκφρασης του γονιδίου της λουσιφεράσης που έφεραν την περιοχή του υποκινητή του γονιδίου της apoE, με τα πιο κοινό ή το λιγότερο κοινό αλληλόμορφο για κάθε πολυμορφισμό, και πραγματοποιήσαμε παροδικές διαμολύνσεις σε ανθρώπινα κύτταρα ηπατοβλαστώματος (HepG2). Δείξαμε ότι δύο πολυμορφισμοί στον υποκινητή του CETP γονιδίου, οι -656C>A και -65G>A, καταστέλλουν και ενεργοποιούν τη δραστικότητα του υποκινητή αντίστοιχα, όταν είναι παρών το λιγότερο κοινό αλληλόμορφο σε κάθε περίπτωση. Ακόμη, η περιοχή +191/+1056 του πρώτου ιντρονίου του γονιδίου του CETP του ανθρώπου, δείχθηκε ότι καταστέλλει σημαντικά την δραστικότητα του υποκινητή. Δείχθηκε, ότι ο συνδυασμός του λιγότερο κοινού αλληλομόρφου -656A, του πολυμορφισμού -656C>A στον υποκινητή του CETP, με το πιο κοινό αλληλόμορφο +279G, του πολυμορφισμού +279G>A στο πρώτο ιντρόνιο του CETP, μειώνει σημαντικά τη δραστικότητα του εν λόγω υποκινητή. Επιπλέον το λιγότερο κοινό αλληλόμορφο -427C, του πολυμορφισμού -427T>C στον υποκινητή του γονιδίου της apoE στον άνθρωπο, δείχθηκε ότι καταστέλλει τη δραστικότητα του υποκινητή της apoE σε στατιστικά σημαντικό βαθμό.

Ο μεταγραφικός παράγοντας HNF-4α είναι μία πυρηνική πρωτεΐνη, που προσδένει DNA αποκλειστικά ως ομοδιμερές και ενεργοποιεί την έκφραση ποικιλίας γονιδίων, κυρίως εκείνων που εμπλέκονται στον ενδιάμεσο μεταβολισμό, συμπεριλαμβανομένου του μεταβολισμού λιπαρών οξέων, γλυκόζης, ξενοβιοτικών, χοληστερόλης και φαρμάκων, καθώς και άλλων παραγόντων που εμπλέκονται στην πήξη του αίματος, τη βιοσύνθεση ουρίας, τις μολύνσεις από ηπατίτιδα Β, και τη διαφοροποίηση του ήπατος. Είναι απαραίτητος τόσο στην πρώιμη όσο και στην ενήλικη ανάπτυξη και εκφράζεται κυρίως στο ενήλικο ήπαρ, το έντερο, το πάγκρεας, τα νεφρά και το παχύ έντερο. Οι καρβοξυλεστεράσες (CES) των θηλαστικών αποτελούν μία οικογένεια γονιδίων, τα προϊόντα των οποίων εντοπίζονται στο ενδοπλασματικό δίκτυο (ΕΔ) πολλών ιστών. Αυτά τα ένζυμα καταλύουν ικανά την υδρόλυση μία ποικιλίας χημικών, που περιέχουν αμίδιο ή εστέρα, καθώς και φαρμάκων και προφαρμάκων, στα αντίστοιχα ελεύθερα οξέα. Τουλάχιστον 20 γονίδια Ces έχουν αναγνωριστεί στο ποντίκι, δύο εκ των οποίων, η καρβοξυλεστεράση 3 άλφα (Ces3a) και η καρβοξυλεστεράση 3 βήτα (Ces3b), αποτελούν τα Ces3 γονίδια στο ποντίκι. Ωστόσο, δε γνωρίζουμε πολλά για τη Ces3b, εκτός από το ότι εκφράζεται κυρίως στο ήπαρ του ποντικού. Προκαταρκτικά δεδομένα από το εργαστήριό μας, χρησιμοποιώντας μικροσυστοιχίες σε ποντίκια που είχε γίνει αποσιώπηση του HNF-4α στο ήπαρ σε σύγκριση με ποντίκια μάρτυρες, έδειξε ένα σημαντικό αριθμό γονιδίων της οικογένειας των καρβοξυλεστερασών στο ποντίκι, τα οποία ενεργοποιούνται ή καταστέλλονται απουσία του HNF-4α από το ήπαρ του ποντικού. Ανάμεσα στις κατασταλμένες καρβοξυλεστεράσες, η Ces3b εμφάνισε τη μεγαλύτερη αλλαγή στα επίπεδα έκφρασης. Έτσι, σκοπός του δευτέρου κεφαλαίου της παρούσας διατριβής, ήταν η εξακρίβωση, εάν η έκφραση της Ces3b στο ποντίκι εξαρτάται από τον HNF-4α μεταγραφικό παράγοντα. Για να αναγνωριστούν πιθανές θέσεις πρόσδεσης για τον HNF-4α στον υποκινητή της Ces3b του ποντικού, πραγματοποιήσαμε παροδικές διαμολύνσεις σε ανθρώπινα εμβρυϊκά νεφρικά κύτταρα (HEK), χρησιμοποιώντας μία σειρά από φορείς έκφρασης του γονιδίου της λουσιφεράσης, οι οποίοι έφεραν διαδοχικές απαλοιφές στο 5’ άκρο ή μία απαλοιφή στο 3’ άκρο για τον υποκινητή του γονιδίου της Ces3b του ποντικού, μαζί με ένα φορέα έκφρασης του cDNA του HNF-4α. Παρατηρήσαμε ότι, η υπερέκφραση του μεταγραφικού παράγοντα HNF-4 ενεργοποίησε σημαντικά τη δραστικότητα του υποκινητή της Ces3b του ποντικού για όλα τα διαφορετικά τμήματα του υποκινητή της Ces3b που εξετάστηκαν, εκτός από το πιο βραχύ τμήμα που έφερε την περιοχή -80 έως +4 του υποκινητή της Ces3b του ποντικού. Μία πιθανή περιοχή πρόσδεσης για τον HNF-4α καθώς και μία πιθανή περιοχή πρόσδεσης για το μεταγραφικό παράγοντα COUP έχουν αναγνωριστεί στην περιοχή έπειτα από τη θέση έναρξης της μεταγραφής +1. Είναι γνωστό, ότι οι HNF-4α και COUP μεταγραφικοί παράγοντες μοιράζονται κοινή αλληλουχία αναγνώρισης, έτσι ενδεχομένως κάποια από αυτές τις δύο περιοχές θα

μπορούσε να ευθύνεται για την πρόσδεση του HNF-4α στα πειράματα διαμόλυνσης που πραγματοποιήσαμε.

Συνοψίζοντας, τα ευρήματα της παρούσας διατριβής αυξάνουν την κατανόησή μας πάνω στους μηχανισμούς, που ρυθμίζουν την έκφραση των γονιδίων που εμπλέκονται στο μεταβολισμό των λιπιδίων και των λιποπρωτεϊνών στο ήπαρ και πώς αυτοί οι μηχανισμοί απορρυθμίζονται μέσω της αποσιώπησης ηπατικών πυρηνικών παραγόντων, ή μέσω φυσικών πολυμορφισμών στους υποκινητές ή στους ενισχυτές των εκάστοτε γονιδίων.

Abstract

Cardiovascular diseases (CVDs) are the number one cause of death globally. There is unequivocal evidence that high density lipoprotein (HDL) cholesterol levels in plasma are inversely associated with the risk of CVDs. Apolipoprotein E (ApoE) appears to possess a dual function: apart from its role in the clearance of triacylglycerol-rich lipoproteins, it participates in the HDL-sized apoE-containing particles biogenesis. Cholesteryl ester transfer protein (CETP) catalyzes the exchange of triglycerides from apoB-containing particles, such as LDL and VLDL, for cholesteryl esters from the core of HDL particles. Its activity determines the plasma levels of HDL cholesterol. Polymorphisms are observed frequently on the apoE and CETP genes and can significantly influence levels of lipoproteins in the plasma, and the latter gene is considered to be candidate for coronary heart disease. Therefore, in the first chapter of the present thesis, we aimed to identify genetic changes in the human apoE and CETP genes, identified by the group of Dr Anne Tybjaerg-Hansen (Copenhagen University Hospital) in The Copenhagen City Heart Study (CCHS), which may contribute to an up- or down-regulation of the corresponding genes, as well as the establishment of the possible mechanisms responsible for this pattern of regulation. For that purpose we generated and utilized luciferase expression vectors bearing the CETP promoter region or/and the CETP first intron region, as well as luciferase expression vectors bearing the apoE promoter region, bearing either the most common or the less common allele for each SNP and we performed transient transfection assays in human hepatoblastoma (HepG2) cells. We showed that two CETP promoter polymorphisms, -656C>A and -65G>A down-regulate and an up-regulate the promoter activity respectively, when the less common allele is present in each case. Moreover, the first intronic region +191/+1056 of the human CETP gene was shown to significantly down-regulate the promoter's activity. Importantly it was shown that the combination of the less common -656A allele, of CETP promoter polymorphism -656C>A, with the most common +279G allele, of the CETP first intron polymorphism +279G>A, significantly reduces the activity of the corresponding promoter. In addition, the less common -427C allele, of human apoE promoter polymorphism -427T>C, was shown to down-regulate the apoE promoter activity at a statistically significant degree.

Hepatocyte nuclear factor 4 alpha (HNF-4α) is a nuclear receptor which activates the expression of a wide variety of genes, especially those involved in intermediary metabolism, including fatty acid, glucose, xenobiotic, cholesterol and drug metabolism, as well as others involved in blood coagulation, urea biosynthesis, hepatitis B infections, and liver differentiation. It appears to be essential both in early as well as in adult development and is mainly expressed in the adult liver, intestine, pancreas, kidney and colon. Mammalian carboxylesterases (CES) comprise a family of genes, whose products are localized in the endoplasmic reticulum of many tissues. These enzymes can efficiently catalyze the hydrolysis of a variety of amide- and ester-containing chemicals, as well as drugs and prodrugs, to the respective free acids. At least 20 mouse Ces genes have been recognized, two of which, carboxylesterase 3 alpha (Ces3a) and carboxylesterase 3 beta (Ces3b), comprise the Ces3-like mouse genes. However, not much is known about carboxylesterase 3 beta (Ces3b) of the mouse, apart from the fact that it is mainly expressed in mouse liver. Preliminary data from our laboratory, using microarrays in HNF-4α liver knock-out mice compared to control mice, revealed a significant number of genes belonging to the family of mouse carboxylesterases, which are up- or down-regulated in the absence of HNF-4α from mouse liver. Among the carboxylesterases down-regulated, Ces3b demonstrated the highest fold change in expression levels. So, the aim of the second chapter of the present thesis was to examine whether mouse Ces3b expression is HNF-4α dependent. To identify potential HNF-4α binding sites in mouse Ces3b promoter, we performed transient transfection assays in human embryonic kidney (HEK) cells, utilizing a series of luciferase reporter plasmids containing consecutive 5' deletions or a 3' deletion of the mouse Ces3b promoter along with an expression vector for the cDNA of HNF-4α. We observed that overexpression of HNF-4 significantly up-regulated mouse Ces3b promoter activity in all different Ces3b promoter fragments examined, apart from the shortest fragment bearing the Ces3b promoter region -80 to +4. Importantly, a potential HNF-4α binding site as well as a potential chicken ovalbumin upstream promoter (COUP) binding site, have been identified in the region downstream the +1 transcription start site. It is well known that HNF-4α and COUP transcription factors share common recognition sequence, so perhaps either of these sites could be responsible for the binding of HNF-4α in our transfection assays.

In summary, the findings of the present thesis increase our understanding of the mechanisms that regulate the expression of genes that are involved in lipid and lipoprotein metabolism in the liver and how these mechanisms are dys-regulated by silencing of hepatic nuclear factors or by natural polymorphisms in the promoters or enhancers of the corresponding genes.

CHAPTER 1:

Functional characterization of polymorphisms in the human apolipoprotein E (apoE) and Cholesteryl Ester Transfer Protein (CETP) promoters found in the Copenhagen City Heart Study (CCHS) to be associated with cardiovascular disease.

1. 1. INTRODUCTION

1.1.1. HDL biogenesis pathway

The biosynthesis of high density lipoprotein (HDL) begins in the liver and in the intestine. Liver and intestine produce and secrete lipid-free apolipoprotein A-I (apoA-I), which interacts thereafter extracellularly and physically with the membrane lipid transporter ATP-Binding Cassete sub-family A, member 1 (ABCA1) transporter. The latter interaction promotes the initial transfer of free cholesterol and phospholipids to apoA-I, forming in this way lipid-poor apoA-I particles (nascent HDL) (Zannis *et al.*, 2004). Nascent HDL acquires additional free cholesterol and phospholipids from extrahepatic, peripheral tissues, such as the macrophages located in the arterial wall through the interaction with ABCA1, and in this way lipid poor ApoA-I particles are converted to discoidal pre-β HDL (β, α4). HDL acquires phospholipids and potentially apolipoproteins (such as apoC-III), during the hydrolysis of triglyceride-rich lipoproteins. Phospholipid surface remnants are transferred from triglyceride-rich lipoproteins to HDL via the phospholipid transfer protein (PLTP) (Masson *et al.*, 2009). Through the action of the enzyme Lecithin: Cholesterol Acyl Transferase (LCAT), which is responsible for HDL cholesterol esterification, discoidal HDL is converted to mature spherical α-HDL (α1-α3). LCAT activity contributes to the formation of a hydrophobic core in HDL; consequently it is important for HDL particle maturation (Besler *et al*., 2012). Efflux of cellular cholesterol to mature spherical α-HDL can be also succeeded by the cell surface ATP-Binding Cassette sub-family G, member 1 (ABCG1) transporter in macrophages (Yancey *et al.*, 2000; Wang *et al.*, 2004) (Figure 1). Following synthesis, α-HDL can interact with the HDL receptor Scavenger Receptor class B type I (SR-BI), thus mediating not only the selective uptake of cholesterol esters and other lipids by the liver (for secretion afterwards into the bile through the ABCG5/G8 transporters), but also the bidirectional movement of unesterified cholesterol (Ji *et al.*, 1997; Gu *et al.*, 2000; Krieger, 2001; Liu *et al.*, 2002). Moreover, HDL esterified cholesterol can be transferred towards apoB-containing lipoproteins in exchange for triglycerides, a process mediated by the cholesteryl ester transfer protein (CETP), and finally be cleared by Low Density Lipoprotein receptor (LDLr)-mediated uptake of apoB-containing lipoproteins to the liver (Besler *et al.*, 2012). Hydrolysis of HDL lipids is mediated by various lipases (endothelial lipase, lipoprotein lipase, hepatic lipase), whereas exchange of lipids between HDL particles and apoB-containing lipoproteins (LDL, VLDL) is performed by PLTP and by cholesteryl ester transfer protein (CETP), regenerating in that way the pre-β HDL particles, that can initiate another cycle of HDL maturation and remodeling (von Eckardstein *et al*., 2005). Via a process called Trans Intestinal Cholesterol Efflux (TICE), cholesterol from HDL can be transported back to the intestine (van der Velde *et al.*, 2010). Lipid-free apoA-I may also interact with ABCA1 transporter present in peripheral tissues and cells, thus contributing to lipid efflux and to potential HDL biogenesis (Zannis *et al.*, 2008). HDL particles in plasma comprise a very heterogenous population in terms of shape, size and

protein/lipid content. ApoA-I constitutes the main protein component of all HDL particles, however, apoA-II, apoA-IV, apoC-III, apoE and apoM are found in certain HDL sub-populations (Zannis *et al.*, 2006). Heterogeneity is also observed in HDL lipids, including oxysterols, phosphor-sphingolipids and lysolipids. Finally, plasma HDL can be separated into discrete subpopulations (pre-β1, pre-β2, α4, α3, α2 and α1), by two dimensional gel electrophoresis, with α1 and α2 particles to be considered to have an association with protection against cardiovascular disease (Zannis *et al.*, 2006).

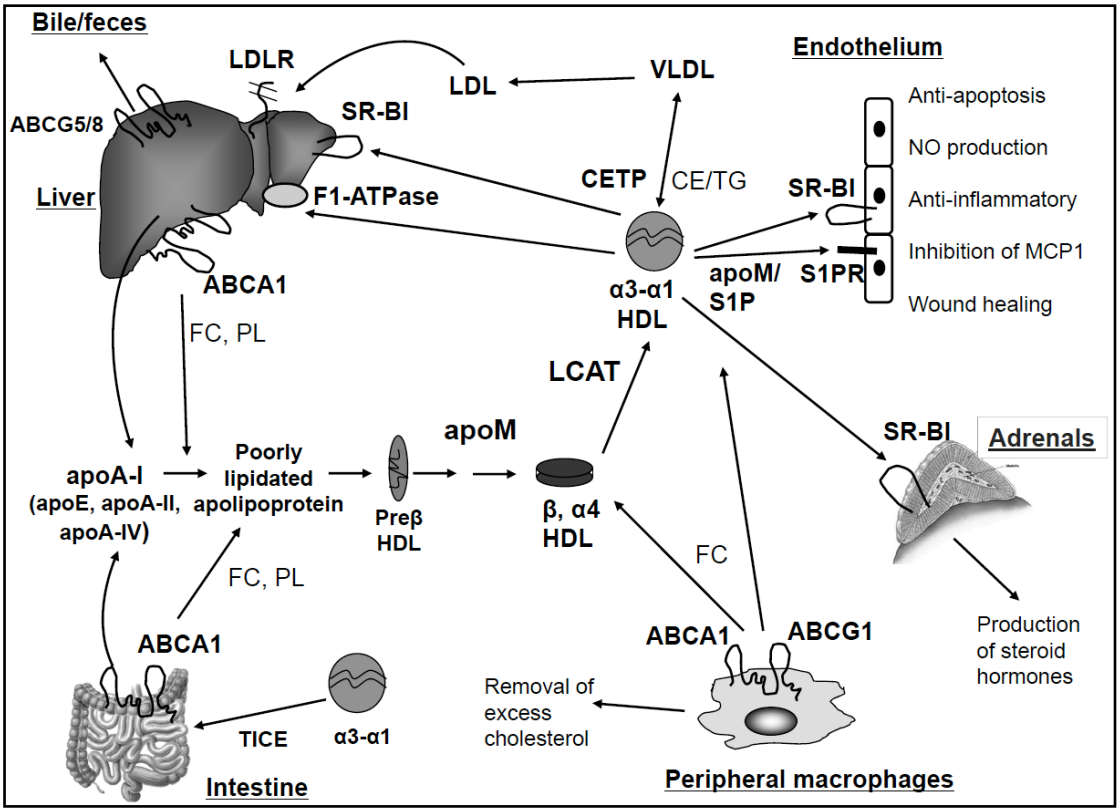

Figure 1. Schematic representation of the HDL metabolism pathway. The figure depicts: the different HDL particles subpopulations (poorly lipidated apoA-I, preβ, α1, α2, α3, α4); the proteins participating in the pathway of HDL biogenesis, catabolism and remodeling; organs, cell types and functions of HDL in each organ or cell type (Kardassis *et al.*, 2014).

1.1.2. Reverse Cholesterol Transport Pathway

The pathway of HDL metabolism has been termed Reverse Cholesterol Transport (RCT) pathway. This is attributed to the ability of HDL particles to transfer excess cholesterol from arterial wall macrophages to the liver for excretion to the feces, contributing to an anti-atherogenic effect of HDL (Kardassis *et al.*, 2014) (Figure 1). The first step of macrophage RCT is the hydrolysis of cytoplasmic cholesterol esters to free cholesterol, with the efflux of the latter to extracellular lipid-poor apoA-I or mature HDL (Besler *et al.*, 2012). Activity of transporters ABCA1, mediating the cholesterol efflux to lipid-poor apoA-I, and ABCG1, mediating the

cholesterol efflux to mature HDL (Tall *et al.*, 2008), facilitates the efficient removal of excess cholesterol from arterial wall macrophages. Following efflux from macrophages, LCAT esterifies HDL-associated cholesterol, by transferring a fatty acyl residue from phospholipids to the 3-beta-hydroxy group of cholesterol (Rader, 2009). Finally, HDL cholesterol is taken up by the liver for excretion in the bile, a process which can occur either directly via SR-BI or indirectly via CETP mediated transfer of cholesteryl esters to apoB-containing lipoproteins, which are thereafter taken up by the liver via the LDLr, as previously described (Besler *et al*, 2012). HDL is considered to be an anti-atherogenic lipoprotein, with low plasma HDL cholesterol levels to be associated with increased risk of myocardial infarction and coronary disease.

1.1.3. Human apolipoprotein E (apoE)

Human apolipoprotein E (apoE), is a glycoprotein of 35 kDa, which plays a crucial role in lipid metabolism, comprising a major component of various plasma lipoprotein classes, such as very low density lipoproteins, chylomicrons, and large high density lipoproteins (Mahley, 1988). The human apoE gene is 3.6 kB in length and is located at the 5'-end of the apoE/apoC-I/apoC-IV/apo-CII gene cluster on chromosome 19 (Olaisen *et. al.*, 1982). It contains 299 amino acid residues and has independently folded amino-terminal as well as carboxy-terminal domains, which are joined together by a flexible hinge region (Hatters *et al.*, 2006; Zhong *et al.*, 2009). The region responsible for the interaction with apoE receptors (residues 136-150) is located in the N-terminal domain, whereas the major lipid-binding region (residues 244-272) is located in the C-terminal domain. There are three isoforms of apoE, apoE2, apoE3 and apoE4, which differ at amino acid positions 112 and 158 (Weisgraber *et al.*, 1981; Mahley *et al.*, 1988; Hatters *et al*., 2006). These three apoE isoforms exhibit altered protein structure owning to the aforementioned amino acid differences, and such alteration influences its receptor binding and lipid association. ApoE3 and apoE4, for example, bind to LDLr with almost 50-fold greater affinity compared to apoE2, resulting in less efficient lipid transport by apoE2, so its presence has been found to be associated with type III hyperlipoproteinaemia. ApoE4 is associated with a modestly increased risk for cardiovascular disease and binds preferentially to large lipoprotein particles, a fact that is attributed to the presence of Arg at position 112, which affects the conformation of the side chain of Arg61, resulting in "domain interaction" between Glu255 in the C-terminal domain and Arg61 located in the N-terminal domain (Bu, 2009) (Figure 2).

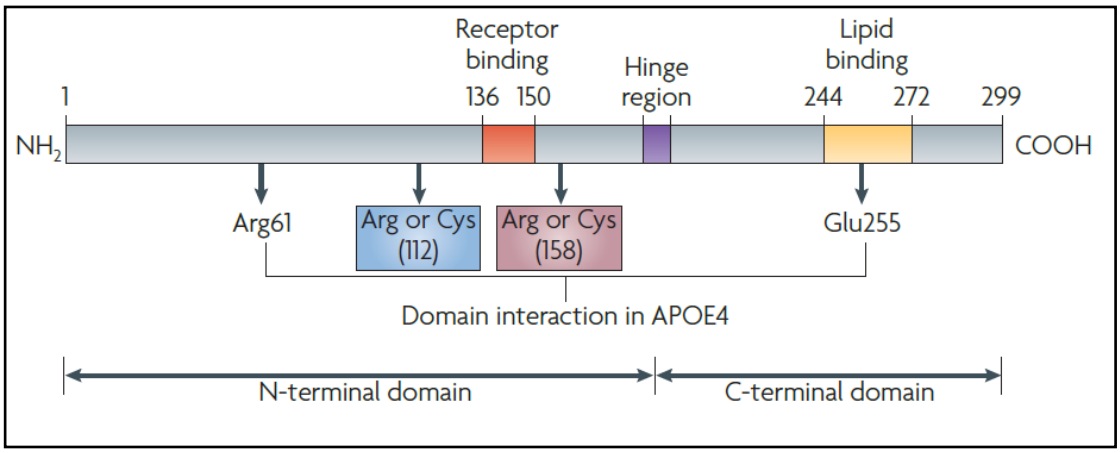

Figure 2. Schematic representation of human apoE. Human apoE is comprised of 299-amino acids and contains two independently folded domains: an amino-terminal domain, which includes the receptor-binding region as well as a carboxy-terminal domain, which includes the major lipid-binding region. Residues 112 and 158, which distinguish the different apoE isoforms, are shown. ApoE2 has Cys at both positions, ApoE3 has Cys at position 112 and Arg at position 158, whereas apoE4 has Arg at both positions. The domain interaction between Arg611 and Glu255 in apoE4 isoform is shown (Bu, 2009).

ApoE is mainly synthesized by the liver, kidney, spleen, brain, lung, skin, as well as various cells such as macrophages (Zannis *et al.*, 2001). ApoE appears to possess a dual function: apart from its role in the clearance of triacylglycerol-rich lipoproteins, it participates in the HDL-sized apoE-containing particles biogenesis (Kypreos *et al.*, 2007). As a ligand for the low density lipopoprotein receptor family, apoE has the ability to scavenge lipoprotein remnants and to mediate the clearance of apoE-rich large high density lipoprotein particles from the serum, a process that is significant for the redistribution of triglycerides and cholesterol between the liver and the peripheral tissues (Gafencu *et al.*, 2007). ApoE enhances the lipid exchange between lipoproteins, mediated by CETP, thus alterations in CETP activity contribute to changes in HDL composition and size (Korhonen *et al.*, 1999; Ordovas *et al.*, 2000). Since both apoE and CETP contribute to the reverse cholesterol transport, the effects of variations at the CETP locus may very well interact with the apoE genotype (Sorlí *et al.*, 2006). The key role of apoE in the clearance of triacylglycerol-rich lipoproteins and in the biogenesis of apoE-containing HDL particles were demonstrated using adenovirus-mediated gene transfer of apoE in apoA-I-deficient mice or in ABCA1-deficient mice. Specifically, infection of apoA-$I^{-/-}$ mice with 2 x 10^9 pfu (plaque forming unit) of an apoE4-expressing adenovirus significantly increased both the triglyceride-rich VLDL/IDL/LDL and HDL fraction, generating discoidal HDL particles. However, $ABCA1^{-/-}$ mice treated the same way, failed to form HDL particles, which suggests that ABCA1 is essential for the biogenesis of apoE-containing HDL. Moreover, combined infection of apoA-$I^{-/-}$ mice with a mixture of adenoviruses expressing both human LCAT (5 x 10^8 pfu) and apoE4 (2 x 10^9 pfu), resulted in clearance of the triglyceride-rich lipoproteins, increase of HDL and

conversion of discoidal form of HDL to spherical. However, similar treatment of ABCA1$^{-/-}$ mice, failed to promote formation of HDL, which suggests that LCAT is necessary for apoE-containing HDL maturation (Zannis *et al.*, 2008). These data proposed the following model: in the first step of the HDL biosynthesis pathway, either lipid-free or minimally lipidated apoE secreted by the liver or other tissues, functionally interacts with the ABCA1 lipid transporter and in this way acquires cholesterol and phospholipids. Intermediate steps follow, which are not completely understood, ending up to the conversion of lipidated apoE into discoidal particles, which accumulate under conditions of apoE overexpression and are able to be visualized through electron microscopy. With the activity of LCAT, the discoidal HDL particles are converted into spherical particles, leading to the biogenesis of apoE-containing HDL (Kypreos *et al.*, 2007) (Figure 3). It has been demonstrated that the formation of spherical apoE-containing HDL was promoted by substitution of hydrophobic residues in the 261 to 269 region of apoE by Ala, and more specifically by substitutions of the residues Leu261, Trp264 and Phe265 (Zannis *et al.*, 2008). The apoE-containing HDL may have anti-inflammatory and anti-oxidant functions that resemble those known for apoA-I-containing HDL, a feature that contributes to the atheroprotective properties of apoE and in addition it may serve several biological functions in the brain (Li *et al.*, 2003).

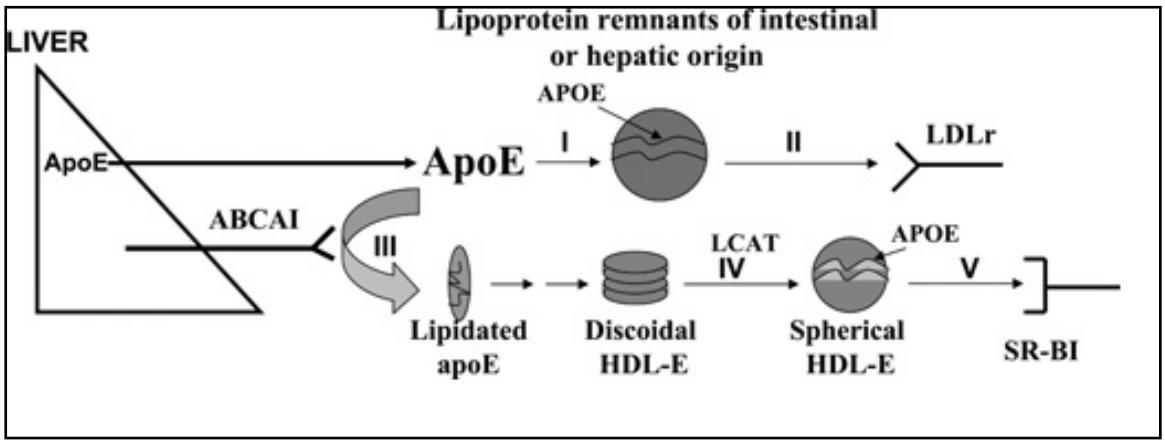

Figure 3. Schematic representation of the participation of apoE in the biogenesis of apoE-containing HDL particles (III-V) and in the clearance of triacylglycerol-rich lipoproteins (I and II). LDLr: Low Density Lipoprotein receptor, ABCA1: ATP-Binding Cassette superfamily A member 1, SR-BI: Scavenger Receptor class B, type I, HDL-E: High Density Lipoprotein-containing apoE.

Several studies have demonstrated, that apoE forms lacking the C-terminal domain, can efficiently clear cholesterol from the plasma of apoE-deficient mice, whereas full-length apoE induced hypertriglyceridaemia, increased VLDL secretion (Kypreos *et al.*, 2001; Zannis *et al.*, 2004) as well as affected the structure of HDL (Kypreos *et al.*, 2005).

Transcriptional regulation of apoE has been shown to require the aggregation of several transcription factors and their interaction both with the proximal promoter and with the distal regulatory regions. The proximal apoE promoter is well

conserved between mice and humans with the same localization of GC box and TATA box elements (Rajavashisth *et al.*, 1985; Horiuchi *et al.*, 1989). Many positive and negative elements that can modulate apoE gene expression have been detected on its promoter region, such as URE1, URE3, Sp1, AP-2, Zic1/Zic2, LXRα/RXRα and LXRβ/RXRα (Kardassis *et al.* 2015). In addition, Mitogen-activated protein kinase/ERK kinase Kinase 1 (MEKK1) and tumor progression locus 2 (Tpl2) were demonstrated to be negative regulators of apoE promoter activity by LPS (Gafencu *et al.*, 2007). Regarding the distal regulatory regions, two homologous enhancers designated as multienhancer 1 (ME.1) and multienhancer 2 (ME.2) have been found to control the expression of apoE in macrophages (Shih *et al.*, 2000). A STAT1

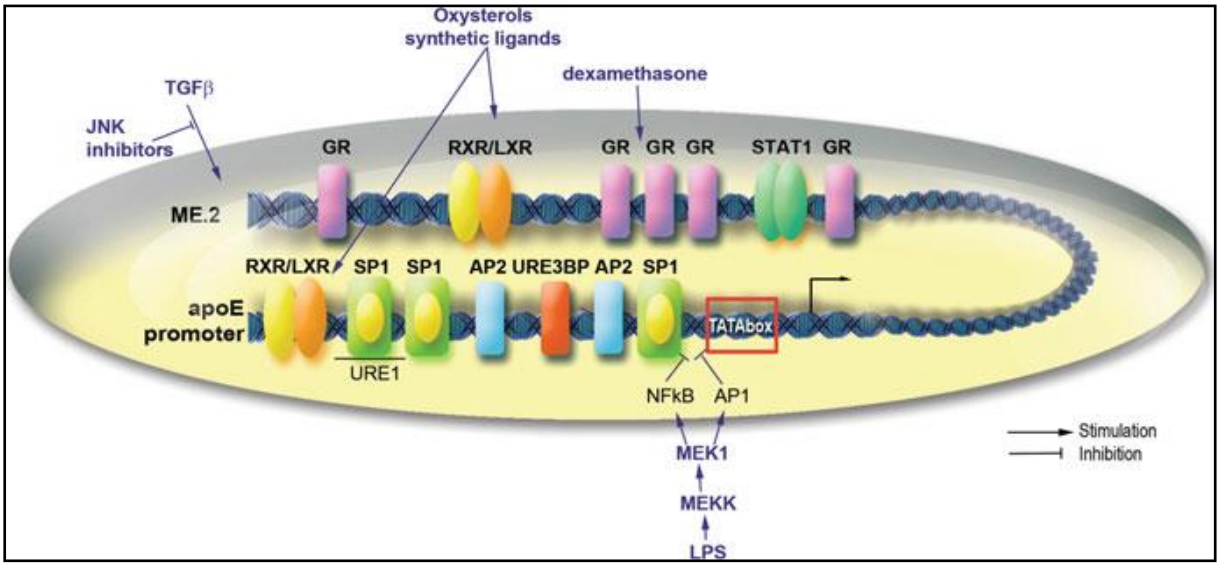

Figure 4. Regulatory regions, signaling molecules (metabolites, cytokines, drugs, growth factors) and transcription factors modulating the apoE gene expression. GR: Glucocorticoid Receptor, STAT: Signal Transducer and Activator of Transcription, RXR: Retinoid X Receptor, LXR: Liver X Receptor, JNK: Jun N-terminal Kinase, AP-2: Activator Protein 2, LPS: Lipopolysaccharide, Sp1: Specificity Protein 1, NF-κB: Nuclear Factor kappa Beta, URE3BP: Upstream Regulatory region 3 Binding Protein, ME.2: Multienhancer 2, TGFβ: Transforming Growth Factor β (Kardassis *et al.* 2015).

binding site has been recognized in ME.2 region but not in ME.1, and it was shown to up-regulate apoE gene expression in macrophages (Figure 4) (Kardassis *et al.* 2015).

So, apoE appears to be an important protein, both for the biogenesis of apoE-containing HDL partricles (Gafencu *et al.*, 2007; Kypreos *et al.*, 2007) and for the enhancement of lipid exchange between lipoproteins, a process mediated by CETP (Korhonen *et al.*, 1999; Ordovas *et al.*, 2000). Since apoE contributes to the reverse cholesterol transport, the effects of variations at the apoE locus could associate with alterations in the HDL biogenesis pathway, thus examination of genetic polymorphisms is critical and could increase our knowledge of the possible mechanisms involved in this process.

1.1.4. Human Cholesteryl Ester Transfer Protein (CETP)

Cholesteryl Ester Transfer Protein (CETP) gene is located in the long arm of chromosome 16 in humans near the LCAT gene locus (Lusis *et al.*, 1987), at position

16q21. The gene consists of 16 exons and encompasses about 25 kB of genomic DNA (Agellon *et al.*, 1990). Although the CETP cDNA sequence has no general homology to the apolipoprotein cDNAs, a pentapeptide (Val-Leu-Thr-Leu-Ala) is conserved among the signal sequences of human CETP, apoA-I, apoA-IV and lipoprotein lipase (Agellon *et al.*, 1990). CETP specifies a 66 to 74 kDa hydrophobic glycoprotein, which is predominantly expressed in the liver, spleen, adipose tissue and skeletal muscle (Brown *et al.*, 2002), and at lower levels in kidney, small intestine, adrenal and heart (Jiang *et al.*, 1991). Different studies revealed that HepG2 cells, macrophages, CaCo2 cells, adipocytes and neuroblastoma cells secrete active CETP (Drayna *et al.*, 1987; Faust *et al.*, 1987; Swenson *et al.*, 1987; Faust *et al.*, 1988; Faust *et al.*, 1989; Faust *et al.*, 1990; Jiang *et al.*, 1991; Pape *et al.*, 1991).

CETP is primarily localized on larger apoA-I-containing HDL particles, and as mentioned before, its role is to catalyze the exchange of triglycerides from apoB-containing particles, such as LDL and VLDL for cholesteryl esters from the core of HDL particles (Anderson *et al.*, 2003). Its activity determines the plasma levels of HDL cholesterol (von Eckardstein *et al*., 2005; Tall *et al.*, 2008), and it is responsible for decreasing the concentration of protective HDL-cholesterol (Tall, 1993). Although deleterious mutations in the CETP gene are quite rare, polymorphisms that are observed frequently on the CETP gene can significantly influence levels of lipoproteins in the plasma, so it is considered to be candidate for coronary heart disease (Corbex *et al.*, 2000). Given the key importance of dyslipidemia for the atherosclerosis development (Rader *et al.*, 2003), many human studies reported a positive association between the atherosclerotic cardiovascular disease mortality and morbidity and plasma CETP activity (Barter *et al.*, 2003; Boekholdt *et al.*, 2004; Zeller *et al.*, 2007; Masson *et al.*, 2009; Ridker *et al.*, 2009). Especially when it comes to the metabolic syndrome, including hypertriglyceridemia and type 2 diabetes mellitus, the predictive value of increased CETP activity and mass for cardiovascular disease events is extremely high (de Vries *et al.*, 2005).

There are several studies on the transcription regulation of CETP by regulatory elements. Two independent groups in 1999 and 2000, demonstrated that CETP gene is regulated at a transcriptional level by both LXRs and SREBPs which bind on the gene's promoter and regulate its expression in response to intracellular cholesterol levels (Gauthier *et al.*, 1999; Luo *et al.*, 2000). In 2009 Lakomy *et al.* reported that LXR agonist T0901317 increased significantly human CETP mRNA levels and CETP production in differentiated macrophages, but not in peripheral blood monocytes. This LXR-mediated CETP up-regulation was abrogated in inflammatory human and mouse macrophages, a condition which was independent of the lipid load. It was demonstrated that LXR-mediated induction of human CETP expression is activated during the differentiation of monocytes to macrophages; however it is inhibited in inflammatory macrophages (Lakomy *et al.*, 2009). Moreover, in 2010

Honzumi *et al.* demonstrated that synthetic LXR agonists enhanced the activity of plasma CETP and decreased HDL cholesterol levels in human CETP transgenic mice and cynomolgus monkeys (Honzumi *et al.*, 2010). Moreover, several other factors were shown to contribute to the regulation of CETP promoter activity, such as the LRH-1, which was shown to enhance the sterol-mediated induction of the CETP gene by LXRs (Luo *et al.*, 2001), the Yin Yang 1 (YY1), which shares the same binding element with SREBPs (Gauthier *et al.*, 1999), the CCAAT/enhancer-binding protein (C/EBP), which activates the expression of CETP gene (Aggelon *et al.*, 1992), the orphan nuclear receptor ARP-1, which hinders the activity of CETP promoter (Gaudet *et al.*, 1995) and the retinoic acid receptor (RAR), which was shown to regulate the expression of the CETP gene in response to all-*trans*-retinoic acid (Jeoung *et al.*, 1999). In 2003, Le Goff *et al.* identified binding sites for both Sp1 and Sp3 transcription factors on the CETP promoter's positions -37, -623 and -690 and demonstrated the importance of these factors for the basal activity of CETP promoter (Le Goff *et al.*, 2003a). In 2013, Gautier *et al.* demonstrated that the CETP gene is regulated by bile acids and their nuclear receptor, farnesoid X receptor (FXR). More specifically, they identified an ER8 FXR response element (FXRE) in the first intron of the CETP gene, which could be responsible for the up-regulation in CETP gene expression in response to bile acids (Gautier *et al.*, 2013). Another study in 2008 demonstrated that DR4 LXRE in the proximal CETP promoter comprise a binding position for FXRα, and upon binding, LXR-mediated transactivation of the CETP promoter is repressed, through a competition mechanism (Park *et al.*, 2008).

So as mentioned above, since CETP activity determines the plasma levels of HDL cholesterol (von Eckardstein *et al*., 2005; Tall *et al.*, 2008), and is responsible for decreasing the concentration of protective HDL-cholesterol (Tall, 1993), it is crucial to examine whether polymorphisms observed on the CETP gene could impact on the its expression levels, thus determing the levels of lipoproteins in the plasma.

1.1.5. Copenhagen City Heart Study (CCHS)

The Copenhagen City Heart Study (CCHS) was initiated in 1976 by Peter Schnohr, Gorm Jensen and AnneTybjærg Hansen. The original purpose of the study was to increase knowledge about the prevention of Coronary Heart Disease (CHD) and stroke. Over the years, questions about heart failure, arthrosis, pulmonary diseases, allergy, dementia, epilepsy, stress, "vital exhaustion", sleep-apnea, and genetics, were added. Regarding the population sample included in this study, this comprises a random sample of 19.698 individuals, 9.381 men and 10.317 women 20-93 years old, drawn from the population of Copenhagen on January 1, 1976. The oversampling of women was desirable, so as to address why women have CHD-events several years older, on average, than men do. The 1976-1978 initial examination had a response rate of 74 %. The second examination took place in 1981-1983 with a response of 70.1 %, the third examination was in 1991-1994 with a

response of 61.2 %, whereas the fourth examination was in 2001-2003 with a response of 49.5 % (Schnohr *et al.*, 2001).

1.1.5.1. Polymorphisms on the promoter of human Cholesteryl Ester Transfer Protein (CETP) found in the CCHS

Seven single nucleotide polymorphisms (SNPs) have been identified in the CCHS, in the promoter region of the CETP gene. Specifically, those SNPs are (numbering refers to the first codon of the CETP translated region): -1073T>C (rs12708968), -998G>A (rs4783961), -854C>T (rs4783962), -658C>A (rs1800776), -656C>A (rs1800775), -282C>A (rs not available) and -65G>A (rs17231520). The exact DNA sequence of the CETP promoter region, as well as the SNPs, is available in the Appendix.

1.1.5.1.1. *-1073T>C substitution*

The SNP -1073T>C (rs12708968) of the CCHS comprises a substitution of the most frequent allele T with the less frequent allele C at position -1046 upstream of +1 in the CETP gene. It was demonstrated that the genotype frequency of -1046T>C, -971G>A, -827C>T and -631C>A were in Hardy-Weinberg equilibrium both in parents and in offspring in a HERITAGE Family Study (Spielmann *et al.*, 2007). In 2003, Le Goff *et al.* demonstrated in transient transfection experiments in HepG2 cells, that there is a binding site for the orphan nuclear receptor CYP7A promoter binding factor (CPF), between positions -1042 to -1050, which facilitates the activation of human CETP promoter activity. The introduction of two point mutations, at positions -1045 and -1046 (T>A substitutions in both positions), which are known to abolish the binding of CPF to the corresponding DNA sequence, indeed abrogated the specificity of binding (Le Goff *et al.*, 2003b). So it was shown, that under the presence of the most common allele T at position -1046, CPF is able to bind and mediate the activation of the CETP promoter.

1.1.5.1.2. *-998G>A substitution*

The SNP -998G>A (rs4783961) of the CCHS comprises a substitution of the most common allele G with the less common allele A, mapped at position -971 upstream of +1 in the CETP gene. In 2005 Frisdal *et al.* performed transient transfection experiments using a set of vectors corresponding to different potential haplotypes for CETP promoter polymorphisms -1337C>T, -971G>A and -629C>A, in order to observe potential changes in CETP activity, in response to a specific haplotype. Indeed, it was demonstrated that the haplotype carrying the less frequent alleles at positions -1337, -971 and -629 (alleles T/A/A respectively) accounted for only 20% of promoter activity ($P<0.0001$) comparing with the haplotype carrying the more frequent alleles (alleles C/G/C respectively) for the three polymorphisms under study. Moreover it was shown that haplotypes carrying only one mutated site (T/G/C, C/A/C and C/G/A) or two mutated sites (T/G/A, C/A/A, T/A/C) had significantly lower promoter activity than the haplotype carrying the

most common alleles (C/G/C). It was shown that the less common allele A, for polymorphism -971G>A, displayed a significant lower *in vitro* promoter activity compared to the more frequent allele, indicating that the polymorphism is functional. It was concluded that the -971A allele enhances the inhibitory action of the -1337T allele on CETP gene promoter activity, whereas the -971G allele is able to reduce the effect of -1337T allele through interaction with the -629A allele (Frisdal *et al.*, 2005). Moreover, in a HERITAGE Family study in 2007, Spielmann *et al.* demonstrated that there was significant interaction between the -971G>A and the -629C>A variants on baseline HDL levels. The strong association between the -629C>A locus and plasma HDL was particularly evident among the -971A allele carriers but no the -971 G/G homozygotes (Spielmann *et al.*, 2007). However, the study by Spielmann *et al.* in 2007, comes into a controversy with a study 5 years earlier by Le Goff *et al.*, who demonstrated in transient transfection experiments in HepG2 cells, that -971G>A polymorphism did not modulate transcriptional activity of the human CETP gene promoter, therefore constituting a non-functional polymorphism. In the same work, it was revealed that the -971G>A polymorphism (A allele frequency: 0.491) was significantly associated with both CETP concentration (P=0.009) and plasma HDL cholesterol (P=0.006) in the Etude Cas-Témoins de l'Infarctus du Myocarde (ECTIM) cohort study, which is a case/control study of myocardial infarction. Specifically, subjects with genotype -971GG displayed both high plasma CETP concentration and low HDL cholesterol levels, while genotype -971AA subjects displayed the inverse relationship. It was shown that -971G>A polymorphism interacts significantly with the functional polymorphism -629C>A and the TaqIB polymorphism with respect to plasma HDL cholesterol levels, however it does not affect plasma CETP concentration. It was concluded that the observed effects of the -971G>A polymorphism on both HDL cholesterol levels and plasma CETP concentration were due to functional variants which are in linkage disequilibrium with it (Le Goff *et al.*, 2002).

1.1.5.1.3. *-854C>T substitution*

The SNP -854C>T (rs4783962) of the CCHS comprises a substitution of the most common allele C with the less common allele T, mapped at position -827 upstream of +1 in the CETP gene. Frisdal *et al.* in 2005 demonstrated in a REGRESS cohort study that allele frequencies of polymorphism -827C>T are in agreement with those observed 3 years earlier by Lu *et al.*, in a population of elderly Japanese men (Lu *et al.*, 2002; Frisdal *et al.*, 2005). In a HERITAGE Family Study in 2007, it was referred that genotype frequency of -827C>T is in Hardy-Weinberg equilibrium both in parents and in offspring (Spielmann *et al.*, 2007).

1.1.5.1.4. *-658C>A substitution*

The SNP -658C>A (rs1800776) of the CCHS comprises a substitution of the most common allele C with the less common allele A, mapped at position -631

upstream of +1 in the CETP gene. In 2000 Dachet *et al.* identified for the first time that position -631 relative to the transcription start site comprises a polymorphism position in the population of the ECTIM study (Dachet *et al.*, 2000). Genotype frequencies of -631 SNP were found to be in Hadry-Weinberg equilibrium with the frequencies of SNP -629C>A, however allele frequencies of the CETP -631 polymorphism increased from Belfast to Toulouse, which are populations of the ECTIM study. Moreover, polymorphism -631 was almost completely concordant with the TaqIB polymorphism on intron 1, whereas it was in complete negative linkage disequilibrium with the CETP -629 polymorphism in the ECTIM study (Dachet *et al.*, 2000). No significant correlation was observed between -631 CETP polymorphism and either plasma HDL cholesterol concentration or CETP mass (Dachet *et al.*, 2000).

1.1.5.1.5. *-656C>A substitution*

The SNP -656C>A (rs1800775) of the CCHS comprises a substitution of the most common allele C with the less common allele A, mapped at position -629 upstream of +1 in the CETP gene. In 2000 Dachet *et al.* demonstrated for the first time that position -629 relative to the transcription start site comprises a polymorphism position in the population of the ECTIM study (Dachet *et al.*, 2000). Genotype frequencies of -629 SNP were found to be in Hadry-Weinberg equilibrium with the frequencies of SNP -631C>A, however allele frequencies on the -629 CETP polymorphism were homogeneous across the populations of the ECTIM study (Dachet *et al.*, 2000). Moreover, polymorphism -629 was almost completely concordant with the TaqIB polymorphism in intron 1, whereas it was in complete negative linkage disequilibrium with the CETP -631 polymorphism in the ECTIM study (Dachet *et al.*, 2000). The -629 polymorphism was significantly associated with plasma concentrations of HDL cholesterol and plasma concentrations of CETP mass. Subjects homozygous for the -629C allele showed higher CETP mass and lower HDL cholesterol than subjects homozygous for the -629A allele, whereas heterozygous subjects displayed intermediate values (Tai *et al.*, 2003; Akbarzadeh *et al.* 2012). The same observations were demonstrated in a study including 220 non-diabetic men without cardiovascular risk, where plasma CETP mass and cholesterol ester transfer (CET) levels were higher, whereas HDL cholesterol and plasma apoA-I levels were lower in CC compared to AA carriers (Dullaart *et al.*, 2008). Consistent with these observations, it was demonstrated in transient transfection experiments in HepG2 cells, that a plasmid construct containing the A allele at position -629 of the CETP promoter showed significantly lower luciferase activity than the plasmid construct containing the C allele (Dachet *et al.*, 2000). It was shown that -629 polymorphism, along with -631, are located in a potential PuF/nm23-H2 nucleoside diphosphate kinase binding site (CCCACCC from -625 to -631), which was recognized by computer consensus motif analysis (Corbex *et al.*, 2000). The promoter sequence surrounding the -629 polymorphism displays similarities with the inverted form of a non-

consensus Sp1 binding site (Hoppe *et al.*, 1998). Dachet *et al.* in 2000 demonstrated that transcription factors Sp1 and Sp3 bind to -629 Sp1 binding site and this binding accounts for the differences detected in promoter activity between alleles C and A, with Sp1 and Sp3 binding to the A but not to the C allele, thus repressing the promoter activity (Dachet *et al.*, 2000). In 2003 Le Goff *et al.* performing the same experiments in HepG2 cells came to the same conclusion about binding of Sp1 and Sp3 to allele A, thus leading to significant reduction in transcriptional activity (Le Goff *et al.*, 2003a). Le Goff *et al.* in 2002 demonstrated that -629C>A polymorphism interacts with the -971G>A (Le Goff *et al.*, 2002). Frisdal *et al.* in 2005 reported that -629C>A is a functional polymorphism in a REGRESS cohort study (Frisdal *et al.*, 2005). In 2007 Spielmann *et al.* demonstrated that -629C>A variant significantly interacts with -971G>A variant as regards baseline HDL levels. Specifically, the strong association between the -629C>A locus and plasma HDL cholesterol was particularly evident among the -971A allele carriers but not the -971 G/G homozygotes in a HERITAGE study (Spielmann *et al.*, 2007). In a study performed in 2008 using subjects recruited from the PREVEND study carried out in the city of Groningen, it was revealed that in CC homozygotes, cholesterol efflux, plasma CETP mass and LCAT activity were higher, while the HDL cholesterol was lower, compared to AA homozygotes (Borggreve *et al.*, 2008). In a case-control study, performed in 2012, including 400 Iranian individuals, it was demonstrated that there was higher allele A frequency among the individuals with metabolic syndrome, compared to the control group, which was associated with high total cholesterol, LDL cholesterol and low CETP levels, a condition related to metabolic syndrome. It was observed that CC genotype had the highest level of CETP mass, whereas AA genotype was related to the lowest concentration, while heterozygotes had intermediate levels of CETP concentration (Akbarzadeh *et al.*, 2012).

1.1.5.1.6. *-282C>A substitution*

The SNP -282C>A (rs not available) of the CCHS comprises a substitution of the most common allele C with the less common allele A, mapped at position -255 upstream of +1 in the CETP gene. However, this specific substitution was not referred to in any of the references provided.

1.1.5.1.7. *-65G>A substitution*

The SNP -65G>A (rs17231520) of the CCHS comprises a substitution of the most common allele G with the less common allele A, mapped at position -38 upstream of +1 in the CETP gene. While the -629 polymorphism is common in all ethnic groups, the -38 polymorphism is found at significant levels (6.4%) only among African Americans (Thompson *et al.*, 2004) and very few individuals were found to have the specific variant (Thompson *et al.*, 2003). In 1995 Gaudet *et al.* identified the existence of an Sp1 binding site at region -26 to -57 upstream of the transcription start site of the CETP gene. This site spans the sequence 5'-

TGTTCCGTGGGGGCTGGGC**G**GACATACATATA-3', in which the binding motif recognized by Sp1 is underlined and the -38G>A variant is in bold. Through site directed mutagenesis of the binding motif 5'-GGGCGG-3' to 5'-GTTCGG-3', they demonstrated that the binding of Sp1 was abolished. Moreover it was shown in transient transfection experiments in HepG2 and Caco-2 cells, that the presence of the two consecutive mutations in the Sp1 binding site significantly reduced the promoter's activity (Gaudet *et al.*, 1995). In 2003 Le Goff *et al.* investigating the same mutations described in the previous study, demonstrated in transient co-transfection experiments in SL2 insect cells, that Sp1 and Sp3 up-regulate the CETP promoter activity (Le Goff *et al.*, 2003a). This is in agreement with the previous study, so it was concluded that both Sp1 and Sp3 are required for activation and positive regulation of CETP gene expression (Gaudet *et al.*, 1995; Liang *et al.*, 1995; Biggs *et al.*, 1996; Prowse *et al.*, 1997). In 2004, it was demonstrated in African Americans individuals from the ACCESS trial, that those homozygous for the less common -38A allele had higher HDL cholesterol levels than those bearing the more frequent -38G allele. It was found that the -38A allele is in high linkage disequilibrium with the -629A allele (Thompson *et al.*, 2004).

1.1.5.2. Polymorphisms on Exon 1 of human Cholesteryl Ester Transfer Protein (CETP) found in the CCHS

1.5.2.1. *-9C>T substitution*

One SNP has been identified downstream of +1 spanning the first exon of the CETP gene in the CCHS. This is the -9C>T (rs not available). It comprises a substitution of the most common allele C with the less common allele T, mapped at position +19. The exact DNA sequence of the region spanning this SNP is available in the Appendix.

1.1.5.3. Polymorphisms in Intron 1 of human Cholesteryl Ester Transfer Protein (CETP) found in the CCHS

1.1.5.3.1. *+279G>A (TaqIB) substitution*

The SNP TaqIB (+279G>A) (rs708272) of the CCHS comprises a substitution of the most common allele G with the less common allele A, mapped at position +424 downstream of +1 in the CETP gene. The G to A substitution leading to the loss of a site for TaqI restriction enzyme when the A allele is present. The TaqIB polymorphism comprises the most extensively studied variant in the human CETP gene, accounting for 6% of the variance in plasma CETP levels (Klerkx *et al.*, 2003). Allele G is designated as B1 allele whereas allele A is designated as B2 allele. Subjects displaying the B1 allele demonstrate higher levels of CETP and lower levels of HDL cholesterol (Tai *et al.*, 2003; Ayyobi *et al.*, 2005; Yilmaz *et al.*, 2005), when compared either with the homozygous or with the heterozygous subjects for the B2 allele (Corbex *et al.*, 2000; Ordovas *et al.*, 2000). In 2007 Porchay-Baldèrelli *et.al.*

demonstrated that B1 allele is associated with higher prevalence of coronary heart disease (Porchay-*Baldèrelli et.al.* 2007), whereas in 2012 and 2013 two separate studies revealed that B2 allele was significantly associated with both low plasma CETP activity and high HDL cholesterol levels (Akbarzadeh *et al.*, 2012; Lu *et al.*, 2013). It was shown in men with coronary artery disease and HDL deficiency, that the CETP TaqI B2B2 genotype is significantly reduced and associated with higher levels of plasma HDL cholesterol levels and lower coronary heart disease risk (Brousseau *et al.*, 2002). Association with coronary heart disease was demonstrated in other studies as well (Ordovas *et al.*, 2000; Brousseau *et al.*, 2002). In 1998 Kuivenhoven *et al.* demonstrated the association of TaqIB with the progression of coronary atherosclerosis by computer-assisted quantitative angiography in the REGRESS study (Kuivenhoven *et al.*, 1998). In a population of Chinese men it was shown that plasma HDL cholesterol levels was in the order B1B1<B1B2<B2B2 (Lu *et al.*, 2013). So the effects of TaqIB on plasma HDL cholesterol and CETP mass are gender-dependent and have been reported to be influenced by body mass index, alcohol consumption and insulin levels (Freeman *et al.*, 1994; Hannuksela *et al.*, 1994). Owning to its location in the first intron, it was proposed that the TaqIB polymorphism does not represent a functional regulatory site but it is a marker for other functional sites (Frisdal *el al.*, 2005). Moreover, its effects on HDL cholesterol and plasma CETP levels are independent (Fumeron *et al.*, 1995; Bernard *et al.*, 1998), which suggests that the TaqIB variant is a marker for at least two functional polymorphisms (Frisdal *el al.*, 2005). Indeed, several studies identified that a polymorphism at position -629 in the upstream region of the CETP gene promoter, can modulate the transcriptional activity of the CETP gene in *in vitro* experiments (Dachet *et al.*, 2000). Le Goff *et al.* in 2002 demonstrated that TaqIB polymorphism interacts with the -971G>A (Le Goff *et al.*, 2002). Different studies have shown that TaqIB polymorphism appears to be in linkage disequilibrium with the promoter polymorphism -629C>A in non-diabetic European Caucasian males with cardiovascular disease (Corbex *et al.*, 2000; Dachet *et al.*, 2000). In 2007, a study performed in individuals from the Tehran Lipid and Glucose Study (TLGS) revealed again that alleles B1 and B2 of TaqIB polymorphism were in linkage disequilibrium with alleles A and C of -629 polymorphism in the CETP gene, with a significant association between B2 and A alleles with both high HDL cholesterol levels and low CETP activity (Daneshpour *et al.*, 2007).

1.1.5.4. Polymorphisms on the promoter of human apolipoprotein E found in the CCHS

Three SNPs have been identified in the promoter region upstream of +1 of the human apoE gene in individuals participating in the CCHS. Those are: -491A>T (rs449647), -427T>C (rs769446) and -219G>T (rs405509). The exact DNA sequence of the apoE promoter region, as well as the SNPs, is available in the Appendix.

1.1.5.4.1. *-491A>T substitution*

The SNP -491A>T (rs449647) of the CCHS comprises a substitution of the most frequent allele A with the less frequent allele T at position -475 upstream of +1 in the apoE gene. In 1998 Artiga *et al.*, identified the existence of -491A>T variant. Transient transfection experiments were performed in constructs bearing the T allele and constructs bearing the A allele and it was demonstrated that allele T contributes to a significant decrease in apoE promoter activity (Artiga *et al.*, 1998). It has been reported that apoE promoter -491A genotype is associated with a higher plasma level of apoE and an increased risk for Alzheimer's disease compared to its -491T counterpart (Laws *et al.*, 1999; Scacchi *et al.*, 2001). Another study in 2011, in which dual-luciferase reporter assays were performed in WRL-68 (human hepatic embryonic) and U-87 (human astrocytic) cell lines, demonstrated that allele T significantly decreased the promoter's activity in both cell lines tested, implying that -491A>T comprises a functional polymorphism (Geng *et al.*, 2011).

1.1.5.4.2. *-427T>C substitution*

The SNP -427T>C (rs769446) of the CCHS comprises a substitution of the most frequent allele T with the less frequent allele C at position -411 upstream of +1 in the apoE gene. In 1998 Artiga *et al.*, identified the existence of -427T>C variant. Transient transfection experiments were performed in constructs bearing the C allele and constructs bearing the T allele, in order to evaluate the promoter's acitivity, however no significant difference was observed between the two alleles (Artiga *et al.*, 1998). In a study performed in a sample of male coronary heart disease patients and controls, allele -427C showed a significant excess in the patients group, with this observation being present only in case where the apoE4 allele was absent, indicating that the presence of -427C allele could represent a risk for developing coronary heart disease in subjects with E2/E2, E2/E3 and E3/E3 genotypes (Corbo *et al.*, 2001). Allele -427C was in linkage disequilibrium with apoE ε2 allele in a different group of patients studied, suffering from atherothrombotic stroke or ischemic heart disease (Artieda *et al.*, 2008). In 2012, a study revealed that -427C allele is associated with higher levels of TGs in a group of Argentinean middle-aged women (Bañares *et al.* 2012) and in 2013 another study identified -427T>C to comprise a functional polymorphism through transient transfection experiments and electrophoretic mobility shift assays in human HepG2 cells, which is independently associated with plasma apoE concentration. It was observed that -427C allele attributed to increased levels of apoE mRNA (Mannila *et al.*, 2013).

1.1.5.4.3. *-219G>T substitution*

The SNP -219G>T (rs405509) of the CCHS comprises a substitution of the most frequent allele G with the less frequent allele T at position -203 upstream of +1 in the apoE gene. In 1998 Artiga *et al.*, identified the existence of -219G>T variant. Transient transfection experiments were performed in constructs bearing the T allele

and constructs bearing the G allele and it was demonstrated that allele T contributes to a significant decrease in apoE promoter activity compared to the most common G allele (Artiga *et al.*, 1998), which is in contrast with a more recent study, in which -219G and -219T alleles showed no difference in apoE promoter activity (Mannila *et al.*, 2013). Two separate studies in 2000 and 2001 revealed that -219G>T polymorphism is associated with increased risk of myocardial infarction (independently of the apoE ε2/ε3/ε4 alleles), premature coronary artery disease and decreased plasma apoE concentrations (Lambert *et al.*, 2000; Viitanen *et al.*, 2001). It has been demonstrated in individuals from the ECTIM study, that -219T allele associates with lower plasma apoE concentration in a dose-dependent manner comparing to -219G allele (Lambert *et al.*, 2000). Another study revealed that -219T allele increases susceptibility of plasma LDL to oxidative modifications and enhances the response of apo-B and LDL cholesterol to the presence of saturated fatty acid (SFA) in the diet of healthy men. In -219T allele carriers, decrease in LDL-C and apoB was significantly higher when they turned from SFA to CHO diet (Moreno *et al.*, 2004). An association has been made between the ε4 allele of apoE and the -219T allele in ischemic heart disease patients (Artieda *et al.*, 2008).

1.1.6. Aim of the study

As mentioned before, CETP comprises a molecule of great importance in the exchange of cholesteryl esters for triglycerides between HDL cholesterol and apoB-containing lipoproteins (VLDL and LDL) and several studies have demonstrated an association between HDL cholesterol levels and CETP. Consistent epidemiologic evidence exists on the protective role of elevated levels of HDL cholesterol against cardiovascular disease, a fact that led to the development of CETP inhibitors with the hope that increasing HDL cholesterol through this mechanism would be beneficial in the reduction of vascular events. However, this was hardly the case, since CETP inhibitors, like torcetrapib, increased mortality rates in individuals at high risk of cardiovascular disease, leading to a debate as to whether HDL cholesterol and/or CETP are in fact causal for heart disease and should remain viable targets for the designing of new drugs (Rader, 2007; Tall *et al.*, 2007). One approach in our understanding of the causal pathways is to examine whether genetic polymorphisms could impact on vascular risk.

ApoE possesses a dual function, by contributing to the clearance of triacylglycerol-rich lipoproteins and by participating in the biogenesis of HDL particles containing apoE (Kypreos *et al.*, 2007), a fact which was well demonstrated using adenovirus-mediated gene transfer of apoE in apoA-I-deficient mice or in ABCA1-deficient mice (Zannis *et al.*, 2008). These studies established the significant role of apoE in lipid metabolism, so it is importnant to elucidate whether genetic polymorphisms could affect its regulation and expression levels, either on a beneficial or on a less favorable way.

So the aim of this study is the identification of genetic changes in the apoE and CETP genes mapped up from the Copenhagen City Heart Study, which may contribute to an up- or down-regulation of the corresponding genes, as well as the establishment of the mechanisms, by which these changes in the DNA sequence affect the expression levels of the aforementioned genes. For that purpose weintend to generate and utilize luciferase expression vectors bearing the CETP promoter region or/and the CETP first intron region, as well as luciferase expression vectors bearing the apoE promoter region, bearing either the most common or the less common allele for each SNP, in transient transfection assays in human hepatoblastoma (HepG2) cells.

1.2. MATERIALS AND METHODS

Materials

Oligonucleotide primers were either by Eurofins Genomics-VBC Biotech (Vienna, Austria) or by Minotech IMBB-FORTH (Crete, Greece). *PfuUltra* high fidelity DNA polymerase (2.5U/µL) was from Agilent Technologies. Restriction enzymes *Sac*I, *Xho*I, *Dpn*I, *Kpn*I and *Sal*I, NEB4 Buffer, BSA 100X, Antarctic Phosphatase (5U/µL) and T4 DNA ligase were from New England Biolabs. Quick Change Site Directed Mutagenesis Kit was by Stratagene. Dulbecco's modified Eagle's medium (DMEM), penicillin/streptomycin and trypsin/EDTA for cell culture were purchased from Invitrogen/Life Technologies. Fetal bovine serum (FBS) was purchased from BioChrom Labs. The luciferase assay system and the cell lysis buffer were purchased from Promega Corp. ONPG (*o*nitrophenyl β-D-galactopyranoside) was purchased from Sigma-Aldrich. pGL3 basic and pGL4.10 [luc2] reporter luciferase vectors were from Promega. CMV-βgal was described previously (Prokova et al. J Biol Chem. 2002 Mar 15;277(11):9342-50; (Mosialou et al., 2010; Prokova et al., 2002). Salmon sperm DNA was from Invitrogen.

Methods

1.2.1. Genomic DNA Purification

The Genomic DNA Purification Kit Fermentas by ThermoFischer Scientific was used to isolate genomic DNA either from HepG2 cells or from Peripheral Blood Mononuclear Cells (PBMCs) according to the manufacturer's protocol. Briefly the protocol was the following:

- Collect cell culture cells (0,4-0,6 X 10^6) by centrifugation at 5000 rpm, for 5 minutes at 4 ^{o} C and remove supernatant from the cell pellet.
- Resuspend the pellet in 200 µL TE Buffer.
- Mix 200 µL of sample with 400 µL of lysis solution and incubate at 65 ^{o} C for 5 minutes.
- Add 600 µL of chloroform, gently emulsify by inversion (3-5 times) and centrifuge the sample at 10000 rpm for 2 minutes.
- Prepare precipitation solution by mixing 720 µL of sterile deionized water with 80 µL of 10X concentrated Precipitation Solution.
- Transfer the upper aqueous phase containing DNA to a new tube and add 800 µL of freshly prepared precipitation solution, mix gently by several inversions at room temperature (18-25 ^{o} C) for 1-2 minutes and centrifuge at 10000 rpm for 2 minutes.
- Remove supernatant completely (do not dry) and dissolve DNA pellet in 100 µL of NaCl solution by gentle vortexing.
- Add 300 µL of cold 100% ethanol, let the DNA precipitate overnight (12-16 hours) at -20 ^{o} C and spin down (10000 rpm for 3-4 minutes). Remove ethanol. Wash

the pellet once with 70% cold ethanol and dry the pellet. Dissolve DNA in 100 µL of sterile deionized water by gentle vortexing.

1.2.2. Quantification of DNA concentration

The ds DNA concentration was measured with the Infinite M200 PRO reader device. The oligonucleotide primer's concentration (ssDNA) was measured with the same device and the final working concentration of the oligonucleotides was calculated according to the molecular weight (Mr) provided along with each oligonucleotide.

1.2.3. Polymerase Chain Reaction (PCR)

1.2.3.1. Human CETP Promoter

In order to isolate the HCETP promoter region of interest (-1054/+27) the following PCR program was used:

1) 94 °C for 3 minutes
2) 94 °C for 1 minute
3) 61 °C for 40 seconds
4) 72 °C for 1 minute and 20 seconds

} x 29 cycles

5) 72 °C for 3 minutes
6) 4 °C for ever
7) End

The PCR reaction was performed with *GoTaq* DNA polymerase (5U/µL) (Invitrogen) at a final volume of 50 µL.

The oligonucleotide primers used are shown in Table 1.

Promoter Region	Insert Length (bps)	Primer Name		Restriction Enzyme Site of Insertion on pGL4.10 [luc2] vector
		Forward	Reverse	
-1054/+27	1081	HCETP -1054F-T new	HCETP +27R-G new	*Sac*I-*Xho*I
-1054/+27	1081	HCETP -1054F-T new	HCETP -9T-AS new	*Sac*I-*Xho*I
-1054/+27	1081	HCETP -1073C-S new	HCETP +27R-G new	*Sac*I-*Xho*I

Table 1. Oligonucleotide primers pairs used to obtain the desired human CETP promoter region from total HepG2 genomic DNA template. Restriction enzyme sites (*Sac*I or *Xho*I) were designed in each primer sequence, so as to succeed the desirable site of insertion of the DNA fragment in pGL4.10 [luc2] plasmid vector.

The oligonucleotide primers sequences are shown in Table 2.

Primer Name	Sequence	Start	End
HCETP -1054F-T new	5'- CCCGAGCTCTCTGAGCCTTGGGAAACAGT -3'	-1054	-1035
HCETP +27R-G new	5'- CCCCCCTCGAGGTTATCAGGCAGTGGTGTG -3'	+27	+8
HCETP -1073C-S new	5'- CCCGAGCTCTCTGAGCCCTGGGAAACAGT -3'	-1054	-1035
HCETP -9T-AS new	5'- CCCCCCTCGAGGTTATCAAGCAGTGGTGTG -3'	+27	+8

Table 2. Oligonucleotide primers sequence. Start position and end position of the primer can be seen, as well as, the restriction site and the extra nucleotides added before, so as for the restriction enzyme to cleave efficiently. The exact DNA sequence is provided in the Appendix.

3-5 nucleotides at the 5' end of the oligonucleotides where added on the side of the recognition site to cleave efficiently. The additional bases were chosen so that palindromes and primer dimers are not formed.

1.2.3.2. Human CETP Intron 1

The first intron of the human CETP gene, was amplified from an already existed plasmid construct in our laboratory, (+191/+1056) Intron 1 CETP-luc on pGL3 basic vector (Promega), with the use of the appropriate oligonucleotide primers.

The PCR program used was the following:

1) 94 ^{o}C for 4 minutes
2) 94 ^{o}C for 30 seconds
3) 60 ^{o}C for 1 minute
4) 72 ^{o}C for 1 minute

   } x 22 cycles

5) 72 ^{o}C for 4 minutes
6) 4 ^{o}C for ever
7) End

The oligonucleotide primers used were the following (Table 3):

Intron 1 Region	Insert Length (bps)	Primer Name		Restriction Enzyme Site of Insertion on pGL4.10 [luc2] vector
		Forward	Reverse	
+191/+1056	865	HCETP INTR1 +191F	HCETP INTR1 +1056R	*Sal*I

Table 3. Oligonucleotide primers pairs used to obtain the desired human CETP intron 1 region from (+191/+1056) INTRON 1 CETP-luc plasmid construct DNA template. Restriction enzyme sites (*Sal*I) were designed in each primer sequence, so as to succeed the desirable site of insertion of the DNA fragment in pGL4.10 [luc2] plasmid vector.

The oligonucleotide primers sequence is shown in Table 4.

Primer Name	Sequence	Start	End
HCETP INTR1 +191F	5'- CCCGTCGACGACACCCACTATGCCAGGAG -3'	+191	+210
HCETP INTR1 +1056R	5'- CCCGTCGACTAGAGGGGAGGGCAGTGGAT -3'	+1037	+1056

Table 4. Oligonucleotide primers sequence. Start position and end position of the primer can be seen, as well as, the restriction site and the extra nucleotides added before, so as for the restriction enzyme to cleave efficiently. The exact DNA sequence is provided in the Appendix.

3-5 nucleotides at the 5' end of the oligonucleotides where added on the side of the recognition site to cleave efficiently. The additional bases were chosen so that palindromes and primer dimers are not formed.

After the PCR reaction was complete, the PCR products were run in agarose gel of the appropriate concentration (usually 1 %), and the size of the DNA fragments of interest was verified with the appropriate DNA ladder (λ BstEII ladder).

1.2.4. PCR clean-up Gel extraction

The NucleoSpin Gel and PCR Clean-up Kit, by Macherey-Nagel (Düren, Germany), was used, in order to gel extract the PCR product from agarose gel. Briefly the protocol recommended was the following:

Excise DNA fragment/solubilize gel slice

- Take a clean scalpel to excise the DNA fragment from the agarose gel. Remove all excess agarose.
- Determine the weight of the gel slice and transfer it to a clean tube.
- For each 100mgr of agarose gel < 2% add 200μL buffer NTI. For gels containing > 2% agarose, double the volume of buffer NTI.
- Incubate sample for 5-10 minutes at 50 ^{o} C. Vortex the sample briefly every 2-3 minutes until the gel slice is completely dissolved.

Bind DNA

Place a NucleoSpin Gel and PCR Clean-up Column into a collection tube and load up to 700 μL sample.

- Centrifuge for 30 seconds at 11000 x g. Discard flow-through and place the column back into the collection tube. Load remaining sample if necessary and repeat the centrifugation step.

Wash silica membrane

- Add 700 μL buffer NT3 to the NucleoSpin Gel and PCR Clean-up Column. Centrifuge for 30 seconds at 11000 x g. Discard flow-through and place the column back into the collection tube.

- Repeat previous washing step.

Dry silica membrane

- Centrifuge for 1 minute at 11000 x g to remove buffer NT3 completely.

Elute DNA

- Place the NucleoSpin Gel and PCR Clean-up Column into a new 1,5 mL microcentrifuge tube. Add 30 µL buffer NE and incubate at room temperature (18-25 ^{o} C) for 1 minute. Centrifuge for 1 minute at 11000 x g.

For DNA fragments larger than 1000 bps, buffer NE should be pre-heated at 70 ^{o} C and incubated for 5 minutes.

1.2.5. Digestion Reaction

One unit is defined as the amount of enzyme required to digest 1 µgr of DNA in 1 hour at 37°C in a total reaction volume of 50 µL.

The restriction enzymes, BSA 10X and NEB4 Buffer used for the restriction reactions were by New England Biolabs.

Briefly, 1 µgr of dsDNA from gel extraction and 5 µgr of pGL4.10 [luc2] vector (Promega) were restricted with SacI-HF restriction enzyme (20U/µL) and XhoI restriction enzyme (20U/µL) and incubated for 3 hours at 37 ^{o} C. In addition, 5 µgr of plasmid construct (-668/+205) ABCA1-luc on pGL3 basic vector (Promega), which was already existed in our laboratory, were restricted with SalI-HF restriction enzyme (20U/µL) and incubated for 3 hours at 37 ^{o} C.

1.2.6. Heat Inactivation

Incubate for 20 minutes at 65 ^{o} C.

After heat inactivation of the restriction enzymes, both insert and plasmid vector were gel extracted with the PCR clean-up Gel extraction protocol mentioned before, and the DNA concentration was quantified with the method previously described.

1.2.7. Dephosphorylation

Antarctic Phosphatase catalyzes the removal of 5’ phosphate groups from DNA. Since phosphatase-treated fragments lack the 5’ phosphoryl termini required by ligases, they cannot self-ligate, so this property can be used to decrease vector background in cloning strategies, in the case when a digestion reaction with a single enzyme has performed.

The Antarctic Phosphatase (5U/µL) by New England Biolabs was used in cloning where vector has been digested with a single restriction enzyme. Briefly the protocol was the following:

1-5 μgr of DNA are added to 1/10 volume of 10X Antarctic Phosphatase Reaction Buffer. Then 1 μL of Antarctic Phosphatase (5 units) are added to the reaction and the final volume is adjusted with deionized H_2O.

Incubation is performed for 15 minutes at 37 ^{o} C for 5' extensions and then heat inactivation for 5 minutes at 70 ^{o} C.

1.2.8. Ligation

The ligation reaction was performed with the use of the T4 DNA ligase enzyme, by New England Biolabs. Different vector to insert molar ratios and different vector concentrations were tried, in order to succeed the desired construct at a final reaction volume of 10 μL or 15 μL.

Incubation took place at 16 ^{o} C or 4 ^{o} C overnight (12-16 hours).

1.2.9. Transformation of bacterial strains with the use of competent cells

The bacterial strain that was used in the experiments is the DH10 beta strain, which is an *Escherichia coli* strain.

- Thaw competent cells on ice.
- Add either a part or the whole ligation reaction (total volume 15μL) to 100μL of DH10 beta competent cells.
- Mix gently by pipetting up and down or flicking the tube 4-5 times to mix the cells and DNA.
- Incubate for 30 minutes on ice.
- Heat shock for 45 seconds at 42 ^{o} C.
- Incubate on ice for 1 minute.
- Add 900μL fresh LB medium at room temperature (18-25 ^{o} C) without antibiotic.
- Incubate for 1 hour at 37 ^{o} C.
- In the meantime, prewarm fresh selection LB plates at 37 ^{o} C.
- Centrifuge at 3000 rpm for 5 minutes at room temperature (18-25 ^{o} C).
- Discard ~900μL of the flow through and resuspend the pellet in the volume of flow through remaining.
- With the use of a bacterial glass rod appropriate for plating, spread ~100μL of the cells and the ligation mixture onto the plates bearing the appropriate antibiotic for selection.

Incubate overnight (12-16 hours) at 37 ^{o} C.

1.2.10. LB plates supplemented with Ampicillin

- Weight the following:
  - 5gr NaCl (Merck, Darmstadt, Germany)

- 2,5gr Yeast Extract (Sigma-aldrich, St. Louis, USA)
- 5gr Peptone

- Dissolve the above in deionized water at a final volume of 500mL.
- After the solution is fully dissolved add 7,5gr Agar (Sigma-aldrich, St. Louis, USA).
- Sterilize short afterwards in an autoclave machine.
- After the sterilization is complete, stir until a temperature of ~37 ^{o}C is achieved, then add 1mL Ampicillin and stir for 2-3 minutes.
- Apply the medium in plates (Sarstedt, Nümbrecht, Germany) and allow them to chill at room temperature (18-25 ^{o}C) for 1 hour.
  Store at 4 ^{o}C for future use.

1.2.11. Site Directed Mutagenesis

In order to achieve *in vitro* site-directed mutagenesis of a specific position at the promoters of both human CETP and human apoE genes, as well as at intron 1 of the human CETP gene, the QuickChange Site-Directed Mutagenesis Kit, by Stratagene was used.

In the beginning, a PCR reaction is performed as following:

- 5 µL of 10x reaction buffer
- X µL (10ngr) of dsDNA template
- X µL (125ngr) of oligonucleotide primer #1
- X µL (125ngr) of oligonucleotide primer #2
- 1 µL of dNTPs mix 10mM
- 1 µL of *PfuUltra* HF DNA polymerase (2.5U/µL) (Agilent Technologies)
- ddH_2O to a final volume of 50 µL

The recommended PCR program is the following:

1) 95 ^{o}C for 1 minute
2) 95 ^{o}C for 50 seconds
3) 60 ^{o}C for 50 seconds
4) 68 ^{o}C for 1minute/kb of plasmid length

   } x 17 cycles

5) 68 ^{o}C for 7 minutes
6) 4 ^{o}C for ever
7) End

After the PCR reaction, 1 µL of DpnI restriction enzyme (20U/µL) (New England Biolabs) is added to the total volume of the aforementioned reaction, in order to digest the parental methylated and hemimethylated DNA. The reaction is incubated at 37 ^{o}C for 1 hour.

Finally, either 2-3 μL or the total volume of the DpnI restriction reaction are used to transform DH10 beta competent cells, according to the transformation protocol described previously. Transformed cells are plated is LB plates supplemented with ampicillin and incubated overnight (12-16 hours) at 37 ^{o}C.

Some colonies from the plate are being chosen for plasmid DNA isolation at short scale (MiniPreps), while whether the mutation has been inserted or not, is being checked by sequencing of the plasmid DNA with the appropriate primers each time.

1.2.11.1. Mutagenesis on human CETP Promoter

As regards the promoter of the human CETP gene, the mutations inserted were the following:

- -998G>A (position: -971)
- -854C<T (position: -827)
- -658C>A (position: -631)
- -656C>A (position: -629)
- -65G>A (position: -38)
- -9C>T (position: +19)
- +279G>A Taq1B (position: +424)

The ds DNA template (either 50 ngr or 100 ngr) for -854C<T mutagenesis was either the plasmid construct (-1054/+27) -854T CETP-luc on pGL4.10 [luc2] vector or the plasmid construct (-1054/+27) -1073C/-854T CETP-luc on pGL4.10 [luc2] vector. The ds DNA template (either 50 ngr or 100 ngr) for -998G>A, -658C>A, -656C>A, -282C>A and -65G>A mutagenesis was the plasmid construct (-1054/+27) CETP-luc on pGL4.10 [luc2] vector.

The primers used, are shown in Table 5.

Primer Name	Sequence
HCETP -854C-FW	5'- GGAGGCAGCACTTGGCCATCTGGTCACAGTTG -3'
HCETP -854C-RV	5'- CAACTGTGACCAGATGGCCAAGTGCTGCCTCC -3'
HCETP-998A-FW	5'- CAGGTGCAAACTCAGGACTAGGGCAGG -3'
HCETP-998A-RV	5'- CCTGCCCTAGTCCTGAGTTTGCACCTG -3'
HCETP-658A-FW	5'- GTATACACCCCCAGAGTTATTTTATGC -3'
HCETP-658A-RV	5'- GCATAAAATAACTCTGGGGGTGTATA -3'
HCETP-656A-FW	5'- CAGAGGCTGTATACCCACCCAGAGTTATTTTATG -3'
HCETP-656A-RV	5'- CATAAAATAACTCTGGGTGGGTATACAGCCTCTG -3'
HCETP-282A-FW	5'- GAATTAATGGAAACATATTAAGCAATTATCCAG -3'
HCETP-282A-RV	5'- CTGGATAATTGCTTAATATGTTTCCATTAATTC -3'
HCETP-65A-FW	5'- CTCATGTTCCGTGGGGGCTGGGCAGACATAC -3'
HCETP-65A-RV	5'- GTATGTCTGCCCAGCCCCCACGGAACATGAG -3'

Table 5. Oligonucleotide primers pairs used to obtain the desired human CETP promoter mutations -854C, -998A, -658A, -656A and -282A and the human CETP exon 1 mutation -65A. With yellow the inserted mutation is indicated.

Regarding the mutagenesis -998G>A, -854C<T, -658C>A, -656C>A, -282C>A and -65G>A the following program was used:

1) 95 °C for 1 minute
2) 95 °C for 50 seconds
3) 60 °C for 50 seconds
4) 68 °C for 6 minutes
5) 68 °C for 7 minutes
6) 4 °C for ever
7) End

(2–4: x 17 cycles)

From these mutagenesis experiments the following constructs were obtained:

- ✓ (-1054/+27) CETP-luc
- ✓ (-1054/+27) -998A CETP-luc
- ✓ (-1054/+27) -658A CETP-luc
- ✓ (-1054/+27) -656A CETP-luc
- ✓ (-1054/+27) -65A CETP-luc

on pGL4.10 [luc2] vector.

However, the mutagenesis of 282C>A was unsuccessful in all the different conditions tried.

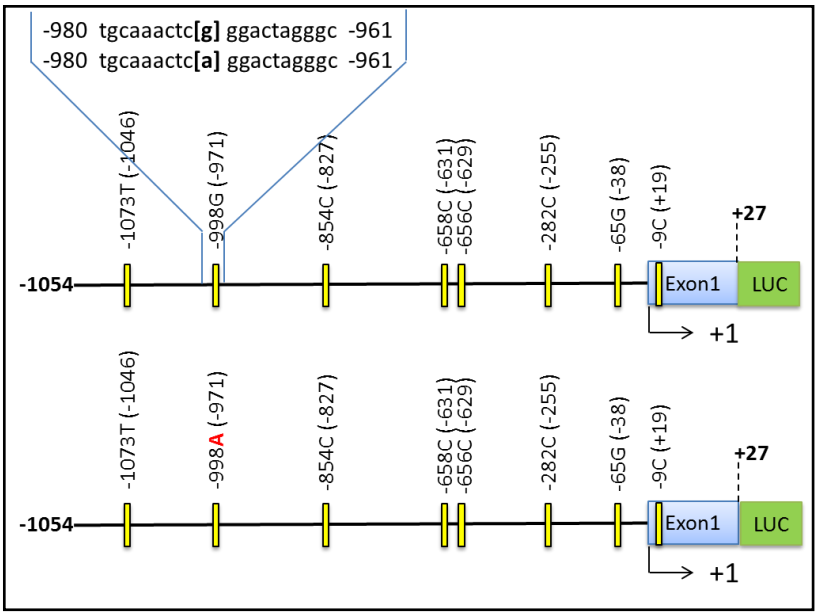

Figure 5. Mutagenesis at position -971 upstream to the transcription start site. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

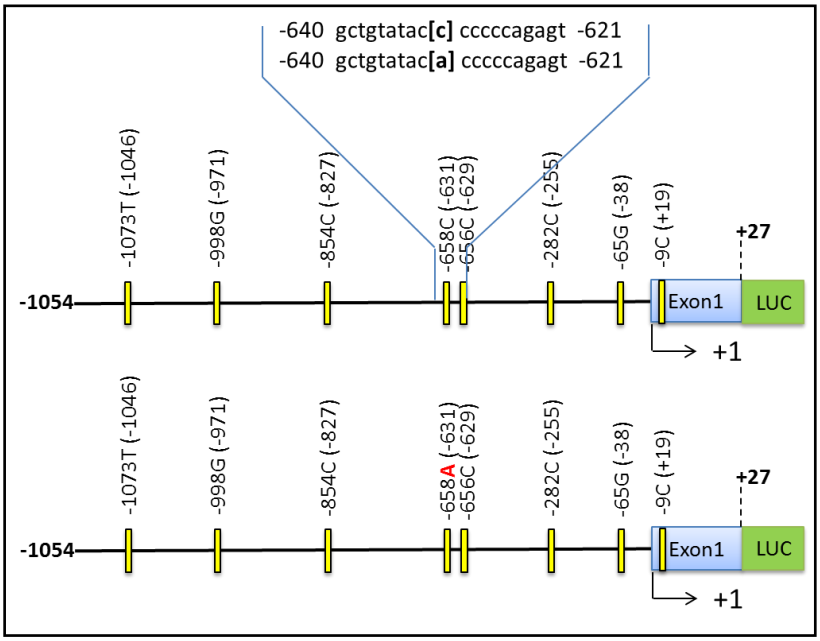

Figure 6. Mutagenesis at position -631 upstream to the transcription start site. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

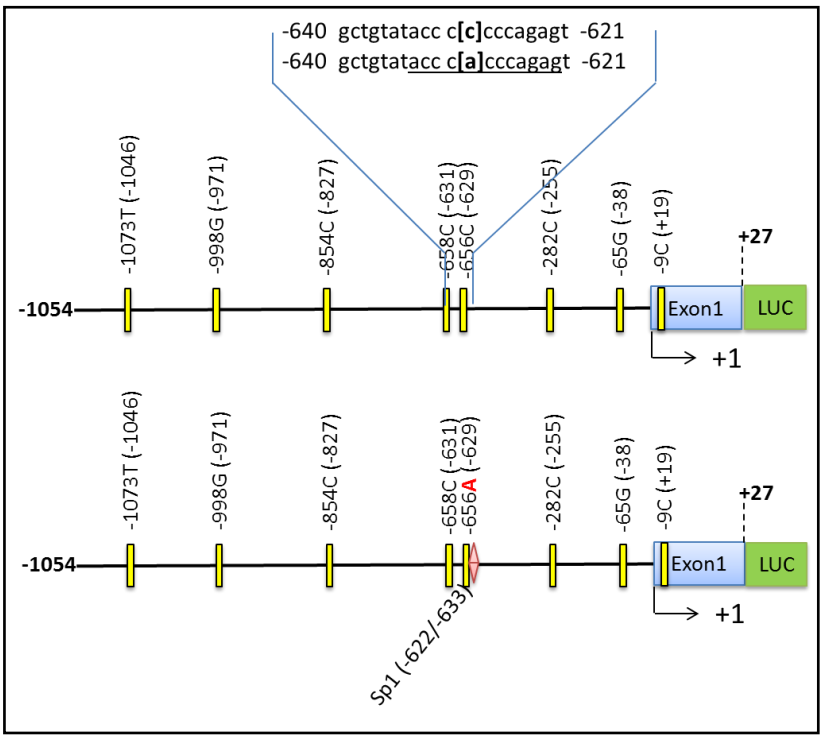

Figure 7. Mutagenesis at position -629 upstream to the transcription start site. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

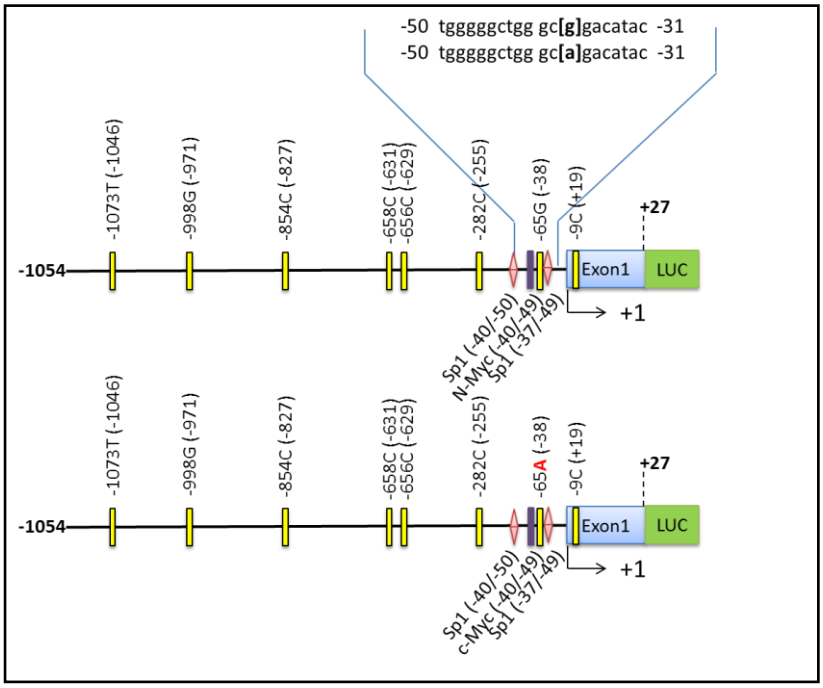

Figure 8. Mutagenesis at position -38 upstream to the transcription start site. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

1.2.11.2. Mutagenesis on human CETP intron 1

As regards the first intron of the human CETP gene, the mutation inserted was the following:

- +279G>A Taq1B (position: +424)

The ds DNA template (either 50 ngr or 100 ngr) for +279G>A was the plasmid construct (-1054/+27) CETP-luc-(+191/+1056) INTRON 1 CETP on pGL4.10 [luc2] vector.

The primers used, are shown in Table 6.

Primer Name	Sequence
HCETP +279A-FW	5'- CAGAATCACTGGGGTTCAAGTTAGGGTTCAGATC -3'
HCETP +279A-RV	5'- GATCTGAACCCTAACTTGAACCCCAGTGATTCTG -3'

Table 6. Oligonucleotide primers pairs used to obtain the desired human CETP intron 1 mutation +279A. With yellow the inserted mutation is indicated.

Regarding the mutagenesis +279G>A the following program was used:

1) 95 °C for 1 minute
2) 95 °C for 50 seconds
3) 60 °C for 50 seconds
4) 68 °C for 7 minutes

   } x 17 cycles

5) 68 °C for 7 minutes
6) 4 °C for ever
7) End

It must be mentioned at this point, that the same conditions were used with the primers HCETP-656A-FW and HCETP-656A-RV shown in Table 5 and eventually the following constructs were obtained:

- ✓ (-1054/+27) -656A CETP-luc-(+191/+1056) INTRON 1 CETP
- ✓ (-1054/+27) CETP-luc-(+191/+1056) +279A INTRON 1 CETP
- ✓ (-1054/+27) -656A CETP-luc-(+191/+1056) +279A INTRON 1 CETP

on pGL4.10 [luc2] vector.

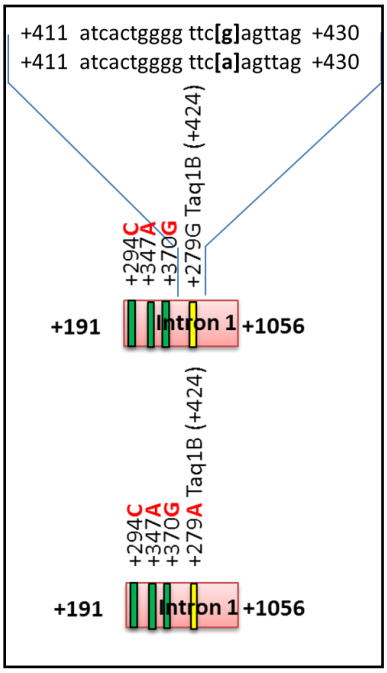

Figure 9. Mutagenesis at position +424 downstream to the transcription start site. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

1.2.11.3. Mutagenesis on human apoE promoter

Regarding the promoter of the human apoE gene, the mutations inserted were the following:

- ➢ -427T>C (position: -411)
- ➢ -219G>T (position: -203)

The ds DNA template (10 ngr, 50 ngr or 100 ngr) for -491A>T, -427T>C and -219G>T mutagenesis was the plasmid construct (-486/+89) apoE-luc on pGL3 basic vector.

The primers used, are shown in Table 7.

Primer Name	Sequence
HapoE-491T-FW	5'- GTACCGCTGGTCTCAATCTCCTGACCTTAAGTG -3'
HapoE-491T-RV	5'- CACTTAAGGTCAGGAGATTGAGACCAGCGGTAC -3'
HapoE-427C-FW	5'- GTGCTGGGATTACAGGCGTGAGCCACCG -3'
HapoE-427C-RV	5'- CGGTGGCTCACGCCTGTAATCCCAGCAC -3'
HapoE-219T-FW	5'- GGAGGAGGGTGTCTGTATTACTGGGCGAGGTG -3'
HapoE-219T-RV	5'- CACCTCGCCCAGTAATACAGACACCCTCCTCC 3'

Table 7. Oligonucleotide primers pairs used to obtain the desired human apoE promoter mutations -491T, -427C and -219T. With yellow the inserted mutation is indicated.

Regarding the mutagenesis -491A>T, -427T>C and -219G>T the following program was used:

1) 95 °C for 1 minute
2) 95 °C for 50 seconds
3) 60 °C for 50 seconds
4) 68 °C for 5 minutes
5) 68 °C for 7 minutes
6) 4 °C for ever
7) End

x 17 cycles

From these mutagenesis experiments the following constructs were obtained:

- ✓ (-486/+89) -427C apoE-luc
- ✓ (-486/+89) -219T apoE-luc

on pGL3 basic vector.

However, the mutagenesis of -491A>T was unsuccessful in all the different conditions tried.

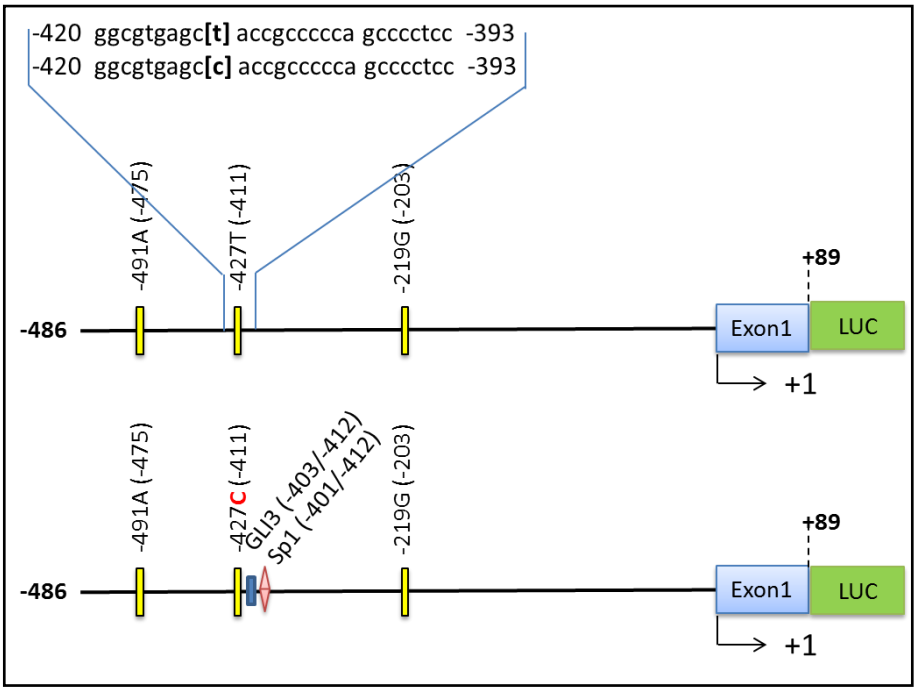

Figure 10. Mutagenesis at position -411 upstream to the transcription start site. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

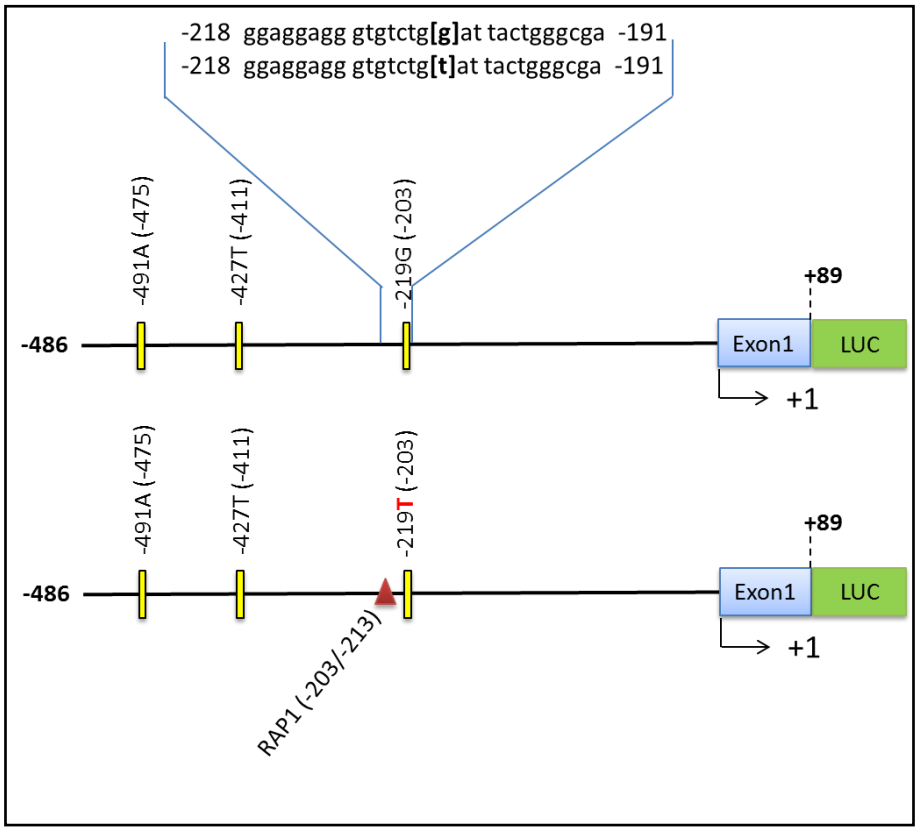

Figure 11. Mutagenesis at position -203 upstream to the transcription start site. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

1.2.12. MiniPreps

In order to isolate plasmid DNA from bacterial cultures at short scale (MiniPreps), the Plasmid DNA Purification Kit by QIAGEN was used.

Cultivate and harvest bacterial cells

- At first, prepare a 2mL bacterial culture, by inoculating the appropriate volume of LB medium (plus antibiotic) with a single colony picked from a freshly transformed plate. Shake the culture overnight (12-16 hours) at 37 ^{o}C.
- Transfer 1,5mL of the bacterial culture in a 1,5mL eppendorf.
- Centrifuge the culture at 13000 rpm for 1 minute at room temperature (18-25 ^{o} C). Carefully discard supernatant.

Cell lysis

- Carefully resuspend the pellet of bacterial cells in 300µL Buffer P1+RNase A (Resuspension Buffer).
- Add 300µL Buffer P2 (Lysis Buffer) to the suspension. Mix gently by inverting the tube 6-8 times. Incubate the mixture at room temperature (18-25 ^{o}C) for 2-3 minutes (max. 5 minutes). Do not vortex, as this will release contaminating chromosomal DNA from cellular debris into the suspension.

- Add 300μL pre-cooled Buffer P3 (Neutralization Buffer) (4 ^{o}C) to the suspension. Immediately mix the lysate gently by inverting the tube 6-8 times until a homogeneous suspension containing an off-white flocculate is formed. Incubate the suspension on ice for 10 minutes.
- Centrifuge at 13000 rpm for 10 minutes at room temperature (18-25 ^{o}C).

Precipitation

- Transfer the lysate from the previous step into a clean 2mL eppendorf.
- Add 800μL pre-cooled isopropanol to precipitate the plasmid DNA. Mix carefully and incubate on ice for 10 minutes.
- Centrifuge at 13000 rpm for 15 minutes at room temperature (18-25 ^{o}C). Carefully discard supernatant.

Wash and dry DNA pellet

- Add 500μL pre-cooled 70% ethanol to the pellet. Vortex briefly and centrifuge at 13000 rpm for 5 minutes at room temperature (18-25 ^{o}C).
- Carefully remove ethanol from the tube with a pipette tip.
- Add again 500μL pre-cooled 70% ethanol to the pellet. Vortex briefly and centrifuge at 13000 rpm for 5 minutes at room temperature (18-25 ^{o}C).
- Carefully remove ethanol from the tube with a pipette tip.
- Allow the pellet to dry at room temperature (18-25 ^{o}C) for 10 minutes.

Reconstitute DNA

- Dissolve pellet in 35μL of sterile deionized H_2O.
- Allow pellet to dissolve by incubating overnight at 4 ^{o}C.

In order to select the positive for the insert colonies, plasmid DNA from each MiniPrep was run in agarose gel of the appropriate concentration (1 %) and the DNA fragment of interest was verified with the use of the appropriate ladder (λ BstEII ladder), while whether the enzymes have performed the restriction reaction successfully or not, was verified by running unrestricted plasmid vector and comparing it with the MiniPreps.

After the selection of the positive, for the insert, colonies, the latter were sent for Sequencing with the appropriate primers, so as to verify whether the DNA sequence was the same as the one provided by PubMed.

The positive colony with the desired sequence was selected to be grown in MidiPreps.

1.2.13. Sequencing

In order to validate that the DNA fragment of interest, which was either obtained through molecular cloning, or through site directed mutagenesis, did not bear any undesirable point mutations, the positive colonies were sent for sequencing in Macrogen Corporation (Amsterdam, The Netherlands), with the use of the appropriate primers each time.

1.2.14. LB medium

- Weight the following:
  - 10gr NaCl (Merck, Darmstadt, Germany)
  - 5gr Yeast Extract (Sigma-aldrich, St. Louis, USA)
  - 10gr Peptone
- Dissolve the above in deionized water at a final volume of 1000mL.
- Sterilize short afterwards in an autoclave machine.

1.2.15. Glycerol Stock

- Prepare a 2mL bacterial culture, by inoculating the appropriate volume of LB medium (plus antibiotic) with a single colony picked from a freshly streaked plate. Shake the culture for overnight (12-16 hours) at 37 ^{o}C.
- In a 1,5mL eppendorf add 300µL sterile glycerol (AppliChem, Darmstadt, Germany) and 700µL of the bacterial culture.
- Vortex briefly so as a homogenous mix is formed.
  Store at -80 ^{o}C for future use.

1.2.16. MidiPreps

In order to isolate plasmid DNA from bacterial cultures at medium scale (MidiPreps), the Plasmid DNA purification NucleoBond PC 100 by Macherey-Nagel (Düren-Germany), was used.

Cultivate and harvest bacterial cells

- At first, prepare a 2mL bacterial culture, by inoculating the appropriate volume of LB medium (plus antibiotic) with a single colony picked from a freshly streaked plate. Shake the culture for 3-4 hours at 37 ^{o}C.
- Then, set up an overnight culture, by transferring the 2mL bacterial culture at a final volume of 100mL LB medium (plus the appropriate antibiotic) and shake this culture overnight (12-16 hours) at 37 ^{o}C.
- Centrifuge the culture at 5000 rpm for 10 minutes at 4 ^{o}C. Carefully discard supernatant. Repeat centrifugation, so as to obtain a bacterial pellet from the total volume of the bacterial culture.

Cell lysis

- Carefully resuspend the pellet of bacterial cells in 8mL Buffer S1+RNase A (Resuspension Buffer).
- Add 8mL Buffer S2 (Lysis Buffer) to the suspension. Mix gently by inverting the tube 6-8 times. Incubate the mixture at room temperature (18-25 ^{o}C) for 2-3 minutes (max. 5 minutes). Do not vortex, as this will release contaminating chromosomal DNA from cellular debris into the suspension.
- Add 8mL pre-cooled Buffer S3 (Neutralization Buffer) (4 ^{o}C) to the suspension. Immediately mix the lysate gently by inverting the tube 6-8 times until a homogeneous suspension containing an off-white flocculate is formed. Incubate the suspension on ice for 5 minutes.

Equilibration of the column

- Equilibrate a NucleoBond AX 100 (Midi) Column with 2,5mL Buffer N2 (Equilibration Buffer). Allow the column to empty by gravity flow. Discard flow-through.

Clarification of the lysate

- Centifuge the lysate at 5000 rpm for 10 minutes at 4 ^{o}C.
- Place a NucleoBond Folded Filter in a funnel of appropriate size. Wet the filter with a few drops of Buffer N2 (Equilibration Buffer) and load the bacterial lysate obtained from the centrifugation onto the wet filter. Collect the flow-through in a separate vessel and proceed to the next step.

Binding

- Load half of the volume of the cleared lysate from the previous step onto the NucleoBond Column. Allow the column to empty by gravity flow.

Washing

- Wash the column with 10mL Buffer N3 (Wash Buffer). Discard flow through.
- Repeat the previous step and discard the flow through.

Elution

- Place the column in a clean tube and elute the plasmid DNA with 5mL Buffer N5 (Elution Buffer).

Precipitation

- Add 3,5mL room-temperature isopropanol to precipitate the eluted plasmid DNA. Mix carefully and centrifuge at 15000 x g for 30 minutes at 4 ^{o}C. Carefully discard supernatant.

At this step, store pellet for a while at 4 ^{o}C and move back to the equilibration of the column step. Equilibrate again the column with 2,5mL Buffer N2 (Equilibration Buffer). Allow the column to empty by gravity flow. Discard flow-through.

Then, proceed to the binding step, and load the second half of the volume of the cleared lysate onto the NucleoBond Column. Allow the column to empty by gravity flow.

Proceed to the Washing, the Elution and the Precipitation (add 3,5mL room-temperature isopropanol to precipitate the eluted plasmid DNA and mix carefully) steps as indicated before.

Place the precipitated plasmid DNA to the tube containing the first pellet (which is stored for a while at 4 ^{o}C and centrifuge at 15000 x g for 30 minutes at 4 ^{o}C. Carefully discard supernatant.

Wash and dry DNA pellet

- Add 2,5mL room-temperature 70% ethanol to the pellet. Vortex briefly and centrifuge at 15000 x g for 10 minutes at room temperature (18-25 ^{o}C).
- Carefully remove ethanol from the tube with a pipette tip. Allow the pellet to dry at room temperature (18-25 ^{o}C) for 10 minutes.

Reconstitute DNA

- Dissolve pellet in 125µL of buffer TE.
- Allow pellet to dissolve by incubating overnight at 4 ^{o}C.

After the desired plasmid construct of interest has been grown in medium scale, the activity of the promoter is tested with the Luciferase Assay method.

TE Buffer

Tris-Cl	10 mM PH 8
EDTA	1 mM

1.2.17. Plasmid Constructions

The plasmid constructs obtained from molecular cloning, on pGL4.10 [luc2] vector, were the following:

- ✓ (-1054/+27) CETP-luc
- ✓ (-1054/+27) -1073C CETP-luc
- ✓ (-1054/+27) -1073C/-854T CETP-luc
- ✓ (-1054/+27) -9T CETP-luc
- ✓ (-1054/+27) CETP-luc-(+191/+1056) Intron 1 CETP

✓ (-668/205) ABCA1-luc-(+191/+1056) Intron 1 CETP

Plasmid maps are available on the Appendix.

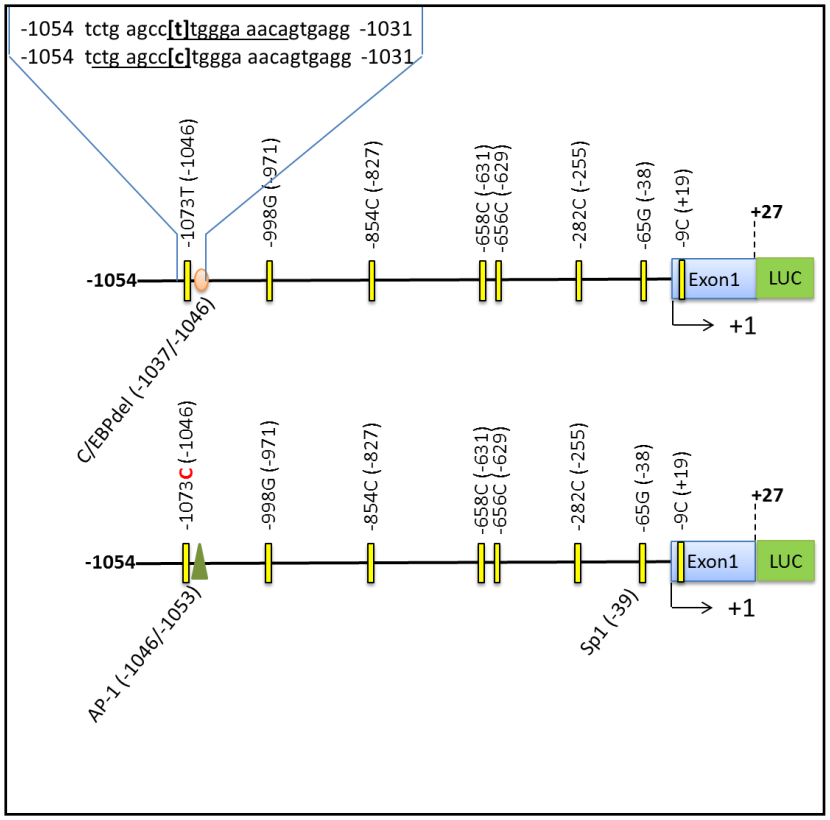

Figure 12. Mutation at position -1046 upstream to the transcription start site, inserted with PCR. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

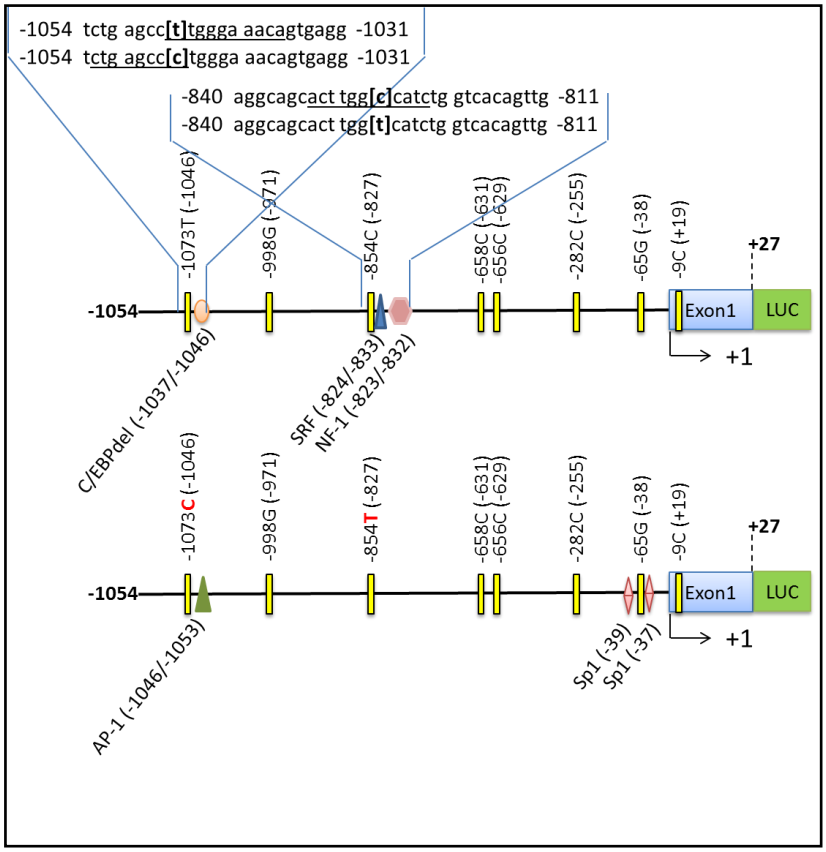

Figure 13. Mutations at positions -1046 and -827 upstream to the transcription start site, the former was inserted with PCR, the latter with Site Directed Mutagenesis. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

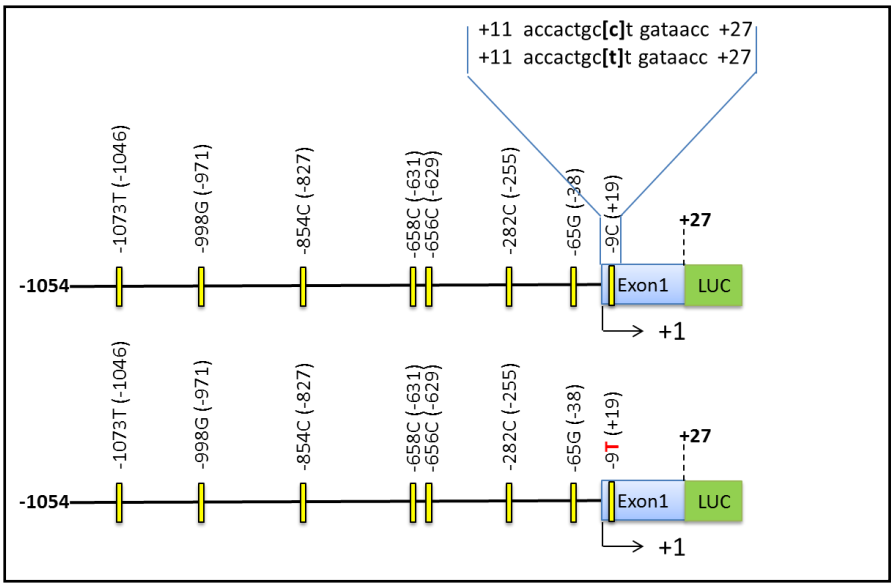

Figure 14. Mutation at position +19 downstream to the transcription start site, inserted with PCR. Potential transcription factor binding sites at this site are indicated, according to the online program AliBaba2.1 from www.gene-regulation.com.

1.2.18. Cell lines

The cell lines used, were the human hepatic cell line HepG2 (liver hepatocellular carcinoma), which were incubated at 37 ^{o} C, 5% CO_2, 100% humidity and at Dulbecco's Modified Eagle Medium 1X (DMEM) (Life Technologies), which contains 4,5 g/L D-glucose, supplemented with 10% Fetal Bovine Serum (FBS) (Biochrom) and 1% penicillin/streptomycin (P/S) (Life Technologies).

1.2.19. Cell culture-Passages

HepG2 cell line is characterized by the adherence of the cells to the surface of the culture flask. Cell passaging is performed when cell confluence is about 80%-90% in the culture flask and Trypsin-EDTA is used to disrupt cell adherence. Briefly the protocol is the following:

- Remove culture medium from the flask by aspiration.
- Wash the monolayer with 2 mL PBS (phosphate buffer saline) 1X.
- Remove PBS 1X by aspiration and add 1 mL cold Trypsin-EDTA.
- Incubate at 37 ^{o} C, 5% CO_2, 100% humidity for 2-3 minutes.
- Deactivate Trypsin-EDTA by adding 3 volumes of complete medium.

In the case of HepG2 cells, use a pipette tip in order to resuspend cells and to break up the clumps.

- Evaluate cell number in the cell suspension using a Neubauer Hemocytometer.
- Culture cells in complete medium under the conditions mentioned before.

1.2.20. Transient Transfections

Transient transfections were performed using the calcium phosphate [$Ca_3(PO_4)_2$] co-precipitation method. The aim of transient transfection assay is to evaluate the

luciferase activity (Luciferase assay). Such transfections are carried out in 6-well-plates using HepG2 cell line.

Day 0

Appropriate number of cells is placed in 6-well-plates, 500000 cells HepG2/well, and incubated at 37 ^{o}C, 5 % CO_2 and 100 % humidity overnight (12-16 hours).

Day 1

Prewarm 2X Hepes Buffered Saline (HBS), $CaCl_2$ 2M, water for injection (WFI) and complete medium. Use freshly diluted plasmid constructs, with water for injection, at a final concentration of 0,5 μgr/μL.

- At the beginning, dilute the desired volume of plasmid constructs at a final concentration of 0,5 μgr/μL.
- Vortex and short spin the diluted plasmid constructs.
- Set up the Pre-Mixes in 1,5 mL tube, by adding WFI, $CaCl_2$ 2M, salmon sperm DNA, CMV-β gal and plasmid constructs of interest. Use the aforementioned order when adding each one of them to the Pre-Mix tube.
  It must be noted at this point, that salmon sperm DNA is being used so as to counterbalance the amount of DNA among the samples.
  In addition, in experiments which include the evaluation of luciferase activity, the expression vector of beta galactosidase (1 μgr/well) is being used, so as to counterbalance the effectiveness of the transfection among the samples.
- Vortex and short spin Pre-Mixes.
- In 2 mL tubes (named as Mix) add equal volume (250 μL) of HBS/tube.
- Place the Mix tube on a vortex device and add the equivalent Pre-Mix drop by drop while vortexing the Mix tube. At this point, set up a timer at 15 minutes and wait.

The conditions for plasmid constructs for the CETP, promoter and first intron, are shown in Table 8, whereas the conditions for plasmid constructs for the apoE promoter are shown in Table 9.

- While waiting, discard old medium from the 6-well-plates and add 1 mL/well of fresh medium.
- After the 15 minute incubation add 250 μL of the Mix to each well, so as to have a duplicate.
- Incubate at 37 ^{o}C, 5 % CO_2 and 100 % humidity overnight (12-16 hours).

HBS 2X

NaCl	274 mM
KCl	10 mM
$Na_2HPO_4.H_2O$	1,5 mM
Dextrose	12 mM
Hepes	42 mM

PH 7±0,1

Plate	Cell line (number of cells/well)	DNA (µgr)/duplicate			$CaCl_2$ 2M (µL)/ duplicate	H_2O (µL)/ duplicate	2X HBS (µL)/ duplicate
	HepG2	Construct of interest (µgr)	CMV- beta gal (µgr)	ssDNA (µgr)			
6-well plate	500000	10	2	-	31	195	250

Table 8. Transient transfection conditions using HepG2 cells for plasmid constructions bearing the CETP gene fragments.

Plate	Cell line (number of cells/well)	DNA (µgr)/duplicate			$CaCl_2$ 2M (µL)/ duplicate	H_2O (µL)/ duplicate	2X HBS (µL)/ duplicate
	HepG2	Construct of interest (µgr)	CMV- beta gal (µgr)	ssDNA (µgr)			
6-well plate	500000	2	2	8	31	195	250

Table 9. Transient transfection conditions using HepG2 cells for plasmid constructions bearing the apoE promoter.

Day 2

- Change the medium by adding fresh medium (1 mL/well).
- Incubate at 37 ^{o}C, 5 % CO_2 and 100 % humidity overnight (12-16 hours).

1.2.21. Reporter Assays

For reporter assays the cells were plated in 6-well-plates and were subsequently transfected by the calcium phosphate method. Luciferase assays were performed using the luciferase assay kit from Promega Corp. according to the manufacturer's instructions. Normalization for transfection efficiency was performed by β-galactosidase assays.

- Discard medium from 6-well-plates.
- Add 1 mL/well 1X PBS (phosphate buffer saline) and shake gently.
- Discard PBS from 6-well-plates.
- Add 150 μL/well (for HepG2) or 200 μL/well (for HEK293T) of 1X Lysis buffer (Promega Lysis Buffer).
- Incubate on a shaker device at 260 rpm, for 10 minutes at room temperature (18-25 ^{o} C).
- Collect cells with the help of a cell scrapper and place the samples at -80 ^{o} C for at least 15 minutes.
- Thaw the samples for 2 minutes at 37 ^{o} C and vortex briefly afterwards.
- Centrifuge for 5 minutes at 13000 rpm at room temperature (18-25 ^{o} C).

After centrifugation, cell extracts are being used to evaluate luciferase and beta galactosidase activity.

Measurement of Luciferase activity

60 μL of cell extract are mixed with 60 μL of substrate (Promega) and the tube is placed in the luminometer device, where released relative light units are being counted.

Measurement of beta galactosidase activity

For each sample a mix is being prepared as following:

- 149 μL Sodium Phosphate Buffer
- 44 μL ONPG
- 2 μL 100X Salt Buffer

So, a final mix is being prepared, according to the number of samples each time.

In a 96-well-plate, add 5 μL of each sample cell extract. 5 μL of 1X Lysis buffer, are being used as blank.

With the use of Infinite M200 PRO reader device, 195 μL of the mix are being injected in each well at 37 ^{o} C. A measurement of the absorbance at 410 nm is being taken every 2 minutes for 25 cycles.

Sodium Phosphate buffer (P Buffer) 0,1M, pH 7.3:

Adjust the pH of Na_2HPO_4 0,1M solution to 7.3 using NaH_2PO_4 o,1M buffer.

O-nitrophenyl-galactopuranoside (ONPG) buffer:

ONPG is being diluted in P buffer at a final concentration of 8 mgr/mL.

100X Salt buffer:

KCl 3M	333 µL
$MgCl_2$	100 µL
β-mercaptoethanol	352 µL
Sterile H_2O	215 µL
Total Volume	1000 µL

1.2.22. Bioinformatics analysis

Bioinformatics analysis was performed, using the latest TRANSFAC version (2014.4). CETP sequence (-1070/+1090) and apoE sequence (-560/+160) from the Copenhagen City Heart Study were screened for both conditions (either most common alleles or less common alleles). Moreover, bioinformatics analysis was performed for the same DNA sequences for human CETP and human apoE using the online program AliBaba2.1 from www.gene-regulation.com, a program which gives a prediction of the transcription factor binding sites of a desired DNA sequence.

1.2.23. Statistical analysis

Data are presented as the average of normalized luciferase values %. Normalization was performed according to beta-galactosidase measurements. Microsoft Excel 2010 program was used for the statistical analysis of the results using t-test analysis, which returns the probability associated with a Student's two-sample equal variance (homoscedastic) t-test with a two-tailed distribution. For p values<0.053 the difference observed was statistically significant.

1.3. RESULTS

1.3.1. Polymorphisms -656C>A and -65G>A upstream of +1 transcription start site of CETP gene regulate the CETP promoter activity

In order to examine whether the polymorphisms that were identified upstream from the transcription start site of the CETP gene in the CCHS can regulate the activity of the CETP promoter, we generated plasmid constructs bearing the most common or the less common alleles at each polymorphic site under study. These constructs are: (-1054/+27) CETP-luc, (-1054/+27) -1073C CETP-luc, (-1054/+27) -1073C/-854T CETP-luc, (-1054/+27) -998A CETP-luc, (-1054/+27) -854T CETP-luc, (-1054/+27) -658A CETP-luc, (-1054/+27) -656A CETP-luc and (-1054/+27) -65A CETP-luc. One polymorphism (-9C>T) was identified downstream from the transcription start site, in the area spanning the first exon. The plasmid construct bearing this polymorphism is: (-1054/+27) -9T CETP-luc. The purpose of these plasmids was to use them in transient transfection assays, so as to examine whether the presence of a less common allele at a specific site affects human CETP promoter activity. For this purpose, we transiently transfected HepG2 cells with the different plasmid constructs for CETP polymorphisms, along with a beta-galactosidase expression vector and we evaluated the relative luciferase activity (%) in each case.

The results from reporter assays are shown in Figure 15 and the constructs reffering to these results in Figure 16. As demonstrated in the graph, the presence of the less common A allele at position -629 (corresponding to -656C>A polymorphism), reduced the promoter's activity to 91% relative to the most common allele ($P<0.053$). Importantly, the presence of the less common allele A at position -38 (corresponding to -65G>A polymorphism) significantly increased the activity of the CETP promoter to 138% relative to the most common allele ($P<0.001$). No statistically significant difference was observed among constructs bearing the less common alleles at positions: -1046 (corresponding to -1073T>C polymorphism), -1046/-827 (corresponding to -1073T>C and -854C>T polymorphisms), -971 (corresponding to -998G>A polymorphism), -827 (corresponding to -854C>T polymorphism), -631 (corresponding to -658C>A polymorphism) and +19 (corresponding to -9C>T polymorphism).

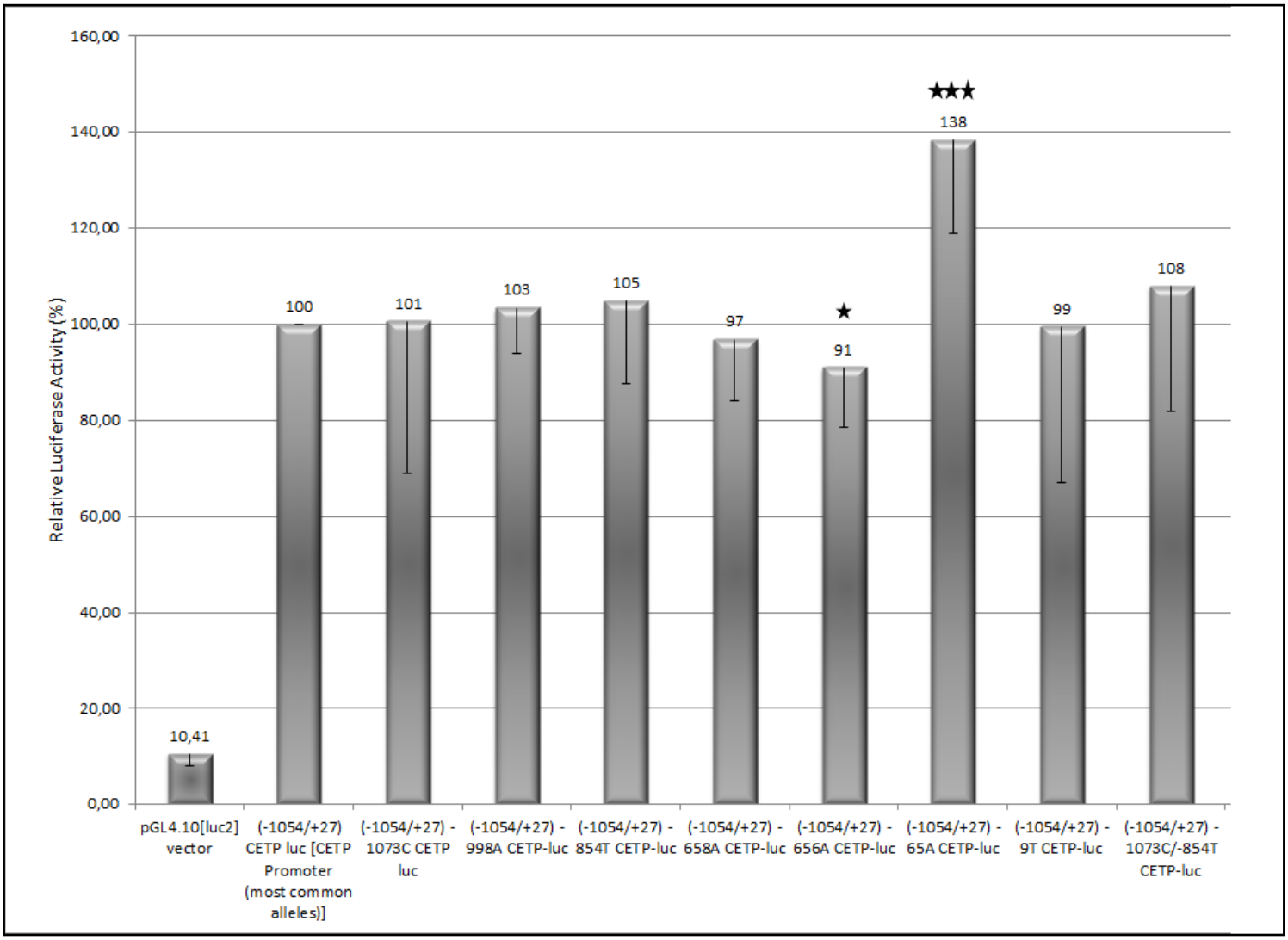

Figure 15. Polymorphisms -656C>A and -65G>A upstream of +1 transcription start site of the human CETP gene regulate the CETP promoter activity. HepG2 cells were transiently transfected with the luciferase reporter plasmid bearing the promoter region -1054/+27 of the CETP gene with the most common alleles (5μgr) for polymorphisms -1073T>C, -998G>A, -854C>T, -658C>A, -656C>A, -65G>A and -9C>T and luciferase reporter plasmids (5μgr) bearing the less common allele for a specific position each time, along with a beta-galactosidase expression vector (1μgr). From left to right, the CETP promoter constructs are the following: (-1054/+27) CETP-luc, (-1054/+27) -1073C CETP-luc, (-1054/+27) -998A CETP-luc, (-1054/+27) -854T CETP-luc, (-1054/+27) -658A CETP-luc, (-1054/+27) -656A CETP-luc, (-1054/+27) -65A CETP-luc, (-1054/+27) -9T CETP-luc and (-1054/+27) -1073C/-854T CETP-luc, whereas empty pGL4.10 [luc2] reporter vector was used as a control. It is observed that the presence of the less common -656A allele reduces the activity of the CETP promoter in a statistically significant manner (P<0.053), whereas the presence of the less common -65A allele, significantly increases the activity of the CETP promoter. A statistical significance is observed for *p* value<0.053. Asterisks represent the statistical significance, * for *p* value<0.053, ** for *p* value<0.01, * for *p* value<0.001 and ns for non-significant.**

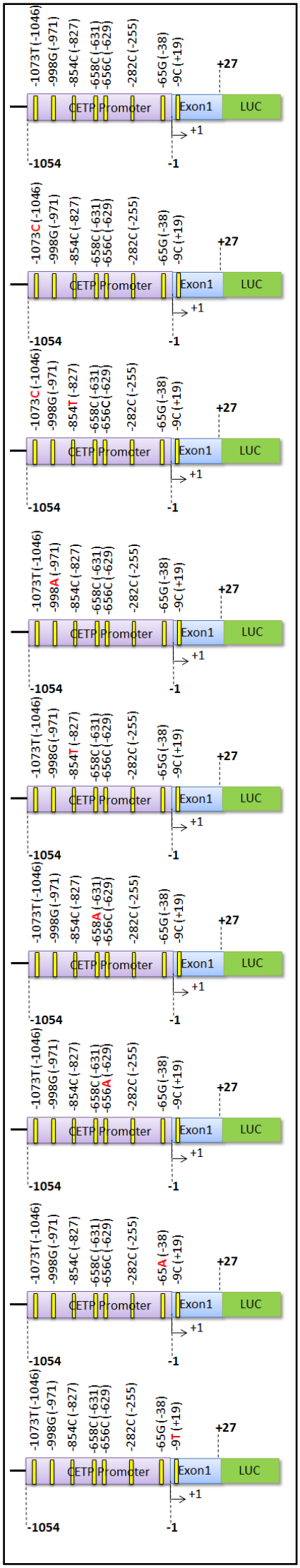

Figure 16. Constructs for CETP promoter (-1073T>C, -998G>A, -854C>T, -658C>A, -656C>A, -65G>A) and first exon (-9C>T) polymorphisms. -1054/+27 CETP gene region was inserted at positions *Sac*I and *Xho*I of the pGL4.10 [luc2] reporter vector. Site Directed Mutagenesis was performed in order to succeed the desired mutations. The CETP promoter region (-1054/-1) is depicted with purple, the exon 1 region (+1/+27) with blue, the luciferase gene with green, the CETP polymorphisms with yellow, whereas with red font less common alleles are shown.

1.3.2. Region +191/+1056 in the first intron of the CETP gene significantly reduced the activity of the CETP promoter

Previous experiments in our laboratory had tried to demonstrate whether the presence of the most common or the less common alleles of the polymorphism TaqIB (IVSI +279G>A), found at position +424 downstream of +1 transcription start site, an area spanning the first intron of the CETP gene, could affect the luciferase activity in reporter assays. Since TaqIB polymorphism is among the best characterized polymorphisms in the CETP gene and its alleles have been associated with disease phenotypes, it was crucial to elucidate whether it possesses activity through binding of potential transcription factors. For that purpose, the fragment +191/+1056 in the first intron of the CETP gene was inserted through molecular cloning technics in the pGL3 basic reporter vector (*Kpn*I and *Hind*III restriction enzymes) upstream of the luciferase gene, and transient transfection experiments were performed in HepG2 in order to evaluate the reporter luciferase activity compared to the empty pGL3 basic reporter vector. This experiment had shown that the construct bearing the +191/+1056 intron 1 region did not display any promoter activity (data not shown).

This observation combined with previous findings (Dachet *et al.*, 2000; Corbex *et al.*, 2000) that the TaqIB polymorphism is in high linkage disequilibrium with the -656C>A polymorphism found in the promoter of the human CETP gene, prompted us to examine the possibility that the above intronic regions of the CETP gene bearing the TaqIB polymorphism could regulate the activity of the CETP promoter bearing the -656C>A polymorphism. The first step towards this goal was to design new plasmid constructs using the pGL4.10 [luc2] vector. Specifically, we inserted the CETP intron 1 fragment +191/+1056 bearing the most common or the less common allele downstream of the luciferase gene in the plasmid constructs (-1054/+27) CETP-luc and (-1054/+27) -656A CETP-luc described above. The constructs obtained at this step were the following: (-1054/+27) CETP-luc-(+191/+1056) INTRON 1 CETP, (-1054/+27) CETP-luc-(+191/+1056) +279A INTRON 1 CETP, (-1054/+27) -656A CETP-luc-(+191/+1056) INTRON 1 CETP and (-1054/+27) -656A CETP-luc-(+191/+1056) +279A INTRON 1 CETP. Transient transfection experiments were performed in HepG2 cells using the four different constructs described above, as well as (-1054/+27) CETP-luc and (-1054/+27) -656A CETP-luc, emply pGL4.10 [luc2] vector as control and beta-galactosidase expression vector for the normalization of luciferase levels in the reporter assays followed.

The results from the transient transfection experiments are shown in Figure 17 and the constructs reffering to these results in Figure 18. As it is demonstrated, the +191/+1056 CETP intron 1 fragment bearing the most common +279G allele decreased significantly the activity of the (-1054/+27) CETP promoter to 44% relative to the activity of the promoter alone (compare bars #2 and 4). Exactly the same

reduction was observed when the (+191/+1056) +279A allele was placed downstream of the (-1054/+27) CETP promoter (compare bars #2 and 5). However, when the +191/+1056 CETP intron 1 fragment bearing the most common +279G allele was tested in combination with the (-1054/+27) CETP promoter bearing the less common -656A allele, the activity of the CETP promoter was totally abolished (1% relative to the activity of the promoter alone, compare bars #4 and 6). Importantly, this strong reduction in promoter activity was not observed when the less common +279A allele of the intron 1 was used (compare bars #5 and 7). These results indicate, that apart from a statistical significant reduction in activity of the CETP promoter observed when the region of intron 1 is introduced to the plasmid constructs, a statistically significant decrease in promoter activity close to basal luciferase levels is observed, only in the case when there is a combination of the less common -656A allele with the most common +279G allele.

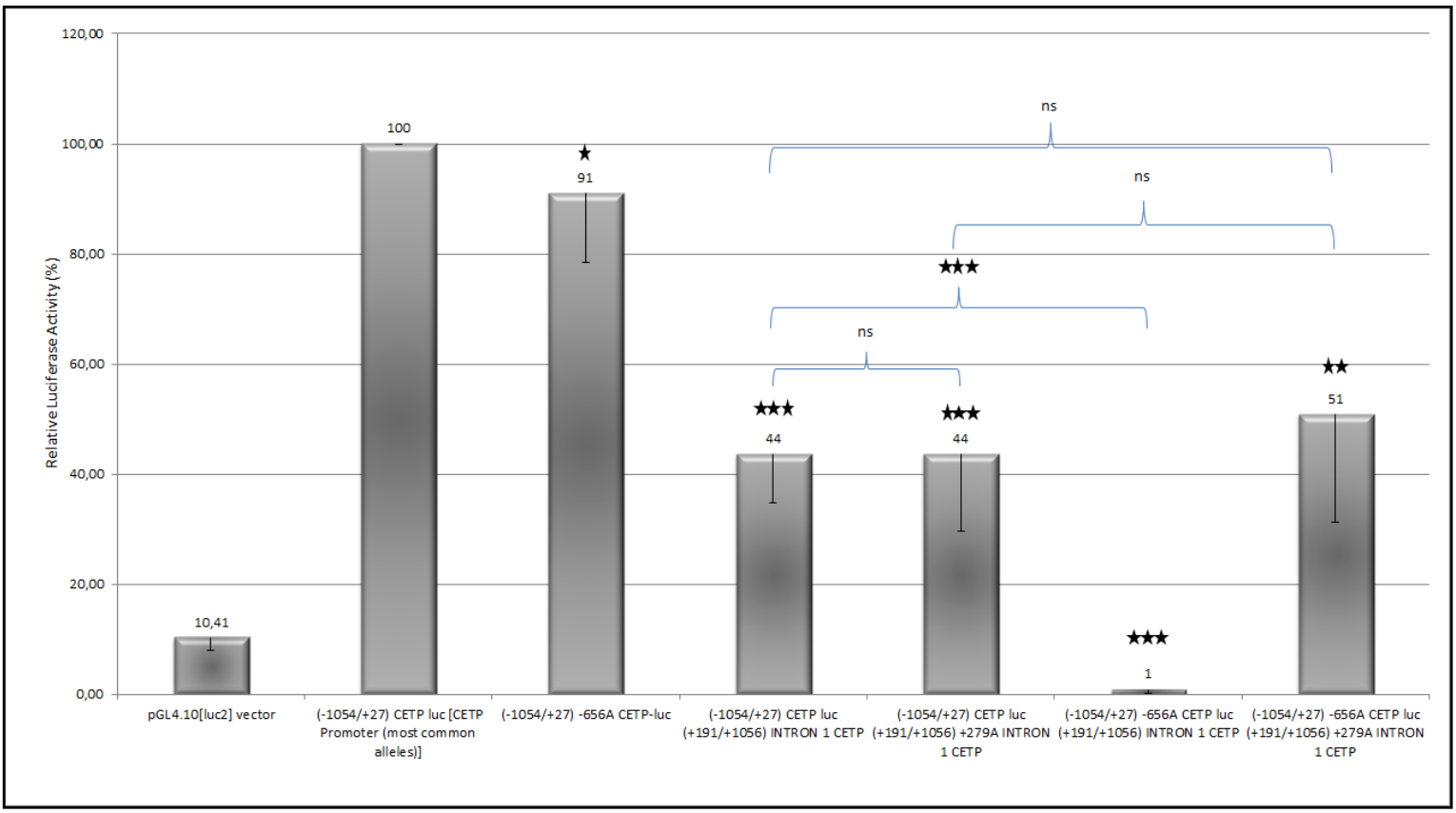

Figure 17. Presence of region +191/+1056 of CETP intron 1 significantly reduces the activity of the CETP promoter. HepG2 cells were transiently transfected with 5μgr of the luciferase reporter plasmids bearing the promoter region -1054/+27 of the CETP gene with the most common alleles, the promoter region -1054/+27 of the CETP gene bearing the less common -656A allele and with those bearing both the promoter region -1054/+27 and the region +191/+1056 of the first intron of human CETP gene at position *Sal*I, after the luciferase gene, for all different haplotypes between polymorphisms -656C>A and +279G>A, along with a beta-galactosidase expression vector (1μgr). From left to right, the CETP promoter constructs are the following: (-1054/+27) CETP-luc, (-1054/+27) -656A CETP-luc, (-1054/+27) CETP-luc-(+191/+1056) INTRON 1 CETP, (-1054/+27) CETP-luc-(+191/+1056) +279A INTRON 1 CETP, (-1054/+27) -656A CETP-luc-(+191/+1056) INTRON 1 CETP and (-1054/+27) -656A CETP-luc-(+191/+1056) +279A INTRON 1 CETP, whereas empty pGL4.10 [luc2] reporter vector was used as a control. It is observed that the presence of +191/+1056 region of CETP intron 1 bearing the most common allele significantly reduces the activity of the CETP promoter, when compared to the construct which lacks the intron 1 region (-1054/+27) CETP-luc (P<0.001). A significant decrease in CETP promoter activity close to basal luciferase levels is observed only in the case of A/G haplotype for polymorphism -656C>A and +279G>A, respectively. A statistical significance is observed for *p* value<0.053. Asterisks represent the statistical significance, * for *p* value<0.053, ** for *p* value<0.01, * for *p* value<0.001 and ns for non-significant.**

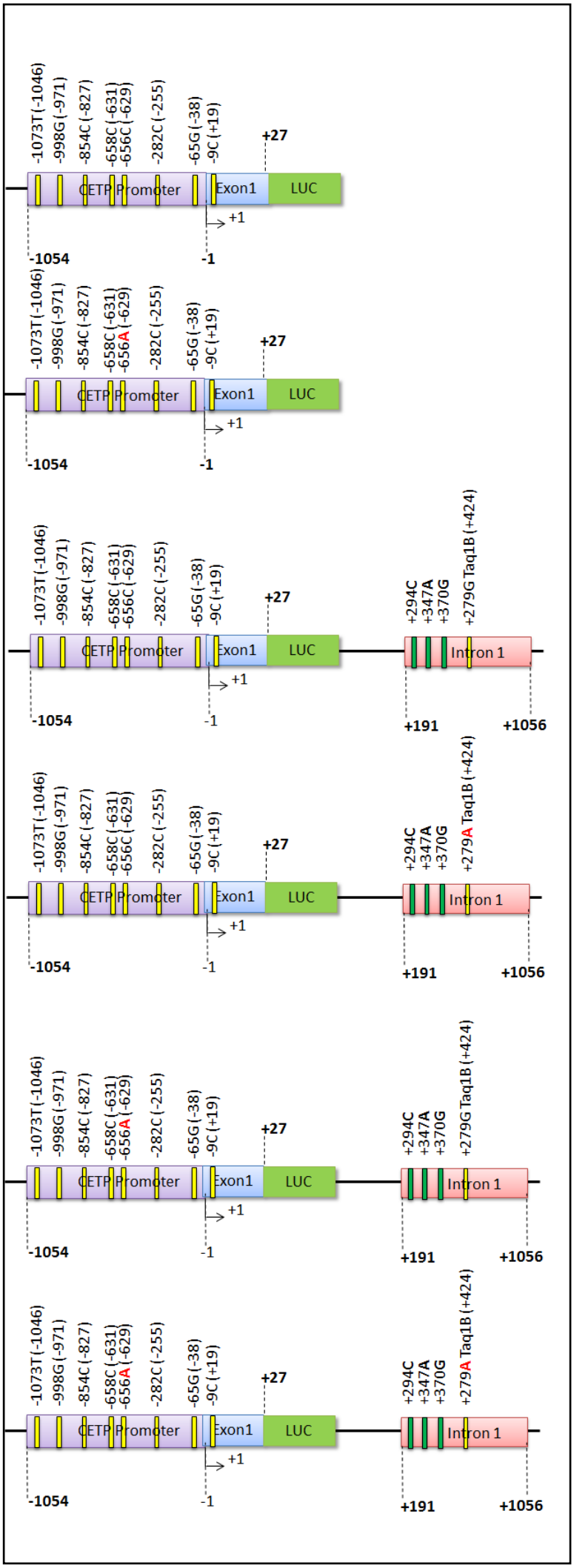

Figure 18. Constructs for CETP promoter (-1073T>C, -998G>A, -854C>T, -658C>A, -656C>A, -65G>A), first exon (-9C>T) and first intron (+279G>A) polymorphisms. -1054/+27 CETP gene region was inserted at positions *Sac*I and *Xho*I of the pGL4.10 [luc2] reporter vector, whereas +191/+1056 CETP Intron 1 region was inserted at position *Sal*I of the same vector. Site Directed Mutagenesis was performed in order to succeed the desired mutations. The CETP promoter region (-1054/-1) is depicted with purple, the exon 1 region (+1/+27) with blue, the luciferase gene with green, the CETP Intron 1 region (+191/+1056) with pink, the CETP polymorphisms with yellow, whereas with red font less common alleles are shown. Extra variants observed (+294C, +347A and +370G) in the region of the first intron of the CETP gene are depicted with dark green.

1.3.3. CETP intron 1-mediated decrease of CETP promoter activity is specific for the promoter of the CETP gene

In order to determine whether the transcriptional inhibitory activity of the +191/+1056 region of the first intron of the human CETP gene is specific for the promoter of the CETP gene, we designed a plasmid construct bearing the -668/+205 promoter region of human ABCA1 gene before the luciferase gene, and the region +191/+1056 of the first intron downstream of the luciferase gene. Specifically the plasmid construct we designed was the following: (-668/+205) ABCA1-luc-(+191/+1056) +279A INTRON 1 CETP. The plasmid construct (-668/+205) ABCA1-luc had been generated previously in our laboratory by Dr Effie Thymiakoy (Thymiakou, 2007).

Transient transfection experiments were performed in HepG2 cells, using the following plasmid constructs: (-1054/+27) CETP-luc, (-1054/+27) CETP-luc-(+191/+1056) INTRON 1 CETP, (-1054/+27) CETP-luc-(+191/+1056) +279A INTRON 1 CETP, (-668/+205) ABCA1-luc, (-668/+205) ABCA1-luc-(+191/+1056) +279A INTRON 1 CETP and beta-galactosidase expression vector for the normalization of luciferase levels. Relative luciferase activity as well as beta-galactosidase activity was evaluated in reporter assays. The results from transient transfection experiments are shown in Figure 19. As it is demonstrated, indeed the +191/+1056 region of CETP first intron has an inhibitory activity that is specific for the CETP promoter because it did not inhibit the activity of the heterologous -668/+205 human ABCA1 promoter. Specifically, we observe the same statistical significant reduction (P=0.001) as before in the CETP promoter activity, in constructs (-1054/+27) CETP-luc-(+191/+1056) INTRON 1 CETP, (-1054/+27) CETP-luc-(+191/+1056) +279A INTRON 1 CETP, when compared to the construct (-1054/+27) CETP-luc. However, this is not the case when we replaced the CETP promoter region [plasmid construct (-1054/+27) CETP-luc-(+191/+1056) +279A INTRON 1 CETP] with the promoter region of the ABCA1 gene [plasmid construct (-668/205) ABCA1-luc-(+191/+1056) +279A INTRON 1 CETP]. In fact, a slight increase, which was statistically significant) in the ABCA1 promoter activity was observed in the presence of the CETP intron 1 (compare bars #4 and 5). It is concluded that the CETP intron 1-mediated decrease in the activity of the CETP promoter is specific for the aforementioned promoter, indicating the possible existence of transcription factors interacting with each other, to regulate CETP transcription.

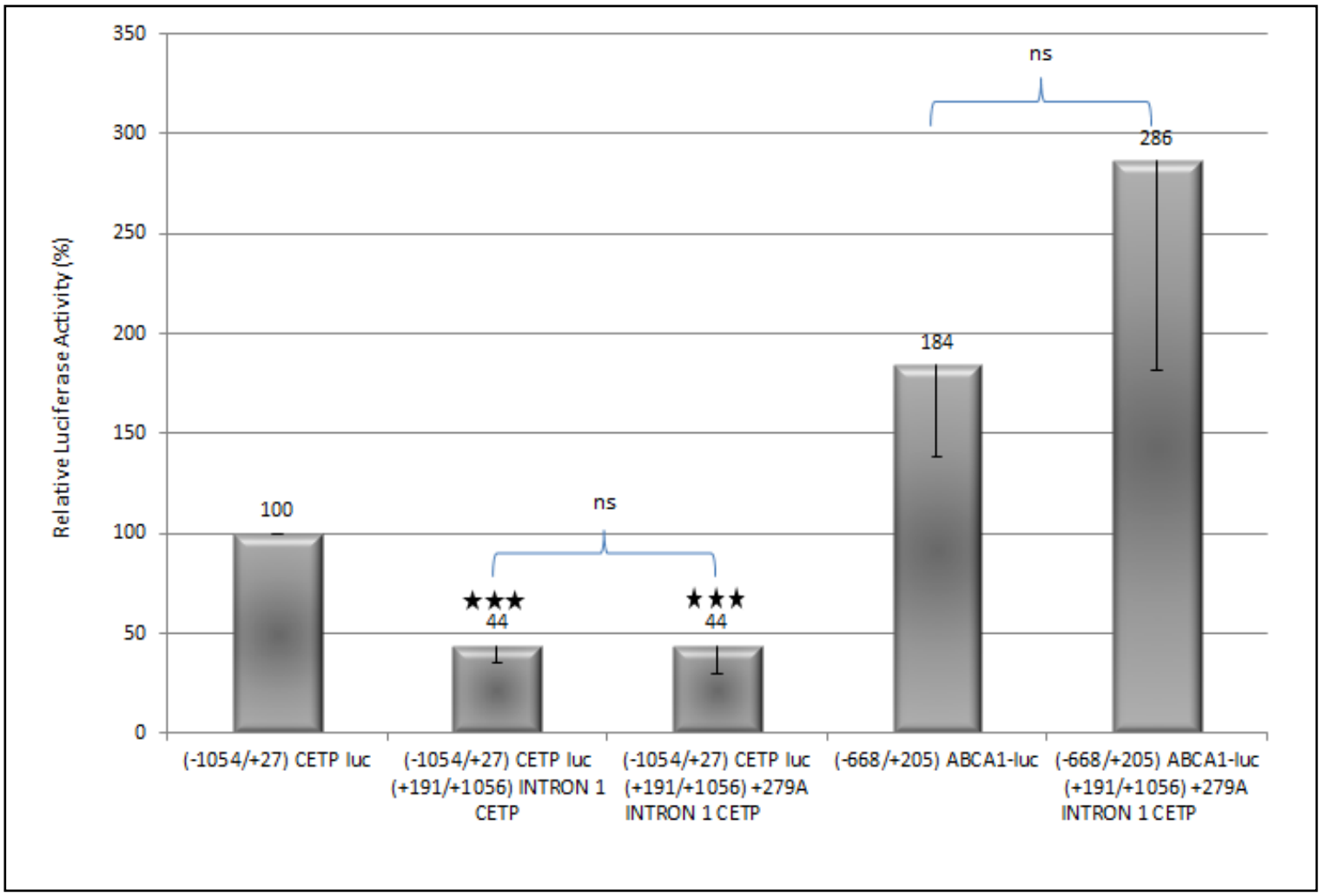

Figure 19. Reduction on the activity of the CETP promoter, mediated by the presence of +191/+1056 region of human CETP intron 1 is specific for CETP promoter. HepG2 cells were transiently transfected with 5μgr of the luciferase reporter plasmids: (-1054/+27) CETP-luc, (-1054/+27) CETP-luc-(+191/+1056) INTRON 1 CETP, (-1054/+27) CETP-luc-(+191/+1056) +279A INTRON 1 CETP, (-668/+205) ABCA1-luc and (-668/+205) ABCA1-luc-(+191/+1056) +279A INTRON 1 CETP, along with a beta-galactosidase expression vector (1μgr). It is observed that the presence of +191/+1056 region of CETP intron 1 bearing either the most common or the less common allele significantly reduces the activity of the CETP promoter, when compared to the construct which lacks the intron 1 region (-1054/+27) CETP-luc (P<0.001). However, upon introduction of +191/+1056 human CETP intron 1 region bearing the less common +279A allele, to a plasmid construction bearing -668/+205 human ABCA1 promoter region, the intron 1 mediated decrease in promoter activity is being abolished, incicating that CETP intron 1 mediated regulation in transcription is specific for CETP promoter and not for other gene promoters. A statistical significance is observed for *p* value<0.053. Asterisks represent the statistical significance, * for *p* value<0.053, ** for *p* value<0.01, * for *p* value<0.001 and ns for non-significant.**

1.3.4. Bioinformatics analysis of region -1070/+1090 of the CETP gene revealed potential transcription factor binding sites

Bioinformatics analysis was performed, using the latest TRANSFAC version (2014.4, Biobase, http://www.gene-regulation.com). The CETP sequence (-1070/+1090) bearing either the most common or the less common alleles was screened and several potential transcription factor binding sites were demonstrated for all the CETP polymorphisms examined (Figure 20). The results, showing the most significant potential sites for transcription factors, are shown in Tables 10 to 20 in the Appendix. Significance was evaluated according to the values in the Difference column shown in tables, with the most significant scores being those with the greater Difference value.

Specifically, for CETP promoter polymorphism -656C>A, putative sites for the following transcription factors were found: RAD21, HNF4A and CP2. For CETP promoter polymorphism -65G>A, putative sites for the following transcription factors were identified: ZNF555, IRXB3, GATA4, POU2F3 and POU3F3. In the case of the TaqIB polymorphism of CETP first intron, putative sites for MOVOB, POU4F3, EHF, MZF1 and GLIS2 transcription factors were identified.

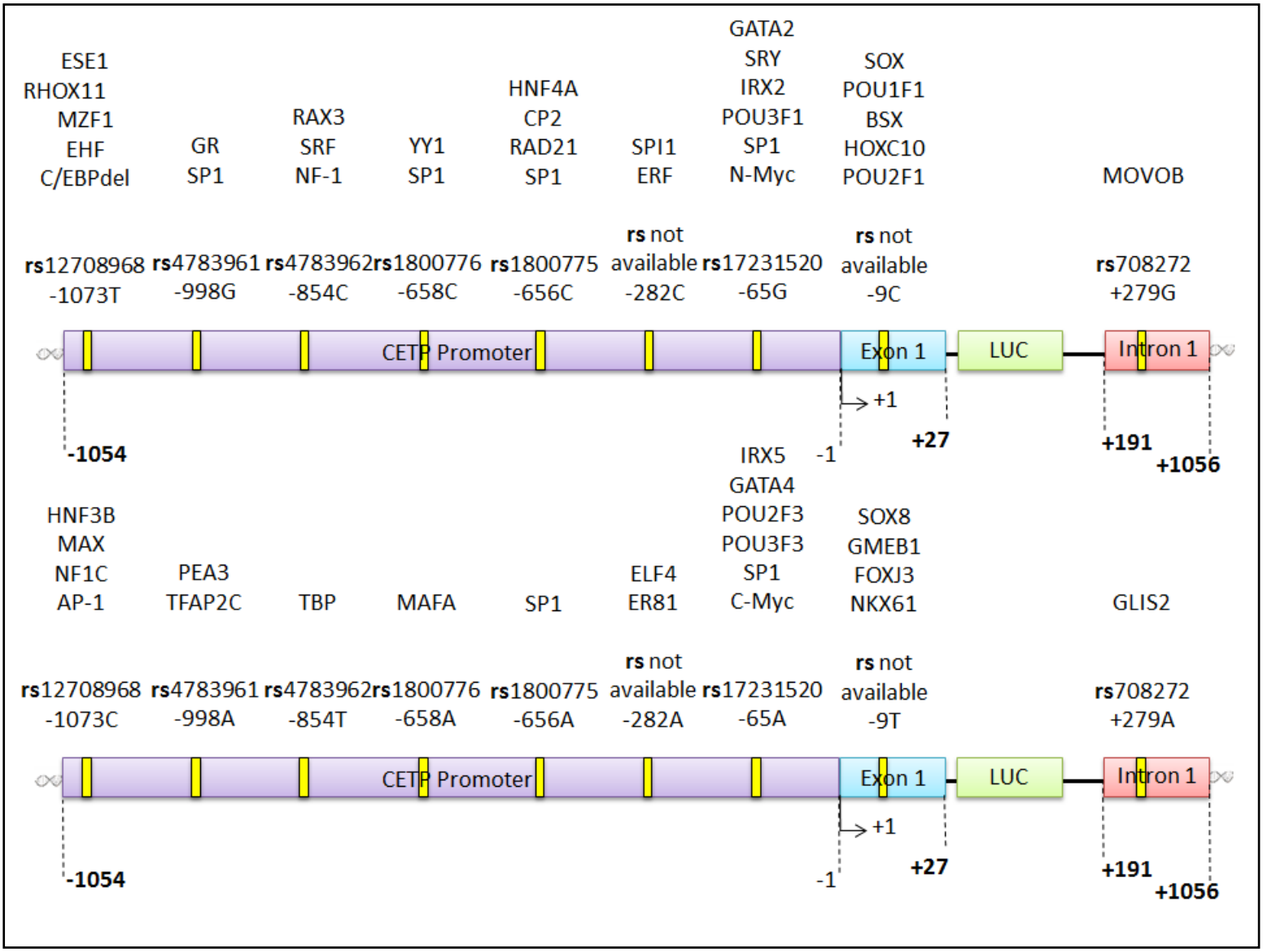

Figure 20. Bioinformatics analysis using the latest TRANSFAC version (2014.4, Biobase, http://www.gene-regulation.com) and the online program AliBaba2.1 from www.gene-regulation.com. Potential transcription factor binding sites are shown, both for most common and for less common alleles for CETP promoter region, first exon and first intron. The CETP promoter region (-1054/-1) is depicted with purple, the exon 1 region (+1/+27) with blue, the luciferase gene with green, the CETP Intron 1 region (+191/+1056) with pink and the CETP polymorphisms with yellow.

1.3.5. Polymorphism -427T>C upstream of +1 transcription start site in the human apoE gene down-regulates the apoE promoter activity

In order to examine, if the polymorphisms observed upstream the +1 transcription start site of the apoE gene in the CCHS can regulate the activity of the apoE promoter, we used a plasmid bearing the -486/+89 human apoE promoter with the most common -491A, -427T and -219G alleles at each polymorphic site under study, which already existed in our laboratory, as a template and performed Site-Directed Mutagenesis in order to generate apoE promoter variants bearing the less common alleles -427C or -219T. Those constructs are: (-486/+89) apoE-luc, (-486/+89) -427C apoE-luc and (-486/+89) -219T apoE-luc. The purpose of these plasmid constructs was to use them in transient transfection assays, so as to examine whether the presence of a less common allele at a specific site regulates human apoE promoter activity. For this purpose, we transiently transfected HepG2 cells with the different apoE promoter plasmids along with a beta-galactosidase expression vector and we evaluated the relative luciferase activity (%) in each case.

The results from the reporter assays are shown in Figure 21 and the constructs reffering to these results in Figure 22. As it is demonstrated, the presence of the less common C allele at position -411 (corresponding to -427T>C polymorphism), significantly reduced the activity of the apoE promoter to 89% relative to the activity of the promoter bearing the most common T allele (compare bars #2 and 3). However, no statistically significant change is observed in the presence of the less common allele T at position -203 (corresponding to -219G>T polymorphism) indicating that this site does not regulate apoE promoter activity.

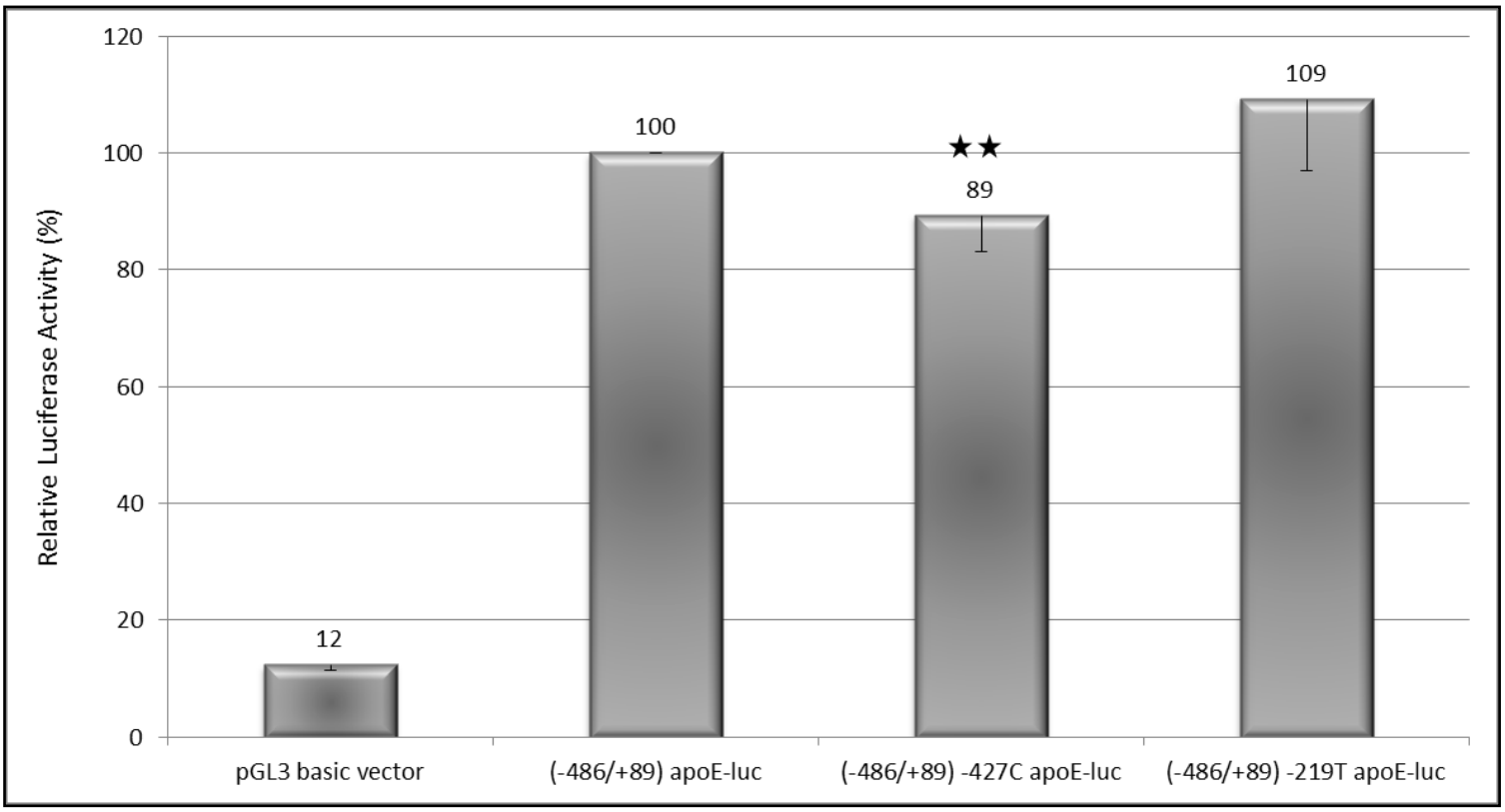

Figure 21. Polymorphisms -427T>C upstream of +1 transcription start site of the human apoE gene down-regulate the apoE promoter activity. HepG2 cells were transiently transfected with the luciferase reporter plasmid bearing the promoter region -486/+89 of the apoE gene with the most common alleles (1μgr) for polymorphisms -491A>T, -427T>C and -219G>T and luciferase reporter plasmids (1μgr) bearing the less common -427C or -219T alleles, along with a beta-galactosidase expression vector (1μgr). From left to right, the apoE promoter constructs are the following: (-486/+89) apoE-luc, (-486/+89) -427C apoE-luc and (-486/+89) -219T apoE-luc, whereas empty pGL3 basic reporter vector was used as a control. It is observed that the presence of the less common -427C allele reduces the activity of the apoE promoter in a statistically significant manner (P<0.01). A statistical significance is observed for *p* value<0.053. Asterisks represent the statistical significance, * for *p* value<0.053, ** for *p* value<0.01, * for *p* value<0.001 and ns for non-significant.**

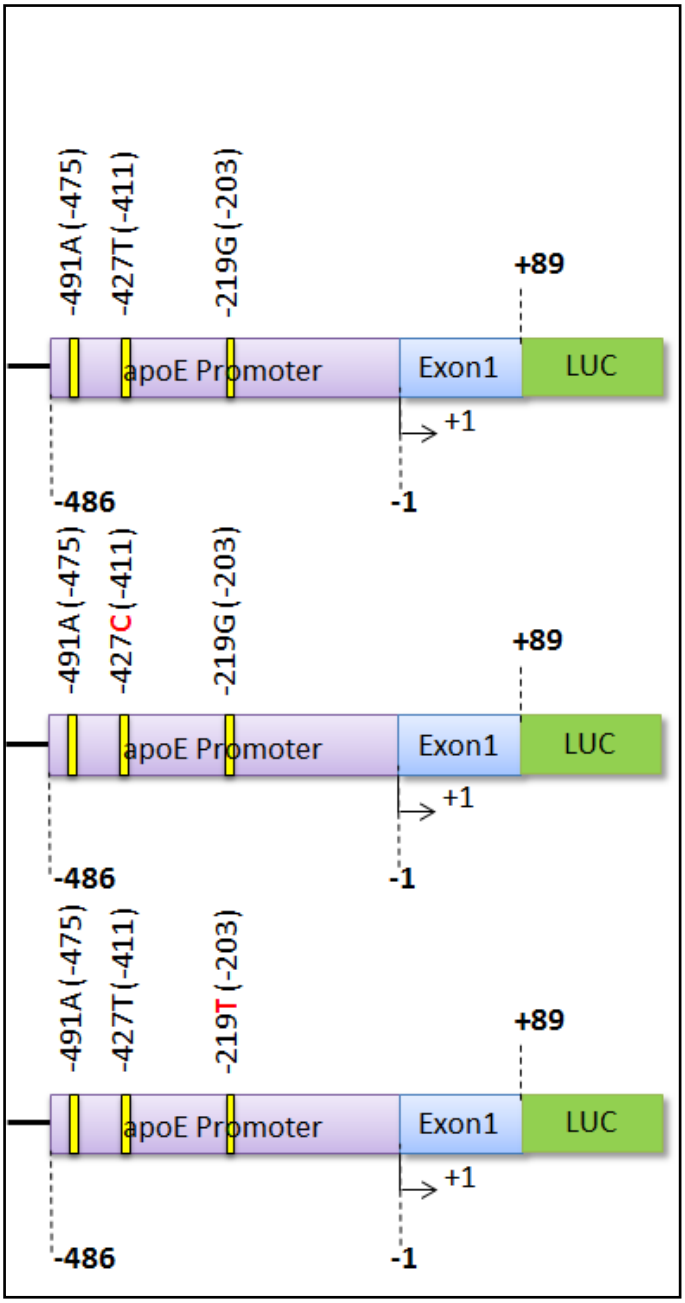

Figure 22. Constructs for apoE promoter (-427T>C, -219G>T) polymorphisms. -486/+89 apoE gene region was inserted at positions *KpnI* and *SacI* of the pGL3 basic reporter vector. Site Directed Mutagenesis was performed in order to succeed the desired mutations. The apoE promoter region (-486/-1) is depicted with purple, the exon 1 region (+1/+89) with blue, the luciferase gene with green, the apoE promoter polymorphisms with yellow, whereas with red font less common alleles are shown.

1.3.6. Bioinformatics analysis of region -560/+160 of the apoE gene revealed potential transcription factor binding sites

Bioinformatics analysis was performed, using the latest TRANSFAC version (2014.4) as above. ApoE sequence -560/+160 bearing either the most common or less common alleles was screened and several potential transcription factor binding sites were demonstrated for all the apoE polymorphisms examined (Figure 23). The results, showing the most significant potential sites for transcription factors, are available in Tables 21 to 23 in the Appendix. Significance was evaluated according to the values in the Difference column shown in tables, with the most significant scores being those with the greater Difference value.

Specifically, regarding the position -411 where -427T>C polymorphism is located in the apoE gene promoter region, several potential transcription factor binding sites were identified. Among the most significant were: NKX2B, AML1, KLF3, PLAG1, SP1, ZFP770 and MZF1.

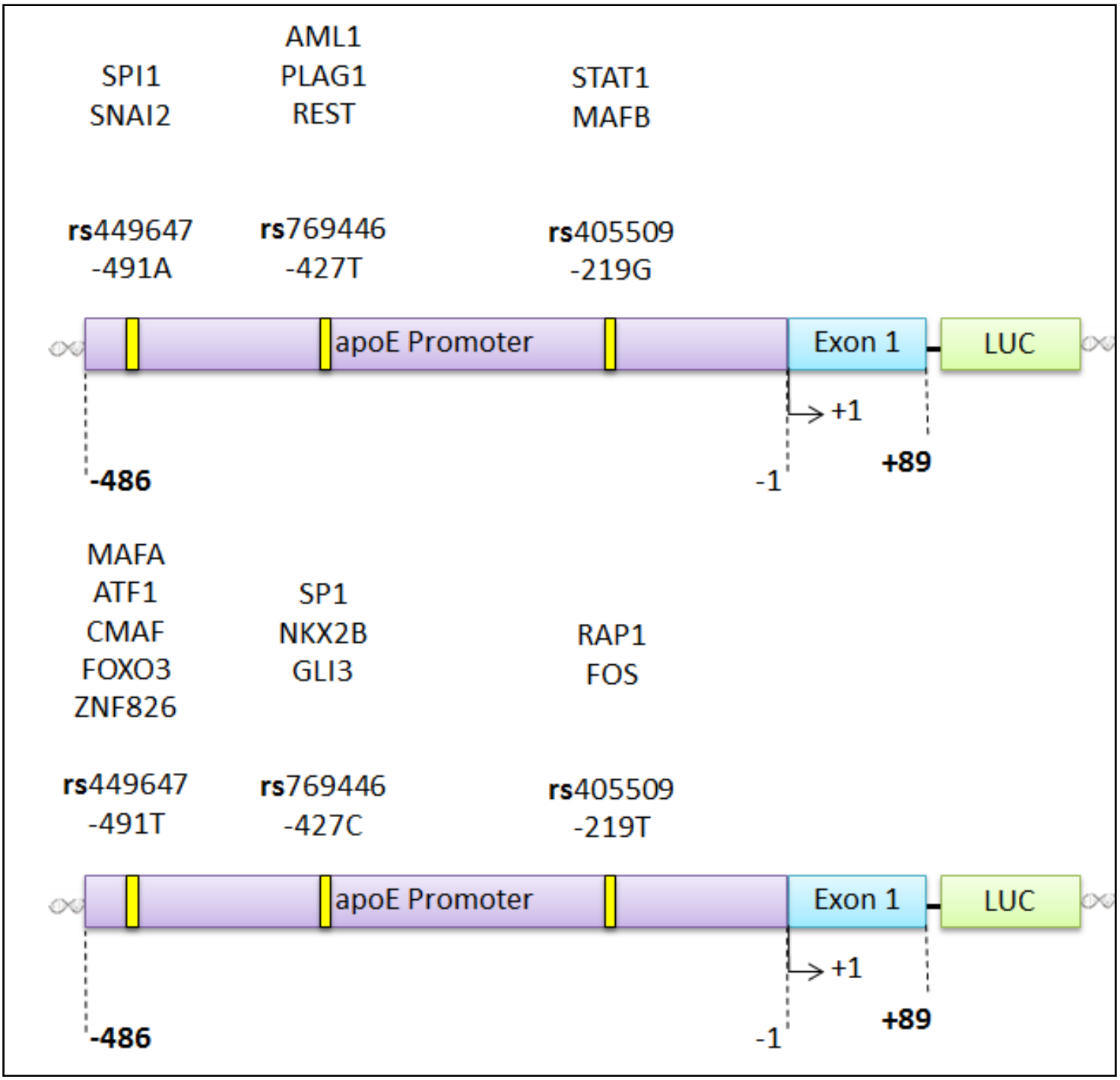

Figure 23. Bioinformatics analysis using the latest TRANSFAC version (2014.4, Biobase, http://www.gene-regulation.com) and the online program AliBaba2.1 from www.gene-regulation.com. Potential transcription factor binding sites are shown, both for most common and for less common alleles for apoE promoter region. The apoE promoter region (-486/-1) is depicted with purple, the exon 1 region (+1/+89) with blue, the luciferase gene with green and the apoE polymorphisms with yellow.

1.4. DISCUSSION

ApoE plays a significant role in lipid metabolism and it appears to possess a dual function. Not only it contributes in the clearance of triacylglycerol-rich lipoproteins, but it also participates in the biogenesis of apoE-containing HDL partricles (Gafencu *et al.*, 2007; Kypreos *et al.*, 2007), a process which is well characterized in experiments using adenovirus mediated gene transfer of apoE in apoA-I-deficient or in ABCA1-deficient mice. ApoE enhances the lipid exchange between lipoproteins, mediated by CETP, thus alterations in CETP activity contribute to changes in HDL composition and size (Korhonen *et al.*, 1999; Ordovas *et al.*, 2000).

CETP is a molecule of significant importance, participating in the RCT pathway, thus mediating the transfer of esterified HDL cholesterol towards apoB-containing lipoproteins in exchange for triglycerides, contributing in this way to the clearance of apoB-containing lipoproteins by the LDLr in the liver (Besler *et al.*, 2012). HDL is considered to be an anti-atherogenic lipoprotein, with low plasma HDL cholesterol levels to be associated with increased risk of myocardial infarction and coronary disease. However, many studies over the years, have demonstrated the connection of genetic abnormalities or polymorphisms in the CETP gene, with disturbances in levels of lipoproteins in plasma, thus rendering CETP as a candidate gene for coronary heart disease, implicated in metabolic syndrome and type 2 diabetes mellitus (Corbex *et al.*, 2000; Barter *et al.*, 2003; Boekholdt *et al.*, 2004; de Vries *et al.*, 2005; Zeller *et al.*, 2007; Masson *et al.*, 2009; Ridker *et al.*, 2009). Since both apoE and CETP contribute to the reverse cholesterol transport, the effects of variations at the CETP locus may very well interact with the apoE genotype (Sorlí *et al.*, 2006).

The CCHS revealed seven polymorphisms in the CETP promoter region (-1073T>C, -998G>A, -854C>T, -658C>A, -656C>A, -282C>A and -65G>A), one polymorphism in the region spanning the first exon of the CETP gene (-9C>T) and one polymorphism in the region of the first intron of the CETP gene (TaqIB INS1 +279G>A). Regarding the apoE gene in the CCHS, three different variants were identified in the promoter region of the corresponding gene (-491A>T, -427T>C and -219G>T. So our aim was to investigate whether these polymorphic sites present in the apoE and CETP promoters are functional and are able to up- or down-regulate the transcription of the corresponding genes.

Our results, regarding the polymorphisms spanning the promoter region of the CETP gene, clearly demonstrate that two different variants are able to regulate the activity of the corresponding promoter in transient transfection experiments performed in HepG2 cells. Specifically, the less common -656A allele of the polymorphism -656C>A of the CCHS, which is located at position -629 upstream of +1 transcription start site, was shown to significantly reduce the activity of the corresponding promoter in reporter assays, compared with the most common -656C

allele. This is in agreement with previous studies from Dachet *et al.* and Le Goff *et al.*, in which it was demonstrated in transient transfection experiments, that the less common A allele at position -629 of the CETP promoter significantly decreases the activity of the promoter. Such repression, was demonstrated to be due to the binding of both Sp1 and Sp3 transcription factors in the presence of the less common A allele, and not in the case where the most common C allele was present, leading to significant reduction in *in vitro* transcriptional activity (Dachet *et al.*, 2000; Le Goff *et al.*, 2003a). Indeed, bioinformatics analysis using the online program AliBaba2.1 from www.gene-regulation.com, revealed that Sp1 transcription factor binds to the region -622 to -633 upstream +1 transcription start site in the CETP gene (Figure 20). Thus we hypothesize that perhaps this is the case in our experiments as well, so the reduction observed in the promoter activity could be due to binding of Sp1 to the less common A allele, thus regulating transcription in a negative manner. Sp1 and Sp3 transcription factors share 95% similarity in their zinc-finger DNA binding domain, and can efficiently bind with identical affinity to both the GC-box (GGGGCGGGG) and GT-box (GGTGTGGGG) (Suske, 1999). However, the physiological roles of Sp1 and Sp3 appear to be quite different. Sp1 is a transcriptional activator (Suske, 1999). Sp3 has been shown to positively regulate transcription (Pagliuca et al., 2000), but it can also act as a repressor (Ammanamanchi *et al.*, 2001). Several studies using mammalian cells have indicated that Sp3 protein is modified by phosphorylation (Ge *et al.*, 2002; Bakovic *et al.*, 2003), sumoylation (Ross *et al.*, 2002; Sapetschnig *et al.*, 2002) and acetylation (Braun *et al.*, 2001; Ammanamanchi *et al.*, 2003), and that these modifications influence the role that Sp3 plays in transcription. It is also reported that the nucleotides adjacent to the GC-box and the GT-box also influence the role of Sp3 in transcription (Suske, 1999). Sp1/Sp3 sites have been shown to be involved in regulating lipid-transfer genes such as LCAT (Hoppe *et al.*, 1998), LPL (Yang *et al.*, 1998), and phospholipid transfer protein (Tu *et al.*, 2001) as well as numerous lipid-related and other genes. Moreover, bioinformatics analysis revealed RAD21, HNF-4α and CP2 transcription factors, to be among the most significant for binding to the specific polymorphic position in the promoter (Figure 20) (Appendix, Table 14). Perhaps, a mechanism exists, involving binding of one of the aforementioned factors, which could regulate transcription mediated by CETP promoter. TRANSFAC analysis revealed, that HNF-4α binds to the most common -656C allele with greater affinity compared to the less common -656A allele, so perhaps the down-regulation in CETP promoter activity observed in the presence of the less common -656A allele, could be owing to lack of a HNF-4α binding at the specific position.

The second variant in the CETP promoter region of the CCHS, which was found to significantly up-regulate transcriptional activity of the corresponding promoter, is the less common -65A allele at position -38 upstream +1 transcription

start site. Specifically, it was demonstrated in reporter assays, that the presence of the less common -65A allele significantly increased the activity of the CETP promoter when compared with the plasmid construction bearing the most common -65G allele at the same position. In 1995 a study by Gaudet *et al.* revealed the existence on a Sp1 binding site spanning the region -26 to -57 upstream +1 transcription start site of the CETP promoter (Gaudet *et al.*, 1995). This region, involves the promoter polymorphism -65G>A at position -38. It was demonstrated that Sp1 can bind at position -37 to -42 bearing the most common promoter sequence, thus up-regulating the promoter activity, however mutations at this site at positions -40 and -41 abolished the binding of Sp1 leading to reduced CETP promoter activity (Gaudet *et al.*, 1995; Le Goff *et al.*, 2003a). Indeed, bioinformatics analysis using the online program AliBaba2.1 from www.gene-regulation.com, revealed two Sp1 sites at positions -37 to -49 and -40 to -50, both including position -38 of CETP polymorphism -65G>A. However, the program revealed no change in Sp1 binding sites, when we substituted the most common -65G to the less common -65A allele (Figure 20), implying that Sp1 binding at the specific positions is independent of the variant. Indeed, Gaudet *et al.* in 1995, performed site directed mutagenesis at positions -40 and -41 in the CETP promoter; however the -38 position remained untouched in the experiments performed (Gaudet *et al.*, 1995). So we performed bioinformatics TRANSFAC analysis which revealed ZNF555, IRXB3, GATA4, POU2F3 and HOXA7 transcription factors, to be among the most significant for binding to the specific polymorphic position in the promoter (Appendix, Table 16). However, several other factors have been identified, which bind to the same position, although at a less significant degree (Appendix, Table 16), so perhaps another mechanism exists, apart from the binding of Sp1, which up-regulates the activity of CETP promoter in the presence of the less common -65A allele.

Future experiments, regarding the -656C>A polymorphism in the CETP promoter, could involve Chromatin Immunoprecipitation (ChIP) assays in order to determine whether Sp1 and Sp3 transcription factors indeed binds at position -629 of the human CETP promoter in the presence of the less common A allele, and not in the presence of the most common C allele, so as to elucidate the specificity of the binding. Moreover, since TRANSFAC analysis identified an HNF-4α binding site in the presence of the most common C allele, but not in the presence of the less common A allele, it is critical to elucidate whether this binding actually exists, performing again ChIP assays. In addition, DNA Affinity Precipitation (DNAP) assay would be another valuable tool, to demonstrate the binding of Sp1, Sp3 and HNF-4α to the specific DNA sequence, in order to validate our results. Regarding the polymorphism -65G>A at position -38 in the CETP promoter, since we excluded the possibility that Sp1 is the factor responsible for the up-regulation in promoter activity observed, we need to show which of the putative ZNF555, IRXB3, GATA4, POU2F3 and HOXA7

transcription factors revealed by the TRANSFAC analysis is the most important for CETP regulation.

Regarding the promoter polymorphisms -1073T>C, -998G>A, -854C>T and -658C>A, as well as polymorphism -9C>T in the first exon of the human CETP gene, no significant change in promoter activity was observed between most common and less common alleles, for each case. However a study in 2003 revealed that under the presence of the most common -1073T allele of polymorphism -1073T>C at position -1046, CPF transcription factor is able to bind and mediate the activation of the CETP promoter (Le Goff *et al.*, 2003b). However, in our experiments, we did not observe any difference in promoter activity between the two alleles at position -1046. In 2005, a study by Frisdal *et al.*, demonstrated that the less common allele A, for polymorphism -971G>A, displayed a significant lower *in vitro* promoter activity compared to the more frequent allele, indicating that the polymorphism is functional (Frisdal *et al.*, 2005). However, this was not the case in our experiments. Although we observed a slight tend towards increased promoter activity in the presence of the less common -998A allele compared to the most common -998G allele at position -971 upstream of +1 transcription start site in the CETP gene, this observation was not able to reach significance. This is in agreement with transient transfection experiments performed by Le Goff *et al.*, in 2002, demonstrating that -971G>A polymorphism did not modulate transcriptional activity of the human CETP gene promoter, therefore constituting a non-functional polymorphism (Le Goff *et al.*, 2002).

Regarding the CETP polymorphism TaqIB (IVS1 +279G>A) which is located at position +424 downstream +1 transcription start site, in a region spanning the first intron of human CETP gene, there are several studies which associated the presence or absence of both alleles at position +424 with disease phenotypes. Owning to its location in the first intron, it was assumed that the TaqIB polymorphism does not represent a functional regulatory site but a marker for other functional sites (Frisdal *el al.*, 2005). Moreover, its effects on HDL cholesterol and plasma CETP levels are independent (Fumeron *et al.*, 1995; Bernard *et al.*, 1998), which suggests that the TaqIB variant is a marker for at least two functional polymorphisms (Frisdal *el al.*, 2005). Indeed, several studies showed that a polymorphism at position -629 in the upstream region of the CETP gene promoter can modulate the transcriptional activity of the CETP gene in *in vitro* experiments (Dachet *et al.*, 2000). In our experiments, we observed a significant decrease in the promoter (bearing the most common alleles for CETP promoter polymorphisms) activity, when we introduced in the plasmid construct the region +191/+1056 of the first intron of human CETP gene. Based on these findings, we propose that there are transcription factors binding to the region of the first intron of CETP, which interact with the promoter region, in a manner so that to down-regulate the promoter activity. The proposed mechanism is

shown in Figure 24. (+191/+1056) CETP first intron region can effectively bind transcription factors (Appendix, Tables 19 and 20), independently of the variant +279G>A, which are able to interact with the (-1054/+27) CETP promoter region and significantly down-regulate transcriptional levels of the corresponding gene (Figure 24 C, D and F). This observed decrease in promoter activity, when the (+191/+1056) CETP first intron region is introduced to the plasmid constructs bearing the promoter region, ranges from 44% to 51% (Figure 17), compared to the activity of the promoter alone. However, when the less common -656A allele in the CETP promoter region is introduced, in combination with the the most common +279G allele in the first intron region, an acute decrease in promoter activity is observed close to 1%, when compared to the activity of the promoter alone (Figure 17). This is assumed, to be due to a specific transcription factor, which even though is able to bind in the CETP first intron region, when the most common +279G allele is present (Appendix, Table 18); it can only interact and drastically eliminate the CETP promoter activity, when the less common -656A allele is present in the promoter region (Figure 24 E). So when the most common -656C allele is introduced in the promoter region (Figure 24 C), or when the less common +279A allele is introduced in the first intron region (Figure 24 F), comparing to the case when -656A/+279G haplotype is present (Figure 24 E), this specific interaction mediated by the transcription factor is abolished, thus only an almost to half reduction of promoter activity (comparing to the promoter bearing the most common alleles alone) is observed.

Several studies have associated TaqIB polymorphism alleles with specific phenotypes. It has been reported that the presence of the most common +279G (B1) allele, is associated with lower HDL cholesterol and higher prevalence of coronary heart disease (Ayyobi *et al.*, 2005; Porchay-Baldèrelli *et al.*, 2007), whereas the presence of the less common +279A (B2) allele is significantly associated with both low plasma CETP activity and elevated levels of plasma HDL cholesterol and apoA-I (Akbarzadeh *et al.*, 2012; Lu *et al.*, 2013). In 2013, a study by Gautier *et al.* revealed the existence of an ER8 element in the first intron of CETP gene at position +317 to +336, which up-regulates CETP activity in response to FXR stimulation in transient co-transfection experiments in HepG2 cells (Gautier *et al.*, 2013). However, in our experiments, we observed that the presence of the first intron significantly down-regulates the promoter activity. Bioinformatics analysis was performed, using the latest TRANSFAC version (2014.4), which revealed a great number of transcription factors, which can bind to the region +191 to +1056 of the first intron of human CETP gene. Among the most significant were RXRRAR, RARA, IRX4, IRX5, IRF6, HOXA3, SOX8, NR2F1, PR and ETS1 (Appendix, Tables 19 and 20). So, it cannot be excluded, that specific transcription factors, which bind in the region of the first intron of human CETP gene, could interact with factors binding in the promoter region of the CETP gene, and through interacting with each other, with a

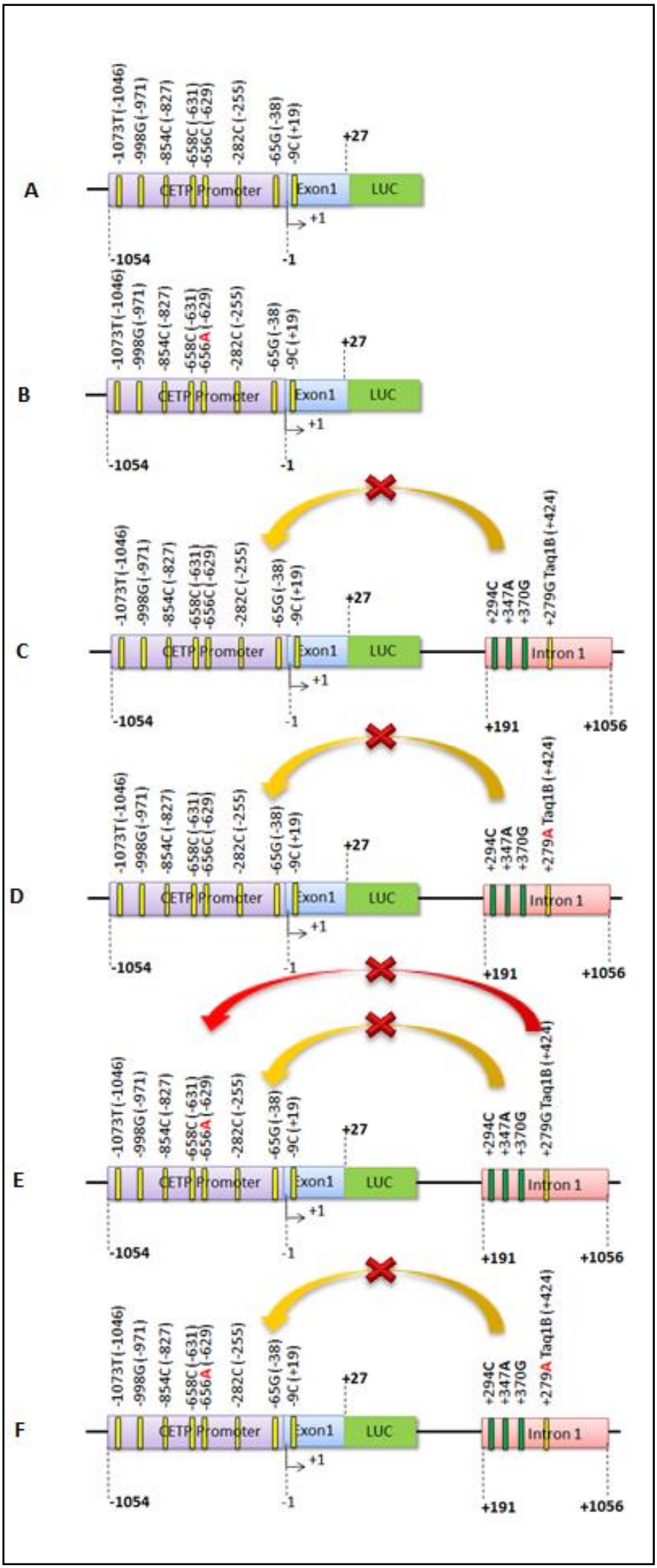

Figure 24. Proposed model for CETP first intron mediated down-regulation of CETP promoter activity. A) CETP promoter and first exon region bearing the most common alleles, B) CETP promoter and first exon region bearing the less common -656A allele at position -629 upstream +1 transcription start site, C) CETP promoter and first exon region bearing the most common alleles ,upstream the luciferase gene and CETP intron 1 region bearing the most common allele, downstream the luciferase gene, D) CETP promoter and first exon region bearing the most common alleles, upstream the luciferase gene and CETP intron 1 region bearing the less common +279A allele at position +424, downstream the luciferase gene, E) CETP promoter and first exon region bearing the less common -656A allele at position -629, upstream the luciferase gene and CETP intron 1 region bearing the most common allele downstream the luciferase gene, F) CETP promoter and first exon region bearing the less common -656A allele at position -629, upstream the luciferase gene and CETP intron 1 region bearing the less common +279A allele, downstream the luciferase gene. The CETP promoter region (-1054/-1) is depicted with purple, the exon 1 region (+1/+27) with blue, the luciferase gene with green, the CETP Intron 1 region (+191/+1056) with pink, the CETP polymorphisms with yellow, whereas with red font less common alleles are shown. Extra variants observed in the region of the first intron of the CETP gene are depicted with dark green. Yellow arrows demonstrate the specific interaction of CETP intron 1 with CETP promoter region mediated by transcription factors, which leads to an almost to half down-regulation of the promoter's activity, compared to the activity of the promoter alone. The red arrow demonstrates the specific interaction between the less common -656A allele in the promoter region and the most common +279G allele in the first intron of the CETP gene mediated by the binding of a specific transcription factors, which leads to a significant down-regulation of the promoter's activity, close to basal luciferase levels.

mechanism yet unknown, can down-regulate the activity of the corresponding promoter. The same bioinformatics analysis for the region +424 of the first intron revealed that MOVOB, POU4F3, EHF, MZF1 and GLIS2 (Figure 20) (Appendix, Table 18), are among the most significant transcription factors, binding to the specific position. Although the mechanisms of this reduction, in promoter activity by the first intron of CETP, are unknown to us, it is critical to elucidate which transcription factors bind efficiently and regulate transcriptional levels, since CETP comprises a significant pharmacological target.

Future experiments, would involve perhaps mutagenesis of the three extra variants observed in the first intron region, to their most common alleles, in order to reinforce our observations. Moreover, since we observe an overall down-regulation in the CETP promoter activity when we introduce the region of the first intron of CETP, a deletion analysis of the +191/+1056 intron 1 region would identify the position of the potential transcription factor binding site(s). In a second step, we can utilize the results provided by TRANSFAC analysis for the specific region, so as to identify potential transcription binding factors. The binding of the most significant factors from the TRANFAC analysis can be examined either by ChIP or by DNAP assays, in order to determine whether a specific binding actually exists in our sequence. If this is the case, we will perform Site-Directed Mutagenesis at the specific site, so as to verify that under the mutated site present, the binding is abolished. For the +424 TaqIB polymorphism, since we hypothesize that a specific transcription factor binds in the presence of the most common +279G allele and interacts with the less common -656A allele in the promoter, we can utilize the data provided from the TRANSFAC analysis and through ChIP and DNAP assays examine the specific binding of the potential transcription factor at position +424 in the presence of the most common +279G allele and at position -629 of the CETP promoter in the presence of the less common -656A allele.

Our results, regarding the polymorphisms spanning the promoter region of the apoE gene, clearly demonstrate that one variant is able to regulate the activity of the corresponding promoter in transient transfection experiments performed in HepG2 cells. Specifically, the less common -427C allele of the polymorphism -427T>C of the CCHS, which is located at position -411 upstream of +1 transcription start site, was shown to significantly reduce the activity of the corresponding promoter in reporter assays, compared with the condition, where the most common -427T allele is present. This observation is in disagreement with transient transfection experiments performed by Artiga *et al.* in 1998, where it was demonstrated, that no significant difference was observed between the two alleles (Artiga *et al.*, 1998). This could be attributed to the heterogeneity of reporter assay experiments as well as to the precision and reproducibility of the transient transfection experiments. Bioinformatics analysis was performed, using the latest

TRANSFAC version (2014.4), and it was revealed that NKX2B, AML1, KLF3, SP1, PLAG1, ZFP770 and MZF1 are among the most significant transcription factors identified (Figure 23) (Appendix, Table 22). Moreover, bioinformatics analysis using the online program AliBaba2.1 from www.gene-regulation.com, revealed two transcription binding factors, which are able to bind in the presence of the less common -427C allele at position -411. Those are Sp1, which binds to the region -401 to -412 upstream +1 transcription start site and GLI3, which binds to the region -403 to -412 in the apoE promoter region (Figure 23). Importantly, we observe that Sp1 has been identified in both bioinformatics analyses, and as shown in Appendix Table 22, Sp1 binds to the less common -427C allele with greater affinity (which is revealed by the greater value in Mut-score, comparing to the WT-score), which implies that perhaps Sp1 is the transcription factor responsible for the down-regulation observed in apoE promoter activity, in the presence of the less common -427C allele. However, there is no study, so far, reporting a mechanism regulating the transcription of apoE promoter, through binding of a transcription factor, when the less common -427C allele is present at position -411 in the apoE promoter.

Future experiments, would involve ChIP assays in order to determine whether Sp1 transcription factor indeed binds at position -411 of the apoE promoter in the presence of the less common C allele, and not in the presence of the most common T allele, so as to elucidate the specificity of the binding. In addition, DNAP assay would be another valuable tool, to demonstrate the binding of Sp1 to the specific DNA sequence, in order to reinforce our results.

Regarding the apoE promoter polymorphism -219G>T at position -203 upstream of +1 transcription start site, we observed no statistical significant change in the apoE promoter activity neither in the case of the most common -219G nor in the case of the less common -219T alleles. This observation is in contrast with a study performed in 1998 by Artiga *et al.*, in which it was shown in transient transfection experiments, that that allele T contributes to a significant decrease in apoE promoter activity comparing to the most common G allele (Artiga *et al.*, 1998). However, this study is in disagreement with a more recent study, in which -219G and -219T alleles resemble no difference in apoE promoter activity (Mannila *et al.*, 2013). Our results regarding the polymorphism -219G>T are in agreement with the results observed by Mannila *et al.* in 2013.

To summarize, in this work we elucidated the critical importance of variants in both apoE and CETP genes from the CCHS. It was demonstrated that specific polymorphisms both in apoE and in CETP promoter regions, are responsible for the up- or down-regulation of the corresponding genes. This is extremely significant for individuals bearing most or less frequent alleles for the specific variants, if one

considers how specific regulation of these genes may affect disease phenotypes and contribution to cardiovascular risk.

1.5. REFERENCES

- Agellon L., E. Quinet, T. Gillette, D. Drayna, M. Brown, A. R. Tall (1990) Organization of the human cholesteryl ester transfer protein gene. *Biochemisty* **29**: 1372-1376.
- Agellon L.B., Zhang P., Jiang X.C., Mendelsohn L., Tall A.R. (1992) The CCAAT/enhancer-binding protein trans-activates the human cholesteryl ester transfer protein gene promoter. *J Biol Chem* **267**: 22336–22339.
- Akbarzadeh M., Hassanzadeh T., Saidijam M., Esmaeili R., Borzouei Sh., Hajilooi M., Mahjub H., Paoli M. (2012) Cholesteryl ester transfer protein (CETP) -629C/A polymorphism and its effects on the serum lipid levels in metabolic syndrome patients. *Mol Biol Rep* **39**: 9529-9534.
- Ammanamanchi S., Brattain M.G. (2001) Sp3 is a transcriptional repressor of transforming growth factor-beta receptors. *J. Biol. Chem.* **276**: 3348–3352.
- Ammanamanchi S., Freeman J.W., Brattain M.G. (2003) Acetylated Sp3 is a transcriptional activator. *J. Biol. Chem.* **278**: 35775–35785.
- Anderson J.L., Carlquist J.F. (2003) Genetic polymorphisms of hepatic lipase and cholesteryl ester transfer protein, intermediate phenotypes, and coronary risk. *Journal of American College of Cardiology* **41(11)**: 1990-1993.
- Artieda M., Gañán A., Cenarro A., García-Otín A.L., Jericó I., Civeira F., Pocoví M. (2008) Association and linkage disequilibrium analyses of *APOE* polymorphisms in atherosclerosis. *Disease Markers* **24**: 65-72.
- Artiga M., Bullido M.J., Sastre I., Recuero M., García M.A., Aldudo J., Vázquez J., Valdivieso F. (1998) Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene. *FEBS Letters* **421**: 105-108.
- Ayyobi A.F., Hill J.S., Molhuizen H.O.F., Lear S.A. (2005) Cholesterol ester transfer protein (CETP) Taq1B polymorphism influences the effect of a standardized cardiac rehabilitation program on lipid risk markers. *Atherosclerosis* **181**: 363-369.
- Bakovic M., Waite K., Vance D.E. (2003) Oncogenic Ha-Ras transformation modulates the transcription of the CTP: phosphocholine cytidylyltransferase alpha gene via p42/44MAPK and transcription factor Sp3. *J. Biol. Chem.* **278**: 14753– 14761.
- Bañares V.G., Bardach A., Peterson G., Tavella M.J., Schreier L.E. (2012) APOE -491 T allele may reduce risk of atherosclerotic lesions among middle-aged women. *Mol Cell Biochem* **362(1-2)**: 123-131.
- Barter P. J., Brewer H. B., Chapman M. J. Jr., Hennekens C. H., Rader D. J., Tall A. R. (2003) Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **23**: 160–167.
- Barter P.J., Caulfield M., Eriksson M., Grundy S.M., Kastelein J.J., Komajda M., Lopez-Sendon J., Mosca L., Tardif J.C., Waters D.D., Shear C.L., Revkin J.H., Buhr K.A., Fisher M.R., Tall A.R., Brewer B. (2007) Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med.* **357**: 2109–2122.

- Bernard S., Moulin P., Lagrost L., Picard S., Elchebly M., Ponsin G., Chapuis F., Berthezene F. (1998) Association between plasma HDL-cholesterol concentration and Taq1B CETP gene polymorphism in non-insulin-dependent diabetes mellitus. *J. Lipid. Res.* **39**: 59–65.
- Besler C., Lüscher T.F., Landmesser U. (2012) Molecular mechanisms of vascular effects of High-density lipoprotein: alterations in cardiovascular disease. *EMBO Mol Med* **4**: 251-268.
- Biggs J. R., J. E. Kudlow, and A. S. Kraft. (1996) The role of the transcription factor Sp1 in regulating the expression of the WAF1/CIP1 gene in U937 leukemic cells. *J. Biol. Chem*. b: 901–906.
- Boekholdt S. M., Kuivenhoven J. A., Wareham N. J., Peters R. J., Jukema J. W., Luben R., Bingham S. A., Day N. E., Kastelein J. J., Khaw K. T. (2004) Plasma levels of cholesteryl ester transfer protein and the risk of future coronary artery disease in apparently healthy men and women: the prospective EPIC (European Prospective Investigation into Cancer and nutrition)-Norfolk population study. *Circulation* **110**: 1418–1423.
- Braun H., Koop R., Ertmer A., Nacht S., Suske G. (2001) Transcription factor Sp3 is regulated by acetylation. *Nucleic Acids Res.* **29**: 4994–5000.
- Brousseau M.E., O'Connor J.J., Ordovas J.M., Collins D., Otvos J.D., Massov T., McNamara J.R., Rubins H.B., Schaefer E.J. (2002) Cholesteryl ester transfer protein TaqI B2B2 genotype is associated with higher HDL cholesterol levels and lower risk of coronary heart disease end points in men with HDL deficiency: Veterans Affairs HDL Cholesterol Intervention Trial. *Arterioscler. Thromb. Vasc. Biol.* **22**: 1148–1154.
- Brown B.G., Cheung M.C., Lee A.C., Zhao X.Q., Chait A. (2002) Antioxidant vitamins and lipid therapy. End of a long romance? *Arterioscler Thromb Vasc Biol* **22**: 1535–1546.
- Bu Guojun (2009) Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nature Reviews Neuroscience* **10**: 333-344.
- Corbex M., Poirier O., Fumeron F., Betoulle D., Evans A., Ruidavets J.B., Arveiler D., Luc G., Tiret L., Cambien F. (2000) Extensive association analysis between the CETP gene and coronary heart disease phenotypes reveals several putative functional polymorphisms and gene-environment interaction. *Genetic Epidemiology* **19**: 64-80.
- Corbo R.M., Scacchi R., Vilardo T., Ruggeri M. (2001) Polymorphisms in the apolipoprotein E gene regulatory region in relation to coronary heart disease and their effect on plasma apolipoprotein E. *Clin Chem Lab Med.* **39(1)**: 2-6.
- Dachet C., Poirier O., Cambien F., Chapman J., Rouis M. (2000) New functional promoter polymorphism, CETP/-629, in cholesterol ester transfer protein (CETP) gene related to CETP mass and high density lipoprotein cholesterol

levels: Role of Sp1/Sp3 in transcriptional regulation. *Arterioscler Thromb Vasc Biol.* **20**: 507-515.

- Daneshpour M., Hedayati M., Azizi F. (2007) TaqI B1/B2 and -629A/C cholesterol ester transfer protein (CETP) gene polymorphisms and their association with CETP activity and high-density lipoprotein cholesterol levels in a Tehranian population. Part of the Tehran Lipid and Glucose Study (TLGS) *Genet. Mol. Biol.* **30(4)**: 1039-1046.
- de Vries R., Perton F. G., Dallinga-Thie G. M., van Roon A. M., Wolffenbuttel B. H., van Tol A., Dullaart R. P. (2005) Plasma cholesteryl ester transfer is a determinant of intima-media thickness in type 2 diabetic and nondiabetic subjects: role of CETP and triglycerides. *Diabetes* **54**: 3554–3559.
- Drayna D., A. S. Jarnagin, J. McLean, W. Henzel, W. Kohr, C. Fielding, and R. Lawn (1987) Cloning and sequencing of human cholesteryl ester transfer protein cDNA. *Nature (London)* **327**: 632-634.
- Dullaart R.P.F., Borggreve S.E., Hillege H.L., Dallinga-thie G.M. (2008) The association of HDL cholesterol concentration with the -629C>A CETP promoter polymorphism is not fully explained by its relationship with plasma cholesteryl ester transfer. *The Scandinavian Journal of Clinical & Laboratory Investigation* **68(2)**: 99-105.
- Faust R. A., J. J. Albers (1987) Synthesis and secretion of plasma cholesteryl ester transfer protein by human hepatocarcinoma cell line, HepG2. *Arteriosclerosis.* **7**: 267-275.
- Faust R. A., J. J. Albers (1988) Regulated vectorial secretion of cholesteryl ester transfer protein (LTP-I) by the CaCo-2 model of human enterocyte epithelium. *J. Biol. Chem.* **263**: 8786-8789.
- Faust R. A., M. C. Cheung, J. J. Albers (1989) Secretion of cholesteryl ester transfer protein-lipoprotein complexes by human HepG2 hepatocytes. *Atherosclerosis* **77**: 77-82.
- Faust R. A., J. H. Tollefson, A. Chait, J. J. Albers (1990) Regulation of LTP-I secretion from human monocyte-derived macrophages by differentiation and cholesterol accumulation in vitro. *Biochim. Biophys. Acta.* **1042**: 404-409.
- Freeman D.J., Griffin B.A., Holmes A.P., Lindsay G.M., Gaffney D., Packard C.J., Shepherd J. (1994) Regulation of plasma HDL cholesterol and subfraction distribution by genetic and environmental factors: associations between the TaqI B RFLP in the CETP gene and smoking and obesity. *Arterioscler. Thromb.* **14**: 336–344.
- Frisdal E., Klerkx A.H.E.M., Le Goff W., Tanck M.W.T., Lagarde J-P., Jukema J.W., Kastelein J.J.P., Chapman M.J., Guerin M. (2005) Functional interaction between -629C/A, -971G/A and -1337C/T polymorphisms in the CETP gene is a major determinant of promoter activity and plasma CETP concentration in the REGRESS study. *Human Molecular Genetics* **14(18)**: 2607-2618.

- Fumeron F., Betoulle D., Luc G., Behague I., Ricard S., Poirier O., Jemaa R., Evans A., Arveiler D., Marques-Vidal P. et al. (1995) Alcohol intake modulates the effect of a polymorphism of the cholesteryl ester transfer protein gene on plasma high-density lipoprotein and the risk of myocardial infarction. *J. Clin. Invest.* **96**: 1664–1671.
- Gafencu A.V., Robciuc M.R., Fuior E., Zannis V.I., Kardassis D. (2007) Molecular basis of cell and developmental biology: Inflammatory signaling pathways regulating apoE gene expression in macrophages. *J. Biol. Chem.* **282**: 21776-21785.
- Gaudet F., Ginsburg G.S. (1995) Transcriptional regulation of the cholesteryl ester transfer protein gene by the orphan nuclear hormone receptor apolipoproteinAI regulatory protein-1. *J Biol Chem* **270**: 29916–29922.
- Gauthier B., Robb M., Gaudet F., Ginsburg G.S., McPherson R. (1999) Characterization of a cholesterol response element (CRE) in the promoter of the cholesteryl ester transfer protein gene: functional role of the transcription factors SREBP-1a, -2, and YY1. *J Lipid Res* **40**: 1284–1293.
- Gautier T., de Haan W., Grober J., Ye D., Bahr M.J., Claudel T. et al. (2013) Farnesoid X receptor activation increases cholesteryl ester transfer protein expression in humans and transgenic mice. *J Lipid Res* **54**: 2195–2205.
- Ge Y., Jensen T.L., Matherly L.H., Taub J.W. (2002) Synergistic regulation of human cystathionine-beta-synthase-1b promoter by transcription factors NF-YA isoforms and Sp1. *Biochim. Biophys. Acta* **1579**: 73–80.
- Geng J., Law P.P.Y., Ng M.C.Y., Li T., Liang L.Y., Ge T.F., Wong K.B., Liang C., Ma R.C., So W.Y., Chan J.C.N., Ho Y.Y. (2011) *APOE* genotype-function relationship: evidence of -491A/T promoter polymorphism modifying transcription control but not Type 2 Diabetes risk. *Plos ONE* **6(10)**: e24669.
- Gu X., Kozarsky K., Krieger M. (2000) Scavenger receptor class B, type I-mediated [3H]cholesterol efflux to high and low density lipoproteins is dependent on lipoprotein binding to the receptor. *J Biol Chem.* **275**: 29993–30001.
- Hannuksela M.L., Liinamaa M.J., Kesaniemi Y.A., Savolainen M.J. (1994) Relation of polymorphisms in the cholesteryl ester transfer protein gene to transfer protein activity and plasma lipoprotein levels in alcohol drinkers. *Atherosclerosis* **110**: 35–44.
- Hatters D. M., Peters-Libeu C. A., Weisgraber K. H. (2006) Apolipoprotein E structure: insights into function. *Trends Biochem. Sci.* **31**: 445–454.
- Honzumi S., Shima A., Hiroshima A., Koieyama T., Ubukata N., Terasaka N. (2010) LXRalpha regulates human CETP expression in vitro and in transgenic mice. *Atherosclerosis* **212**: 139–145.
- Hoppe K.L., Francone O.L. (1998) Binding and functional effects of transcription factors Sp1 and Sp3 on the proximal human lecithin: cholesterol acyltransferase promoter. *J Lipid Res.* **39**: 969–977.

- Horiuchi K., Tajima S., Menju M., Yamamoto A. (1989) Structure and expression of mouse apolipoprotein E gene. *J Biochem* **106**: 98–103.
- Jeoung N.H., Jang W.G., Nam J.I., Pak Y.K., Park Y.B. (1999) Identification of retinoic acid receptor element in human cholesteryl ester transfer protein gene. *Biochem Biophys Res Commun* **258**: 411–415.
- Ji Y., Jian B., Wang N., Sun Y., Moya M.L., Phillips M.C., et al. (1997) Scavenger receptor BI promotes high density lipoprotein mediated cellular cholesterol efflux. *J Biol Chem*. **272**: 20982–20985.
- Jiang X., P. Moulin, E. Quinet, I. J. Goldberg, L. K. Yacoub, L. B. Agellon, D. Compton, R. Polokoff, A. R. Tall (1991) Mammalian adipose tissue and muscle are major sources of lipid transfer protein mRNA. *J Biol. Chem.* **266**: 4631-4639.
- Kardassis D., Mosialou I., Kanaki M., Tiniakou I., Thymiakou E. (2014) Metabolism of HDL and its regulation. *Current Medicinal Chemistry* **21(25)**: 2864-2880.
- Kardassis D., Gafencu A., Zannis V.I., Davalos A. (2015) Regulation of HDL genes: transcriptional, posttranscriptional, and posttranslational. *Handb Exp Pharmacol.* **224**: 113-179.
- Klerkx A.H.E.M., Tanck M.W.T., Kastelein J.J.P., Molhuizen H.O.F., Jukema J.W., Zwinderman A.H., Kuivenhoven J.A. (2003) Haplotype analysis of the CETP gene: not TaqIB, but the closely linked -629C→A polymorphism and a novel promoter variant are independently associated with CETP concentration. *Hum. Mol. Genet.* **12**: 111–123.
- Krieger M. (2001) Scavenger receptor class B type I is a multiligand HDL receptor that influences diverse physiologic systems. *J Clin Invest.* **108**: 793–797.
- Korhonen T., Hannuksela M.L., Seppanen S., Kervinen K., Kesaniemi M.J., Savolainen M.J. (1999) The effect of the apolipoprotein E phenotype on cholesteryl ester transfer protein activity, plasma lipids and apolipoprotein A I levels in hypercholesterolaemic patients on colestipol and lovastatin treatment. *Eur J Clin Pharmacol* **54**: 903 – 910.
- Kuivenhoven J.A., Jukema J.W., Zwinderman A.H., de Knijff P., McPherson R., Bruschke A.V., Lie K.I., Kastelein J.J.P. (1998) The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. The Regression Growth Evaluation Statin Study Group. *N. Engl. J. Med.* **338**: 86–93.
- Kypreos K. E., Van Dijk K. W., van Der Z. A., Havekes L. M., Zannis V. I. (2001) Domains of apolipoprotein E contributing to triglyceride and cholesterol homeostasis in vivo: carboxyl-terminal region 203–299 promotes hepatic very low density lipoprotein–triglyceride secretion. *J. Biol. Chem.* **276**: 19778–19786.
- Kypreos K. E., van Dijk K. W., Havekes L. M., Zannis, V. I. (2005) Generation of a recombinant apolipoprotein E variant with improved biological functions:

hydrophobic residues (Leu-261, Trp-264, Phe-265, Leu-268, Val-269) of apoE can account for the apoE-induced hypertriglyceridemia. *J. Biol. Chem.* **280**: 6276–6284.

- Kypreos K.E., Zannis V.I. (2007) Pathway of biogenesis of apolipoprotein E-containing HDL *in vivo* with the participation of ABCA1 and LCAT. *Biochem. J.* **403**: 359-367.
- Lakomy D., Rebe C., Sberna A.L., Masson D., Gautier T., Chevriaux A. et al. (2009) Liver X receptormediated induction of cholesteryl ester transfer protein expression is selectively impaired in inflammatory macrophages. *Arterioscler Thromb Vasc Biol* **29**: 1923–1929.
- Lambert J.C., Brousseau T., Defosse V., Evans A., Arveiler D., Ruidavets J.B., Haas B., Cambou J.P., Luc G., Ducimetière P., Cambien F., Chartier-Harlin M.C., Amouyel P. (2000) Independent association of an APOE gene promoter polymorphism with increased risk of myocardial infarction and decreased APOE plasma concentrations-the ECTIM Study. *Human Molecular Genetics* **9(1)**: 57-61.
- Laws S.M., Taddei K., Martins G., Paton A., Fisher C., et al. (1999) The 2491AA polymorphism in the APOE gene is associated with increased plasma apoE levels in Alzheimer's disease. *Neuroreport* **10**: 879–882.
- Le Goff W., Guerin M., Nicaud V., Dachet C., Luc G., Arveiler D., Ruidavets J.B., Evans A., Kee F., Morrison C., Chapman M.J., Thillet J (2002) A novel cholesteryl ester transfer protein promoter polymorphism (-971G/A) associated with plasma high-density lipoprotein cholesterol levels: Interaction with the TaqIB and -629C/A polymorphisms. *Atherosclerosis* **161(2)**: 269-279.
- Le Goff W., Guerin M., Petit L., Chapman M.J., Thillet J. (2003a) Regulation of human CETP gene expression: role of SP1 and SP3 transcription factors at promoter sites -690, -629, and -37. *J Lipid Res* **44**: 1322–1331.
- Le Goff W., Guerin M., Chapman M.J., Thillet J. (2003b) A CYP7A promoter binding factor site and Alu repeat in the distal promoter region are implicated in regulation of human CETP gene expression. *Journal of Lipid Research* **44**: 902-910.
- Li X., Kypreos K., Zanni E.E., Zannis V.I. (2003) Domains of apoE required for binding to apoE receptor 2 and to phospholipids: Implications for the functions of apoE in the brain. *Biochemistry* **42**: 10406–10417.
- Liang Y., D. F. Robinson, J. Dennig, G. Suske, W. E. Fahl. (1996) Transcriptional regulation of the SIS/PDGF-B gene in human osteosarcoma cells by the Sp family of transcription factors. *J. Biol. Chem.* **271**: 11792–11797.
- Liu T., Krieger M., Kan H.Y., Zannis V.I. (2002) The effects of mutations in helices 4 and 6 of apoA-I on scavenger receptor class B type I (SR-BI)-mediated cholesterol efflux suggest that formation of a productive complex between

reconstituted high density lipoprotein and SR-BI is required for efficient lipid transport. *J Biol Chem.* **277**: 21576–21584.

- Lu H., Inazu A., Moriyama Y., Higashikata T., Kawashiri M., Yu W., Huang Z., Okamura T., Mabuchi H. (2002) Haplotype analyses of cholesteryl ester transfer protein gene promoter: a clue to an unsolved mystery of TaqIB polymorphism. *J Mol Med* **81**: 246-255.
- Lu Y., Tayebi N., Li H., Saha N., Yang H., Heng C.K (2013) Association of CETP Taq1B and -629C>A polymorphisms with coronary artery disease and lipid levels in the multi-ethnic Singaporean population. *Lipids in Health and Disease* **12**: 1-13.
- Luo Y., Tall A.R. (2000) Sterol upregulation of human CETP expression in vitro and in transgenic mice by an LXR element. *J Clin Invest* **105**: 513–520.
- Luo Y., Liang C.P., Tall A.R. (2001) The orphan nuclear receptor LRH-1 potentiates the sterolmediated induction of the human CETP gene by liver X receptor. *J Biol Chem* **276**: 24767–24773.
- Lusis A. J., S. Zollman, R. S. Sparkes, I. Klisak, T. Mohandas, D. Drayna, R. M. Lawn (1987) Assignment of the human gene for cholesteryl ester transfer protein to chromosome 16q12-16q21. *Genomics* **1**: 232-235.
- Mahley R.W. (1988) Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* **240(4852)**: 622-630.
- Mannila M.N., Mahdessian H., Franco-Cereceda A., Eggertsen G., de Faire U., Syvänen A.C., Eriksson P., Hamsten A., Ferdinand M. van't Hooft (2013) Identification of a functional apolipoprotein E promoter polymorphism regulating plasma apolipoprotein E concentration. *Arterioscler Thromb Vasc Biol.* **33**: 1063-1069.
- Masson D., Jiang X.C., Lagrost L., Tall A.R. (2009) The role of plasma lipid transfer proteins in lipoprotein metabolism and atherogenesis. *J Lipid Res* **50**: S201-S206.
- Moreno J.A., Pérez-Jiménez F., Marín C., Gómez P., Pérez-Martínez P., Moreno R., Bellido C., Fuentes F., López-Miranda J. (2004) Apolipoprotein E gene promoter -219G→T polymorphism increases LDL-cholesterol concentrations and susceptibility to oxidation in response to a diet rich in saturated fat $_{1\text{-}3}$. *Am J Clin Nutr* **80**: 1404-1409.
- Olaisen B., Teisberg P., Gedde-Dahl T. Jr. (1982) *Hum Genet.* **62**: 233–236.
- Ordovas J.M., Cupples L.A., Corella D., Otvos J.D., Osgood D., Martinez A., Lahoz C., Coltell O., Wilson P.W., Schaefer E.J. (2000) Association of cholesteryl ester transfer protein–TaqIB polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. *Arterioscler. Thromb. Vasc. Biol*. **20**: 1323–1329.

- Pagliuca A., Gallo P., Lania L. (2000) Differential role for Sp1/Sp3 transcription factors in the regulation of the promoter activity of multiple cyclin-dependent kinase inhibitor genes. *J. Cell. Biochem.* **76**: 360– 367.
- Pape M. E., R. G. Ulrich, T. J. Rea, K. R. Marotti, G. W. Melchior (1991) Evidence that the nonparenchymal cells of the liver are the principal source of cholesteryl ester transfer protein in primates. *J. Biol. Chem.* **266**: 12829-12831.
- Park S.S., Choi H., Kim S.J., Kim O.J., Chae K.S., Kim E. (2008) FXRalpha down-regulates LXRalpha signaling at the CETP promoter via a common element. *Mol Cells* **26**: 409–414.
- Porchay-Baldèrelli I., Pèan F., Bellini N., Jaziri R., Marre M., Fumeron F. (2007) The CETP TaqIB polymorphism is associated with the risk of sudden death in Type 2 Diabetic patients. *American Diabetes Association* 1-13.
- Prowse D. M., L. Bolgan, A. Molnar, and G. P. Dotto. (1997) Involvement of the Sp3 transcription factor in induction of p21Cip1/WAF1 in keratinocyte differentiation. *J. Biol. Chem.* **272**: 1308–1314.
- Rader D.J. (2007) Illuminating HDL--is it still a viable therapeutic target? *N Engl J Med.* **357**: 2180–2183.
- Rader D. J., Daugherty A. (2008) Translating molecular discoveries into new therapies for atherosclerosis. *Nature* **451**: 904–913.
- Rader D.J. (2009) Lecithin/cholesterol acyltransferase and atherosclerosis: another high-density lipoprotein story that doesn't quite follow the script. *Circulation* **120**: 549-552.
- Rajavashisth T.B., Kaptein J.S., Reue K.L., Lusis A.J. (1985) Evolution of apolipoprotein E: mouse sequence and evidence for an 11-nucleotide ancestral unit. *Proc Natl Acad Sci USA* **82**: 8085–8089.
- Ridker P. M., Paré G., Parker A. N., Zee R. Y., Miletich J. P., Chasman D. I. (2009) Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction: genomewide analysis among 18,245 initially healthy women from the Women's Genome Health Study. *Circ Cardiovasc Genet.* **2**: 26–33.
- Ross S., Best J.L., Zon L.I., Gill G. (2002) SUMO-1 modification represses Sp3 transcriptional activation and modulates its subnuclear localization. *Mol. Cell* **10**: 831– 842.
- Sapetschnig A., et al. (2002) Transcription factor Sp3 is silenced through SUMO modification by PIAS1. *EMBO J.* **21**: 5206–5215.
- Scacchi R., Gambina G., Martini M.C., Ruggeri M., Ferrari G., et al. (2001) Polymorphisms of the apolipoprotein E gene regulatory region and of the LDL receptor gene in late-onset Alzheimer's disease in relation to the plasma lipidic pattern. *Dement Geriatr Cogn Disord* **12**: 63–68.

- Schnohr Peter, G. Jensen, P. Lange, H. Scharling, M. Appleyard (2001) The Copenhagen City Heart Study. Tables with data from the third examination 1991-1994. *European Heart Journal 3* **(suppl H)**: H1-H83.
- Shih S.J., Allan C., Grehan S., Tse E., Moran C., Taylor J.M. (2000) Duplicated downstream enhancers control expression of the human apolipoprotein E gene in macrophages and adipose tissue. *J Biol Chem* **275**: 31567–31572.
- Sladek F.M., Ruse M.D. Jr, Nepomuceno L., Huang S.M., Stallcup M.R. (1999) Modulation of transcriptional activation and coactivator interaction by a splicing variation in the F domain of nuclear receptor hepatocyte nuclear factor 4alpha1. Mol Cell Biol. 19(10): 6509-6522.
- Sorlí J.V., Corella D., Francés F., Ramírez J.B., González J.I., Guillén M., Portolés O. (2006) The effect of the *APO*E polymorphism on HDL-C concentrations depends on the cholesterol ester transfer protein gene variation in a Southern European population. *Clinica Chemica Acta* **366**: 196-203.
- Spielmann N., Leon A.S., Rao D.C., Rice T., Skinner J.S., Bouchard C., Rankinen T. (2007) CETP genotypes and HDL-cholesterol phenotypes in the HERITAGE family study. *Physiol Genomics* **31**: 25-31.
- Suske G. (1999) The Sp-family of transcription factors. *Gene* **238**: 291– 300.
- Swenson T., J. Simmons, C. Hesler, C. Bisgaier, A. R. Tall (1987) Cholesteryl ester transfer protein is secreted by HepG2 cells and contains asparagine-linked carbohydrate and sialic acid. *J. Biol. Chem.* **262**: 16271-16274.
- Tall AR. (1993) Plasma cholesteryl ester transfer protein. *J Lipid Res* **34**: 1255–1274.
- Tall A.R., Yvan-Charvet L., Wang N. (2007) The failure of torcetrapib: was it the molecule or the mechanism? *Arterioscler Thromb Vasc Biol.* **27**: 257–260.
- Tall A.R., Yvan-Charvet L., Terasaka N., Pagler T., Wang N. (2008) HDL, ABC transporters, and cholesterol efflux: implications for the treatment of atherosclerosis. *Cell Metab* **7**: 365-375.
- Tai E.S., Ordovas J.M., Corella D., Deurenberg-Yap M., Chan E., Adiconis X., Chew S.K., Loh L.M., Tan C.E. (2003) The TaqIB and -629C>A polymorphisms at the cholesteryl ester transfer protein locus: associations with lipid levels in a multiethnic population. The 1998 Singapore National Health Survey. *Clin Genet* **63**: 19-30.
- Thompson J.F., Lira M.E., Durham L.K. et al. (2003) Polymorphisms in the CETP gene and association with CETP mass and HDL levels. *Atherosclerosis* **167**: 195–204.
- Thompson J.F., Lloyd D.B., Lira M.E., Milos P.M. Cholesteryl ester transfer protein promoter single-nucleotide polymorphisms in Sp1-binding sites affect transcription and are associated with high-density lipoprotein cholesterol. *Clin Genet* **66**: 223-228.

- Tu A.Y., Albers J.J. (2001) Functional analysis of the transcriptional activity of the mouse phospholipid transfer protein gene. *Biochem Biophys Res Comm* **287**: 921–926.
- Van der Velde A.E., Brufau G., Groen A.K. (2010) Transintestinal cholesterol efflux. *Curr. Opin. Lipidol.* **21(3)**: 167-171.
- Viitanen L., Pihlajamäki J., Miettinen R., Kärkkäinen A., Lehto S., Laakso M. (2001) Apolipoprotein E gene promoter (-219G/T) polymorphism is associated with premature coronary heart disease. *J Mol Med* **79**: 732-737.
- Von Eckardstein A and Kardassis D. High density lipoproteins. From biological understanding to clinical exploitation. *Handbook of Experimental Pharmacology* **224**.
- Von Eckardstein A., Hersberger M., Rohrer L. (2005) Current understanding of the metabolism and biological actions of HDL. *Curr. Opin. Clin. Nutr. Metab. Care* **8(2)**: 147-152.
- Wang N., Lan D., Chen W., Matsuura F., Tall A.R. (2004) ATP binding cassette transporters G1 and G4 mediate cellular cholesterol efflux to high-density lipoproteins. *Proc Natl Acad Sci U S A.* **101**: 9774–9779.
- Weisgraber K. H., Rall S. C. Jr., Mahley R. W. (1981) Human E apoprotein heterogeneity. Cysteine-arginine interchanges in the amino acid sequence of the apo-E isoforms. *J. Biol. Chem.* **256**: 9077–9083.
- Yancey P.G., de la Llera-Moya M., Swarnakar S., Monzo P., Klein S.M., Connelly M.A., Johnson W.J., Williams D.L., Rothblat G.H. (2000) High density lipoprotein phospholipid composition is a major determinant of the bi-directional flux and net movement of cellular free cholesterol mediated by scavenger receptor BI. *J. Biol. Chem.* **275(47)**: 36596-36604.
- Yang W-S., Deeb S.S. (1998) Sp1 and Sp3 transactivate the human lipoprotein lipase gene promoter through binding to a CT element: synergy with sterol regulatory element binding protein and reduced transactivation of a naturally occurring promoter variant. *J Lipid Res* **39**: 2054–2064.
- Yilmaz H., Isbir T., Agachan B., Karaali Z.E. (2005) Effects of cholesterol ester transfer protein Taq1B gene polymorphism on serum lipoprotein levels in Turkish coronary artery disease patients. *Cell Biochem Funct.* **23(1)**: 23-28.
- Zannis V.I., Cole F.S., Jackson C.L., Kurnit D.M., Karathanasis S.K. (1985) Distribution of apolipoprotein A-I, C-II, C-III, and E mRNA in fetal human tissues. Time-dependent induction of apolipoprotein E mRNA by cultures of human monocyte-macrophages. *Biochemistry* **24**: 4450–4455.
- Zannis V.I., Kan H.Y., Kritis A., Zanni E.E., Kardassis D. (2001) Transcriptional regulatory mechanisms of the human apolipoprotein genes in vitro and in vivo. *Curr Opin Lipidol.* **12(2)**: 181-207.

- Zannis V. I., Chroni A., Kypreos K. E., Kan H. Y., Cesar T. B., Zanni E. E., Kardassis D. (2004) Probing the pathways of chylomicron and HDL metabolism using adenovirus-mediated gene transfer. *Curr. Opin. Lipidol.* **15**: 151–166.
- Zannis V.I., Chroni A., Krieger M. (2006) Role of apoA-I, ABCA1, LCAT, and SR-BI in the biogenesis of HDL. *J. Mol. Med.* **84(4)**: 276-294.
- Zannis V.I., Koukos G., Drosatos K., Vezeridis A., Zanni E.E., Kypreos K.E., Chroni A. (2008) Discrete roles of apoA-I and apoE in the biogenesis of HDL species: Lessons learned from gene transfer studies in different mouse models. *Annals of Medicine* **40**: 14-28.
- Zeller M., Masson D., Farnier M., Lorgis L., Deckert V., Pais de Barros J. P., Desrumaux C., Sicard P., Grober J., Blache D., et al. (2007) High serum cholesteryl ester transfer rates and small high-density lipoproteins are associated with young age in patients with acute myocardial infarction. *J. Am. Coll. Cardiol.* **50**: 1948–1955.
- Zhong N., Weisgraber K. H. (2009) Understanding the association of apolipoprotein E4 with Alzheimer's disease: clues from its structure. *J. Biol. Chem.* **284**: 6027–6031.

1.6. APPENDIX

1.6.1. Sequences according to the Copenhagen City Heart Study

1.6.1.1. Human Cholesteryl Ester Transfer Protein (CETP) Promoter and Intron 1 sequence

HCETP -1054F-T new

HCETP -1073C-S new

```
-1070 gcaatcagct tatgtctctg agcc[t]tggga aacagtgagg gtcagcccgg ctccccacgt
                                          -998G>A (rs4783961)
-1010 gcttctgggc agctttggta ttggagcagg tgcaaactc[g] ggactagggc aggacccccct
 -950 gagaggcgac tgagcaaggc catcccgact catgtttcct tggccctgcc cggggcacag
 -890 catcctgccc acatccctgc agccctggct ccttcctagg ggctctgagg aggcagcact
       -854C>T (rs4783962)
 -830 tgg[c]catctg gtcacagttg ctgcagggca gttcttggcc ccagctgtag gtaaagtact
 -770 gtatgttgta atttttgaa agataacacg ttcacacaac tcagaattga aatgccacag
 -710 acattccccc tgctccgccc ctttcccccg gatacccagt ttctcccgga ggcagccaat
       -658C>A(rs1800776)     -656C>A(rs1800775)
 -650 gatctcagag gctgtatac[c] c[c]cccagagt tattttatgc atatcaagga aagtctacat
 -590 agaggactgt ttctggggta cccagatgca gcgtcaaatg ccatggaata ctacagtgag
 -530 gacattatcc tttcaagctt tcaaatcaga gcaagggaaa ggtcgatgct agagtttctc
 -470 tagcacccat gaagccctct cccttttct actgagtttt actttacagg caacagcagg
 -410 cttcaagctt ggggtcattg tcgggcaaca gtatctggca agaattcaat gtctttttct
 -350 catagtcatt gtattttggc ctctttctat ttatggcaac tgagagagaa agcttattcc
                                         -282C>A (rs no. not available)
 -290 tagatatatg tatttaagta aaaaataaat gaatt[c]atgg aaacatatta agcaattatc
 -230 cagataacat aagggatggc aaaaatggtg cagatggtgg aggggagaca agtagaagtt
 -170 ggggtgctct tgttgaatgt ctggctctga actctagagg aggccgcagg ggctgggcag
 -110 gaaggaggtg aatctctggg gccaggaaga ccctgctgcc cggaagagcc tcatgttccg
              -65G>A (rs17231520)
  -50 tgggggctgg gc[g]gacatac atatacgggc tccaggctga acggctcggg ccacttacac
           -9C>T (rs no. not available)
  +11 accactgc[c]t gataaccatg ctggctgcca cagtcctgac cctggccctg ctgggcaatg
```

HCETP +27R-G new

HCETP -9T-AS new

```
  +71 cccatgcctg ctccaaaggc acctcgcacg aggcaggcat cgtgtgccgc atcaccaagc
 +131 ctgccctcct ggtgtgtaag tatcagtgca tctgtctgcc ctgccagggg tcttttcatg
```

HCETP INTR1 +191F

```
 +191 gacacccact atgccaggag cctccctggc ctgaagccag ccctgaagcc ggctgccaca
 +251 ctagcccaga gagaggagtg ccctgggagg gagatgggct gagtggagct gtcatcaccc
 +311 cctcctgacc tcgccttcaa ggtcaagttc tttggtgaga aggtcctagc tgcattgcaa
```

```
                                                  Taq1B (IVS1+279G>A) (rs708272)
 +371 acagccaggt atagggattt gtgtttgtct gagacccaga atcactgggg ttc[g]agttag
 +431 ggttcagatc tgagccaggt taggggggttg agtcaggggg taaagattag gaggttggtg
 +491 tatatttggt gttgggggtc actctatggc caaagtcagg ggttgccatg agctcaggtg
 +551 acggaggctc catcactgac tgttgtgtga ctttggccag ctcccctgcc ctctctgggc
 +611 ctcagtctct tgctcaatat aataagggta tagggaggct aaatgataca atttctaaaa
 +671 tagagtatcg ccaagttcaa aagccagaat tatagacccc caggactaca gacagtgtca
 +731 cagcatcgtc tgggtgaggc tagggttagt gtgcggctgg gctcagggct gccccatttg
 +791 ctaggatcgt ggggttccca tgtgtcagga tccagaggct agggtatgat caggatctct
 +851 agctggggtc agggtcagag ctctctgtgt cccctagaat tgccatcaac cttaaaccca
 +911 gagggaggcc cagtccaacc cctcagcttt aagacctgct gggagcctca tctcagagag
 +971 gctgagtcat ggccaaggcc agttgggggt gggagcaggg ggcttggtgt gggcctgcag
+1031 cccctcatcc actgccctcc cctctagtga accacgagac tgccaaggtg atccagaccg
```

HCETP INTR1 +1056R

With yellow labeling, at positions +294, +347 and +370 downstream of +1, the extra polymorphisms observed in our constructs are shown. More specifically those are substitutions of the most common allele: +294T>C (rs3816117), where we have the C allele in our constructs, +347G>A (rs711752), where we have the A allele in our constructs and +370A>G (rs not available), where we have the G allele in our constructs.

1.6.1.2. Human apolipoprotein E (apoE) Promoter sequence

Primer Forward

```
                                  -491A>T (rs449647)
 -500 accatgttgg ccaggctggt ctcaa[a]ctcc tgaccttaag tgattcgccc actgtggcct

                                  -427T>C (rs769446)
 -440 cccaaagtgc tgggattaca ggcgtgagc[t] accgccccca gcccctccca tcccacttct
 -380 gtccagcccc ctagccctac tttctttctg ggatccagga gtccagatcc ccagcccccct
 -320 ctccagatta cattcatcca ggcacaggaa aggacagggt caggaaagga ggactctggg

                                                        -219G>T (rs405509)
 -260 cggcagcctc cacattcccc ttccacgctt ggcccccaga atggaggagg gtgtctg[g]at
 -200 tactgggcga ggtgtcctcc cttcctgggg actgtggggg gtggtcaaaa gacctctatg
 -140 ccccacctcc ttcctccctc tgccctgctg tgcctggggc agggggagaa cagcccacct
  -80 cgtgactggg ggctggccca gcccgcccta tccctggggg agggggcggg acagggggag
  -20 ccctataatt ggacaagtct gggatccttg agtcctactc agccccagcg gaggtgaagg
  +41 acgtccttcc ccaggagccg gtgagaagcg cagtcggggg cacggggatg agctcagggg
```

Primer Reverse

1.6.2. Plasmid maps

1.6.2.1. Plasmid maps for CETP promoter polymorphisms

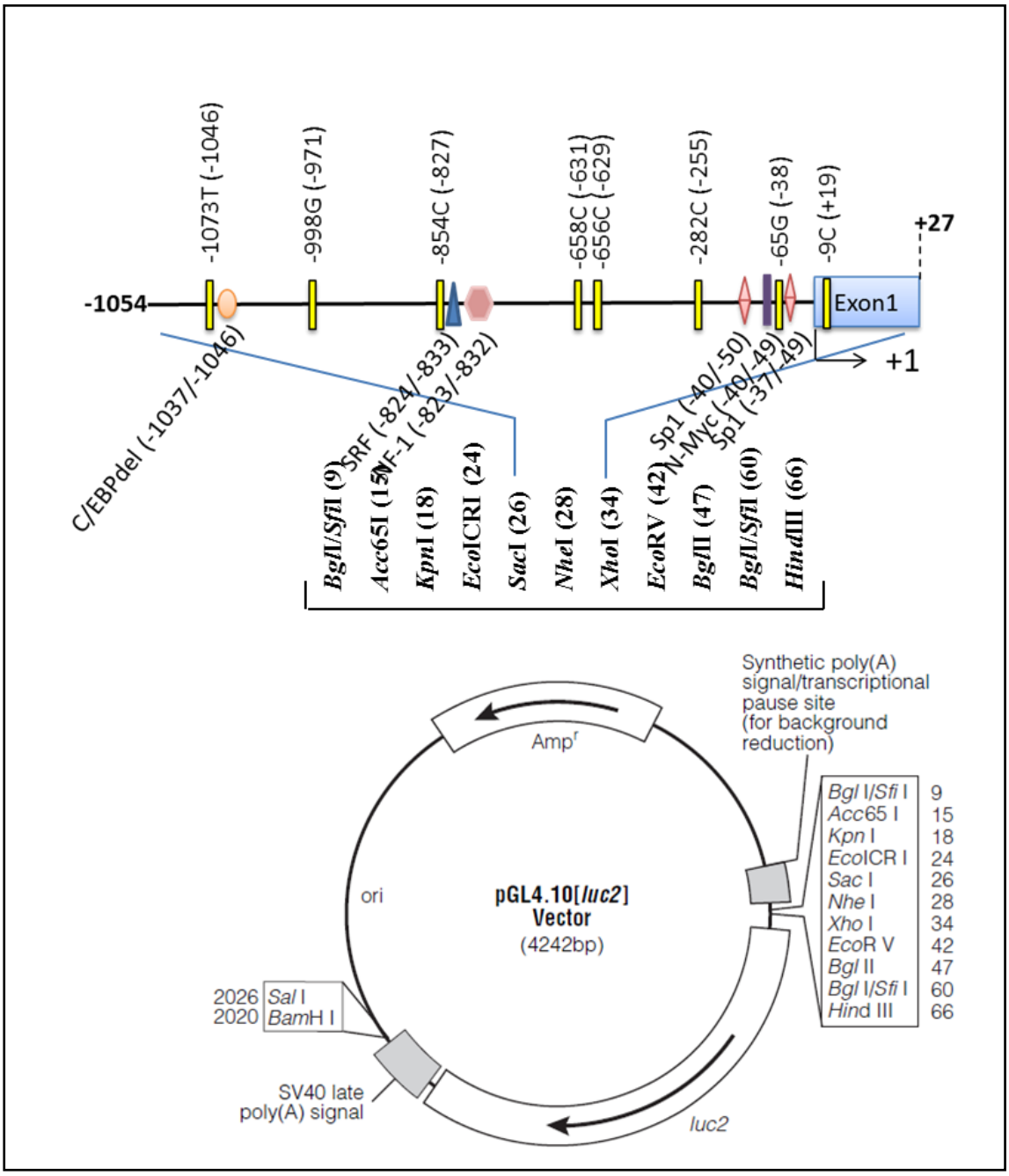

Figure 25. (-1054/+27) CETP-luc. Human CETP gene fragment -1054/+27, bearing the most common alleles for all SNPs under study, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

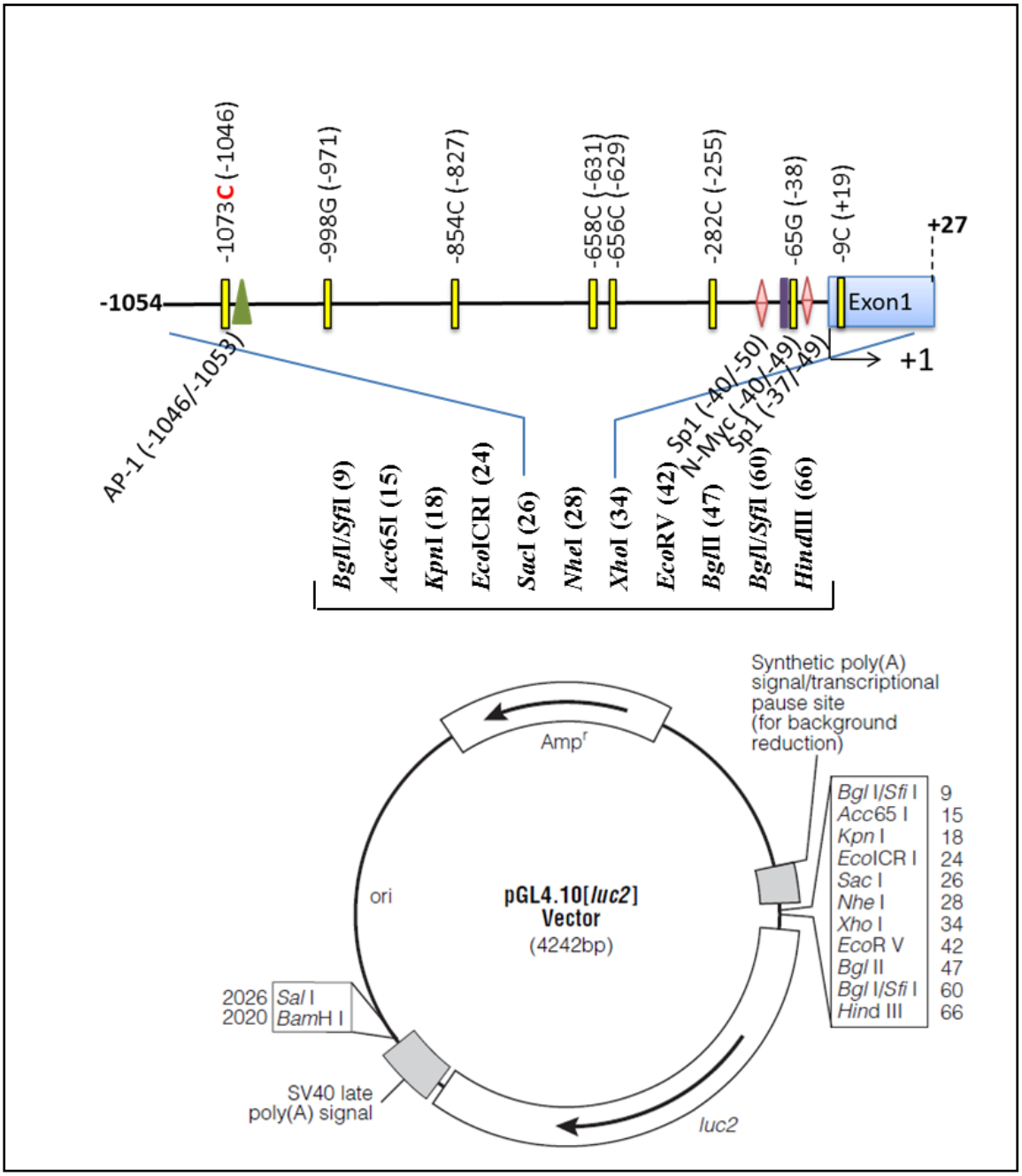

Figure 26. (-1054/+27) -1073C CETP-luc. Human CETP gene fragment -1054/+27, bearing the less common allele at position -1046, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

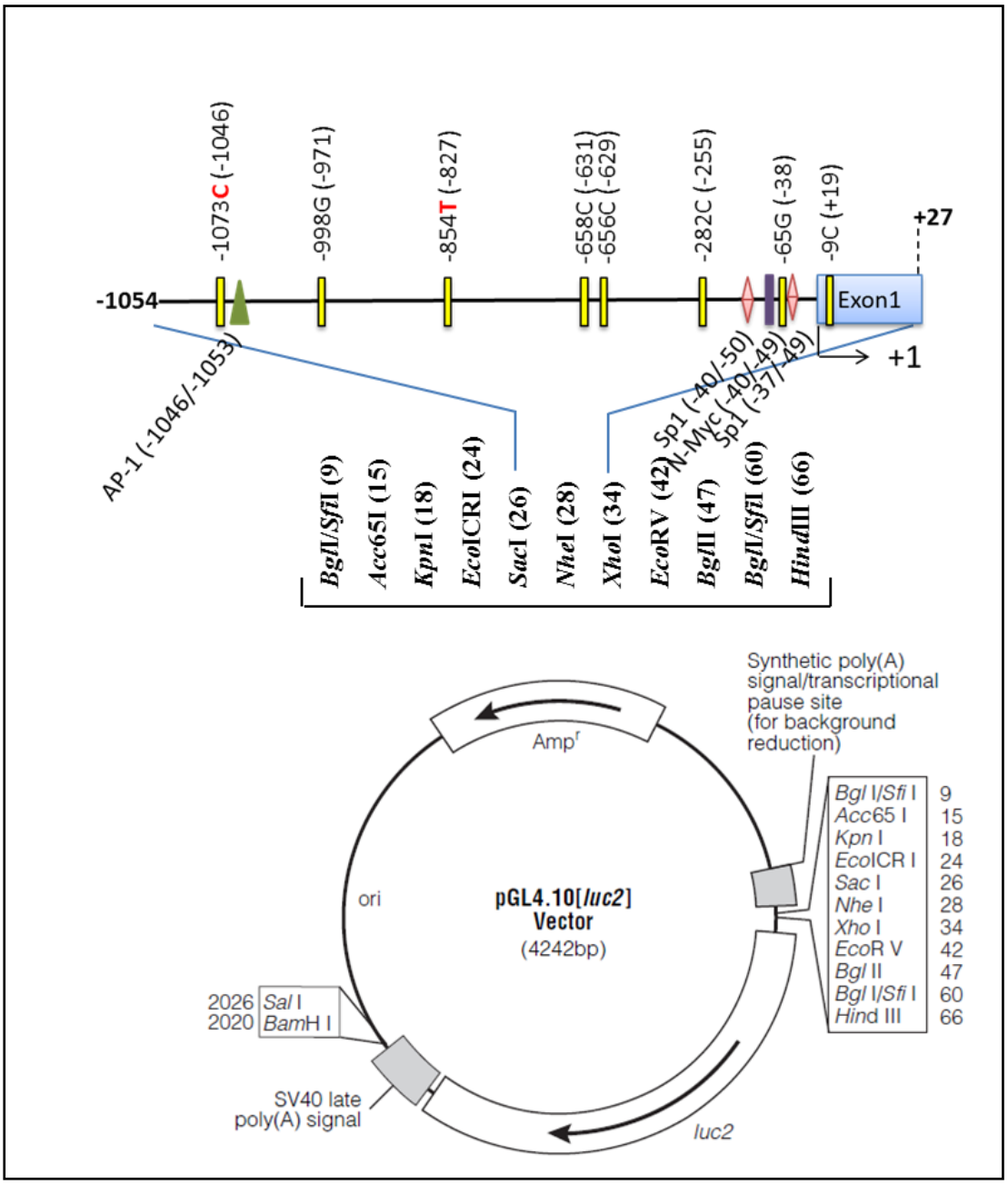

Figure 27. (-1054/+27) -1073C/-854T CETP-luc. Human CETP gene fragment -1054/+27, bearing the less common allele at positions -1046 and -827, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

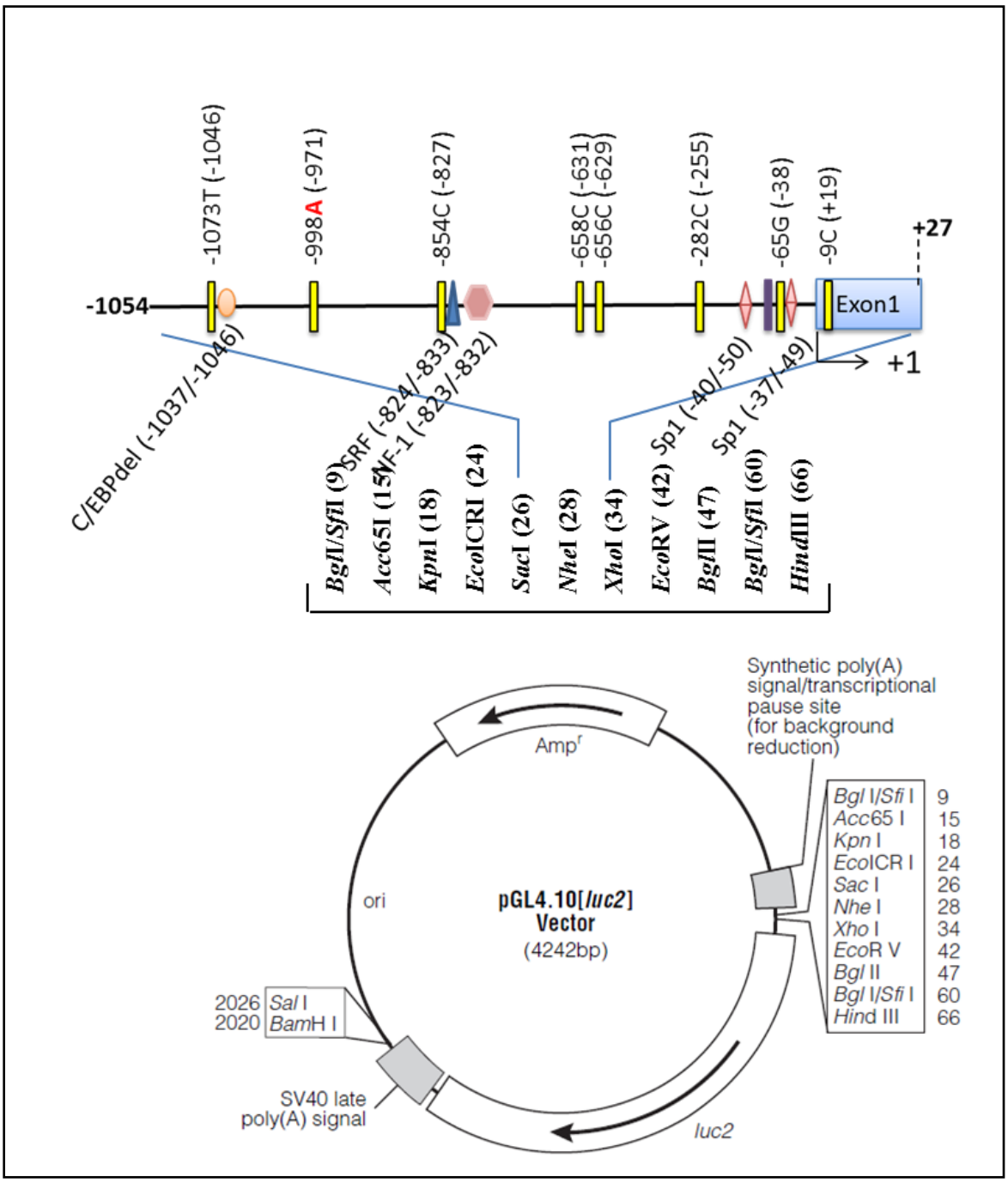

Figure 28. (-1054/+27) -998A CETP-luc. Human CETP gene fragment -1054/+27, bearing the less common allele at position -971, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

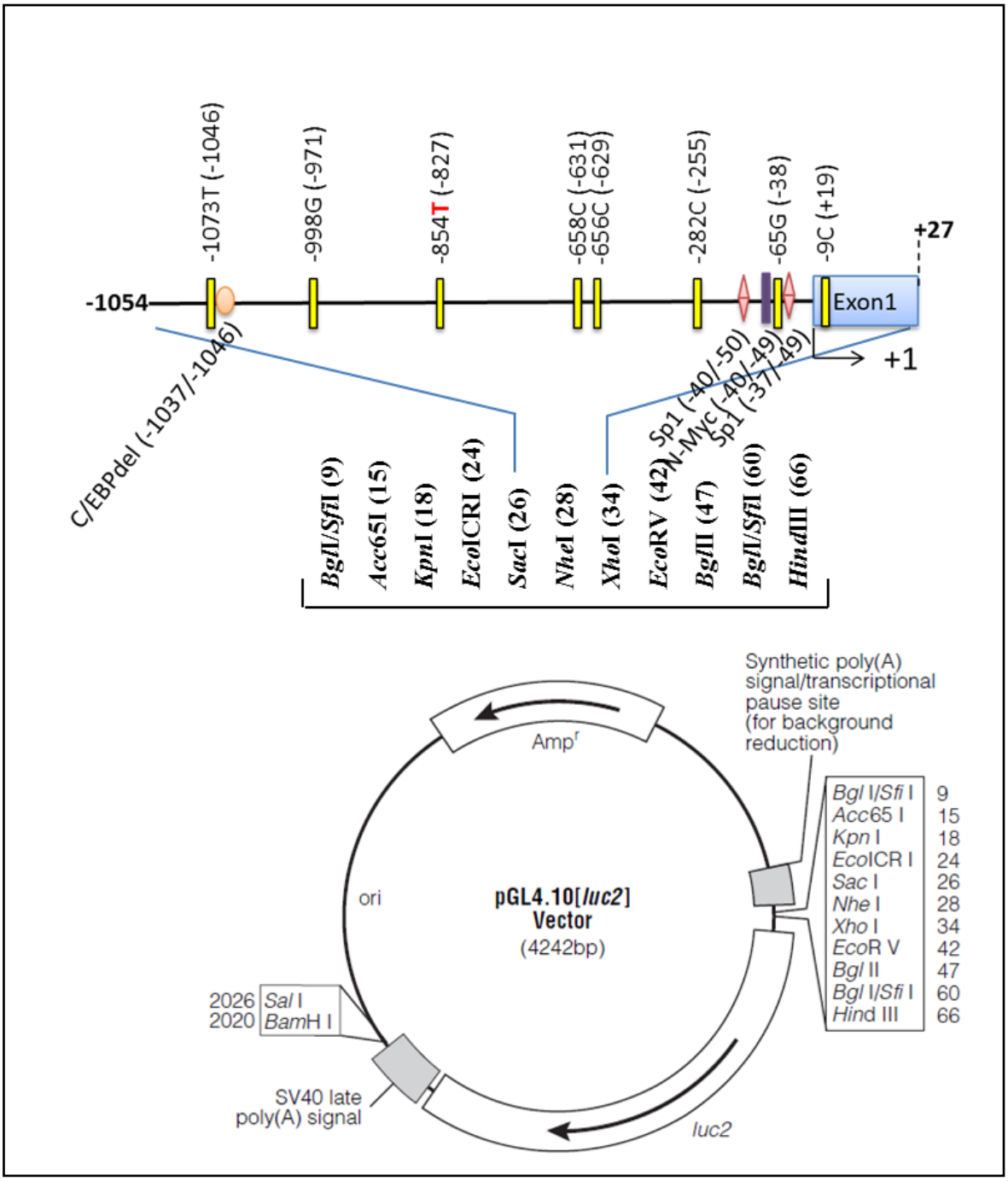

Figure 29. (-1054/+27) -854T CETP-luc. Human CETP gene fragment -1054/+27, bearing the less common allele at position -827, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

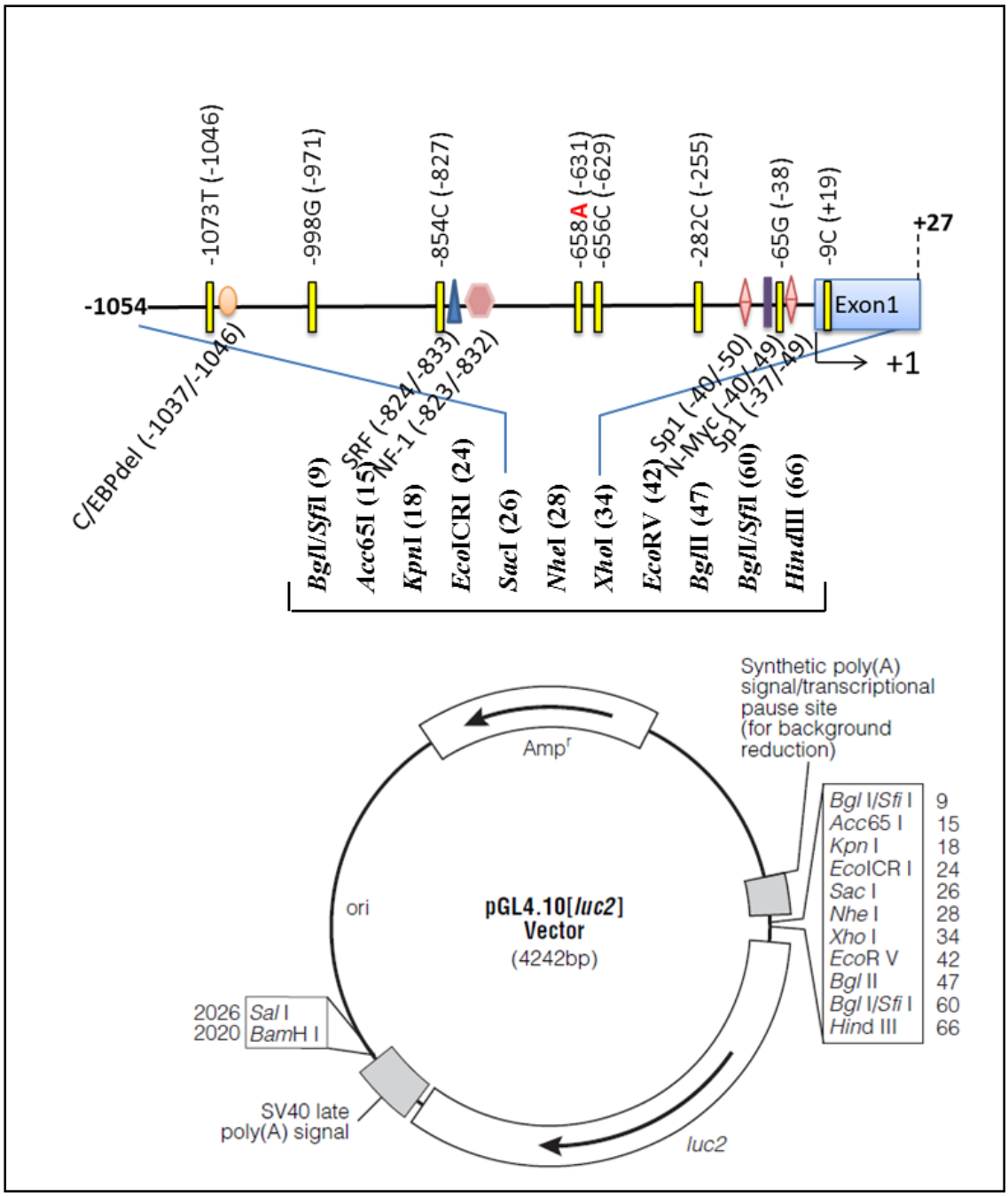

Figure 30. (-1054/+27) -658A CETP-luc. Human CETP gene fragment -1054/+27, bearing the less common allele at position -631, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

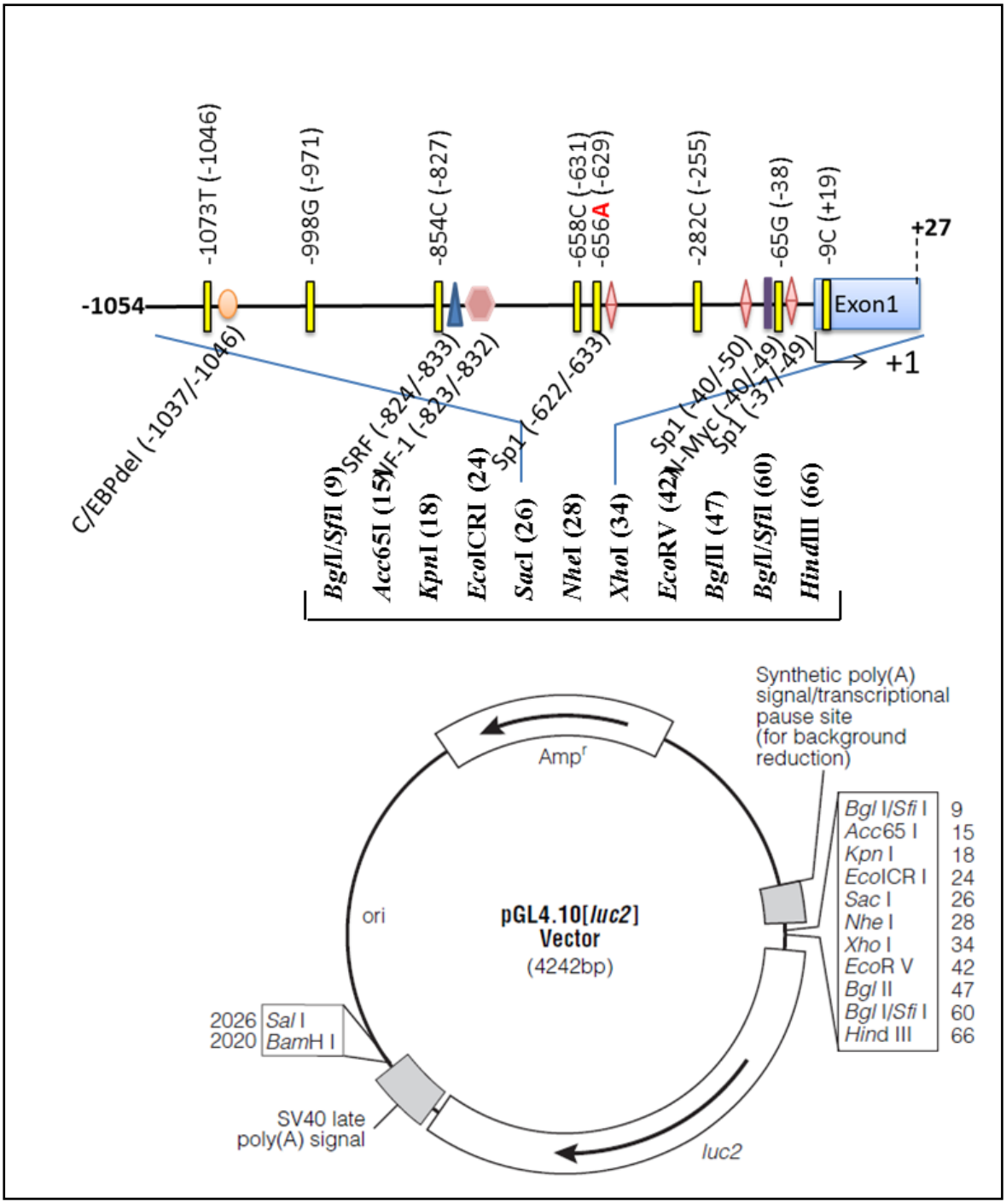

Figure 31. (-1054/+27) -656A CETP-luc. Human CETP gene fragment -1054/+27, bearing the less common allele at position -629, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

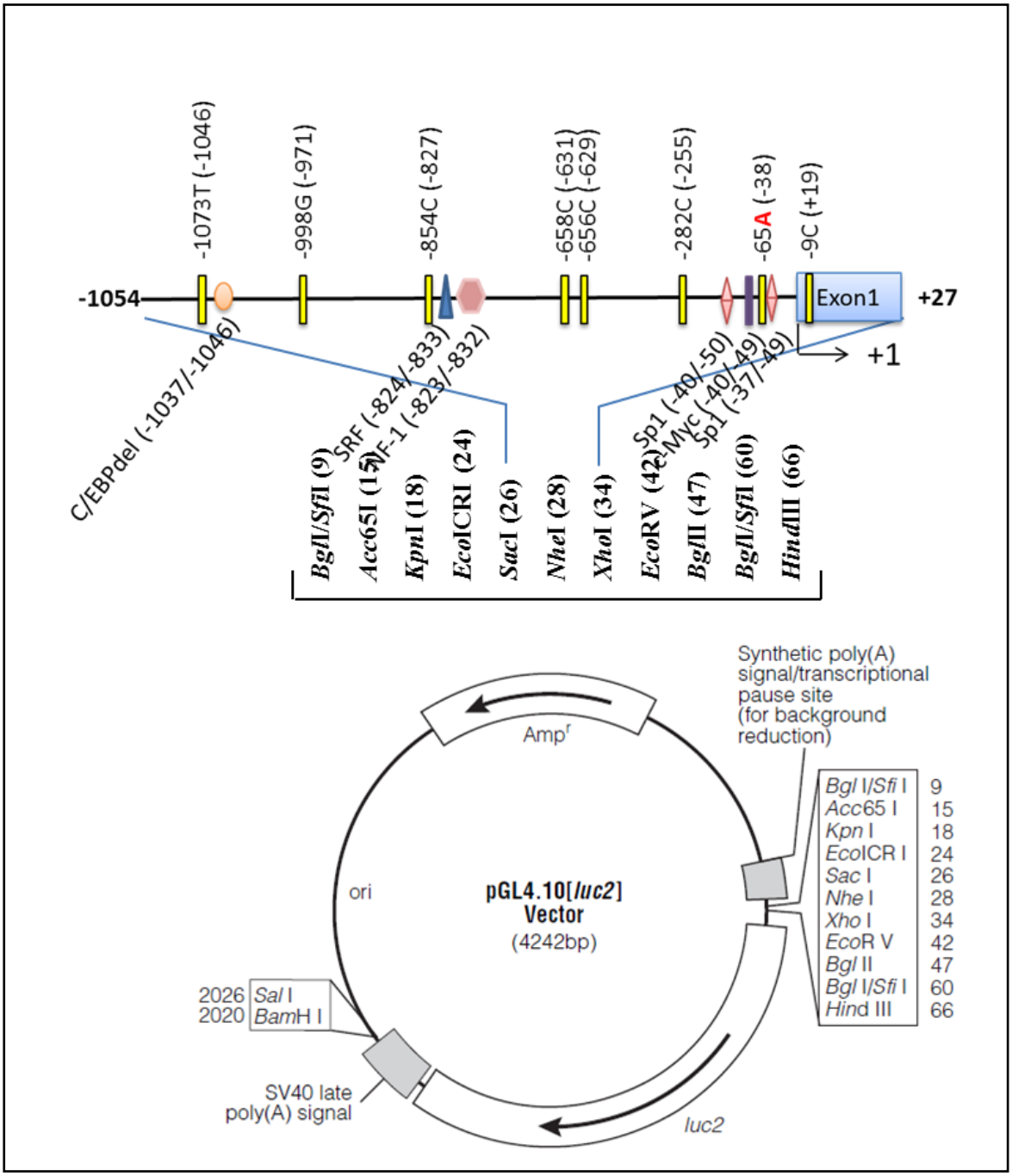

Figure 32. (-1054/+27) -65A CETP-luc. Human CETP gene fragment -1054/+27, bearing the less common allele at position -38, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

1.6.2.2. Plasmid map for CETP polymorphism found in the first exon

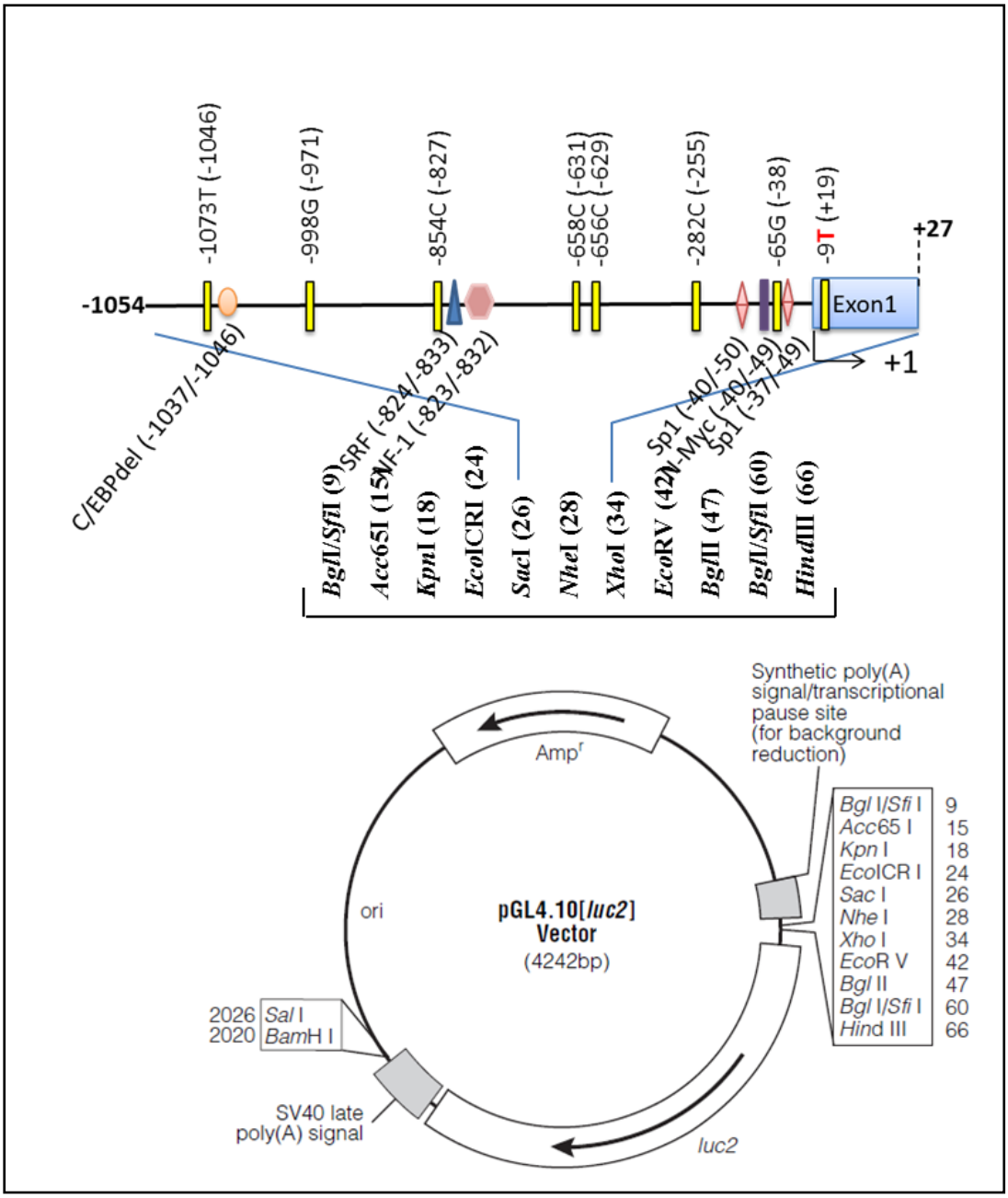

Figure 33. (-1054/+27) –9T CETP-luc. Human CETP gene fragment -1054/+27, bearing the less common allele at position +19, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

1.6.2.3. Plasmid maps for CETP polymorphisms found in the promoter and in the first intron sequence

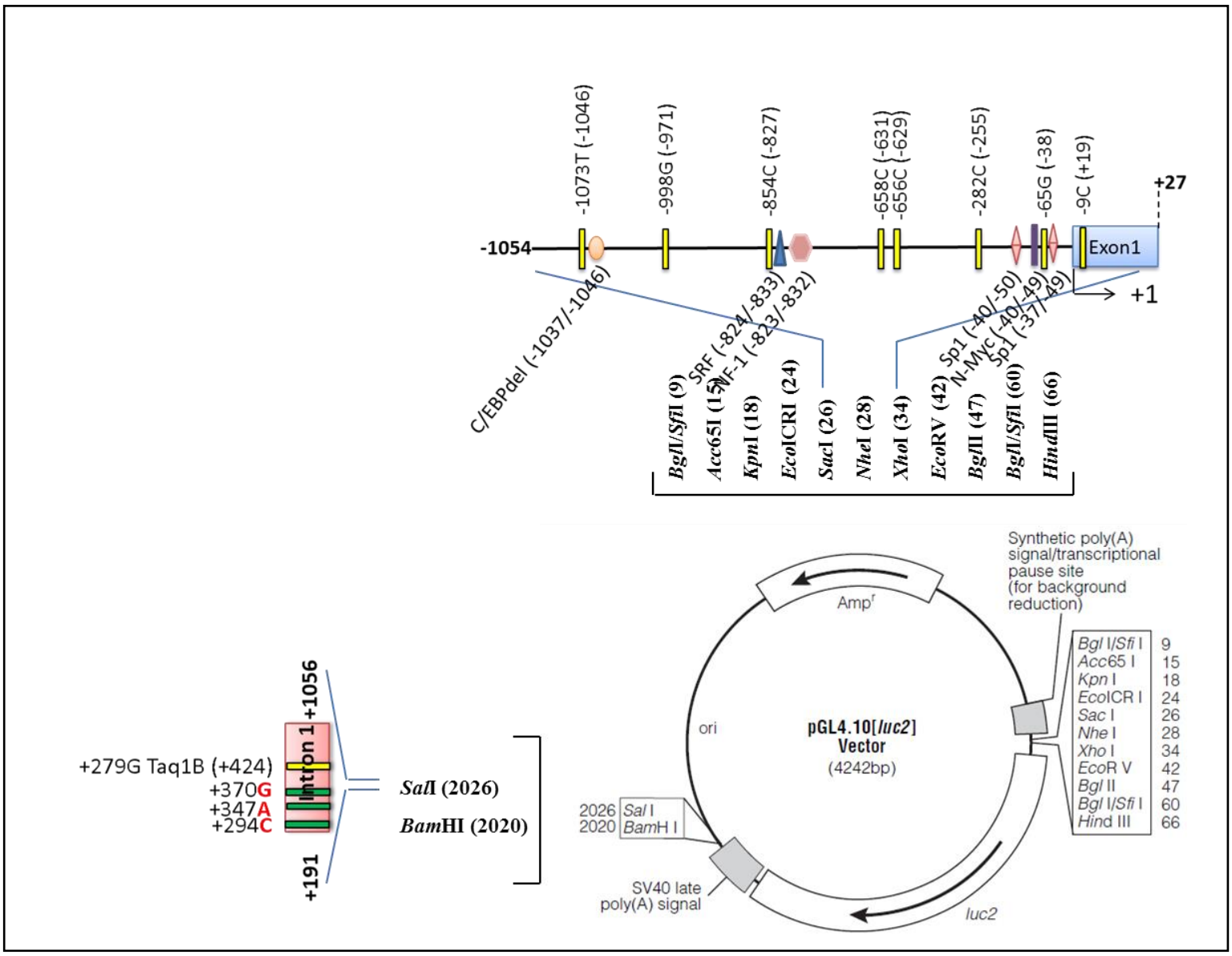

Figure 34. (-1054/+27) CETP-luc-(+191/+1056) INTRON 1 CETP. Human CETP gene fragment -1054/+27, bearing the most common alleles was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. Human CETP Intron 1 fragment +191/+1056 bearing the most common allele was inserted in the same vector using *Sal*I restriction enzyme. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. The three extra SNPs observed in the area spanning the first exon are shown (+294C, +347A and +370G). Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

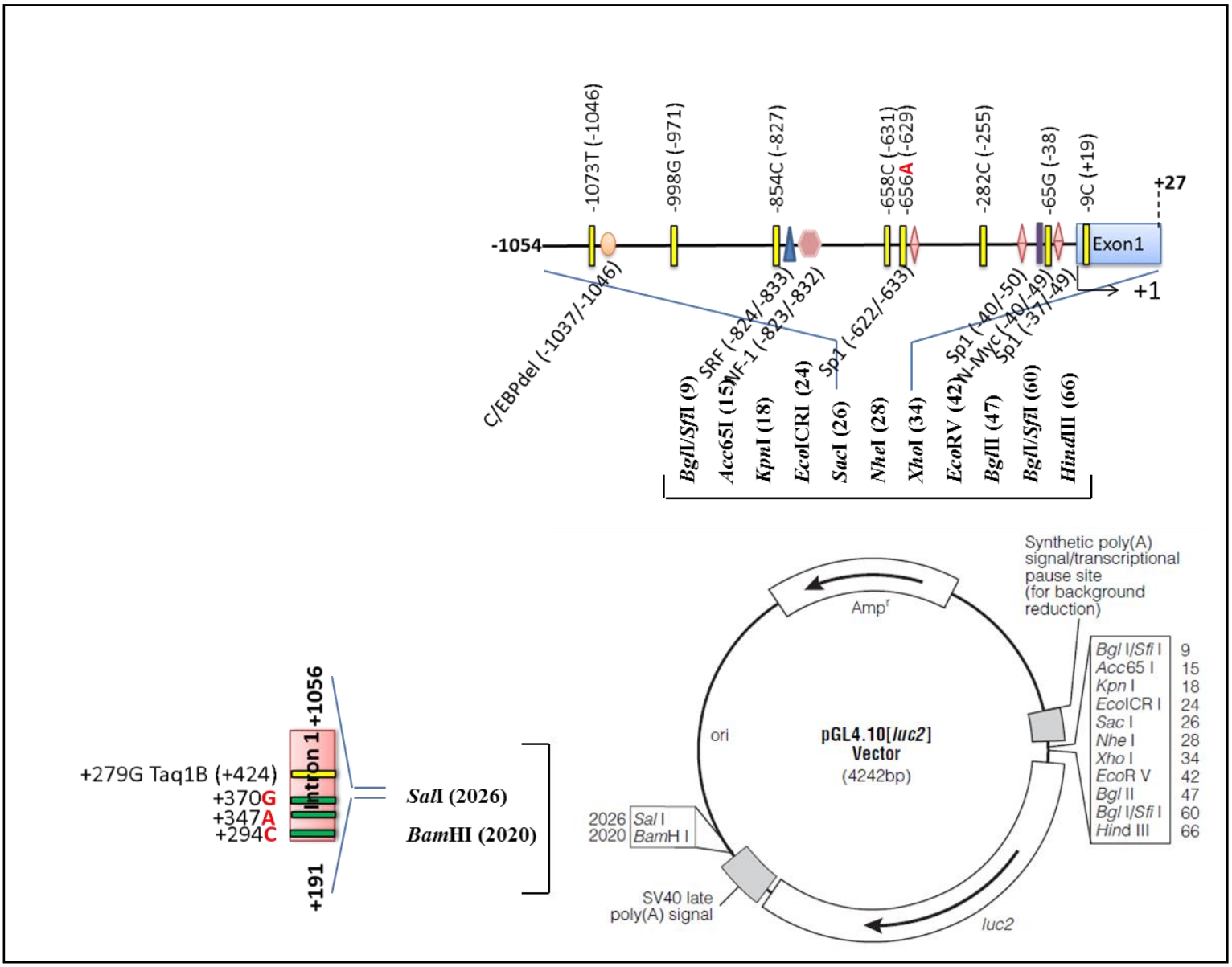

Figure 35. (-1054/+27) -656A CETP-luc-(+191/+1056) INTRON 1 CETP. Human CETP gene fragment -1054/+27, bearing the less common allele at position -629, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. Human CETP Intron 1 fragment +191/+1056 bearing the most common allele was inserted in the same vector using *Sal*I restriction enzyme. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. The three extra SNPs observed in the area spanning the first exon are shown (+294C, +347A and +370G). Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

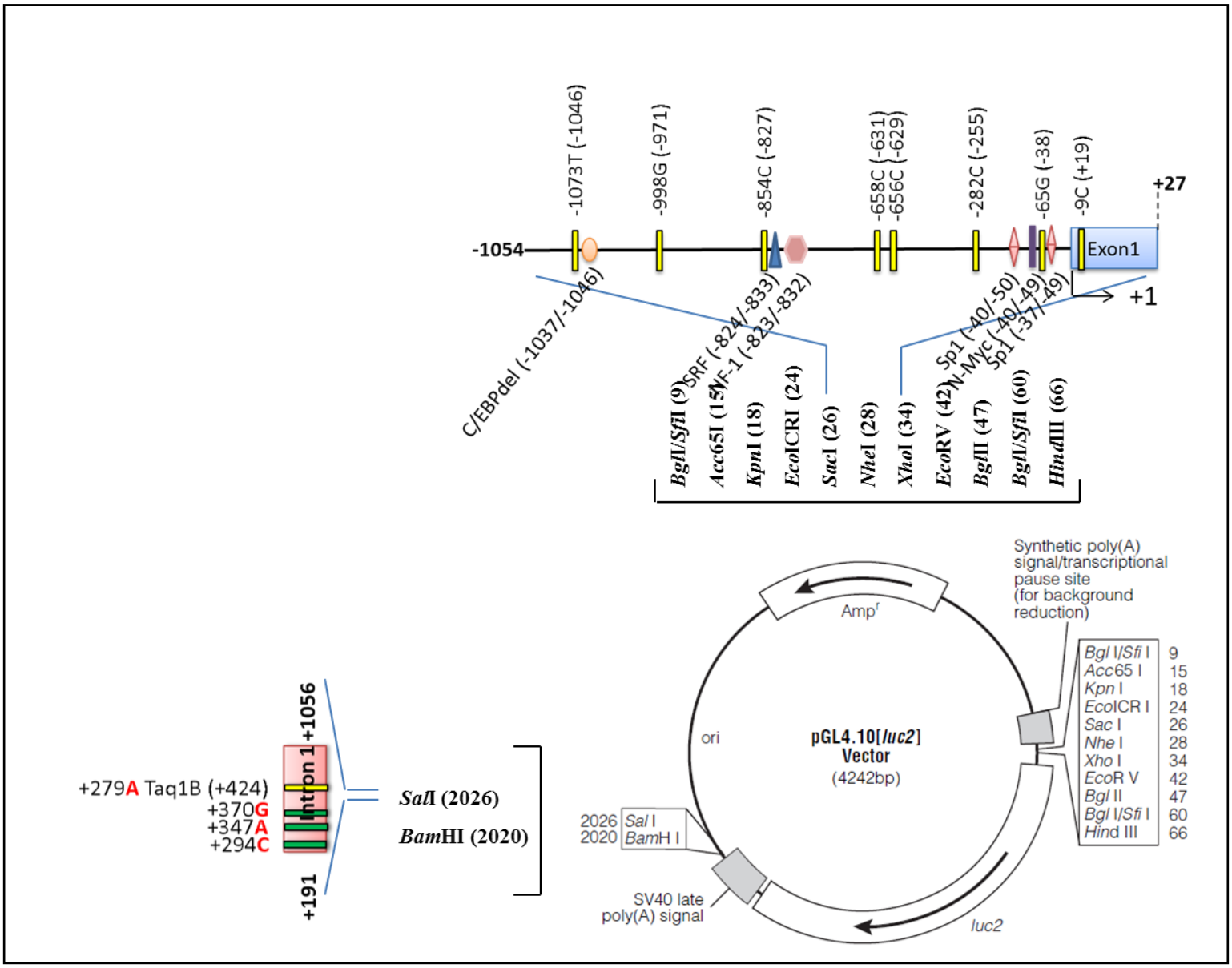

Figure 36. (-1054/+27) CETP-luc-(+191/+1056) +279A INTRON 1 CETP. Human CETP gene fragment -1054/+27, bearing the most common alleles was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. Human CETP Intron 1 fragment +191/+1056 bearing the less common allele at position +424, was inserted in the same vector using *Sal*I restriction enzyme. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. The three extra SNPs observed in the area spanning the first exon are shown (+294C, +347A and +370G). Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

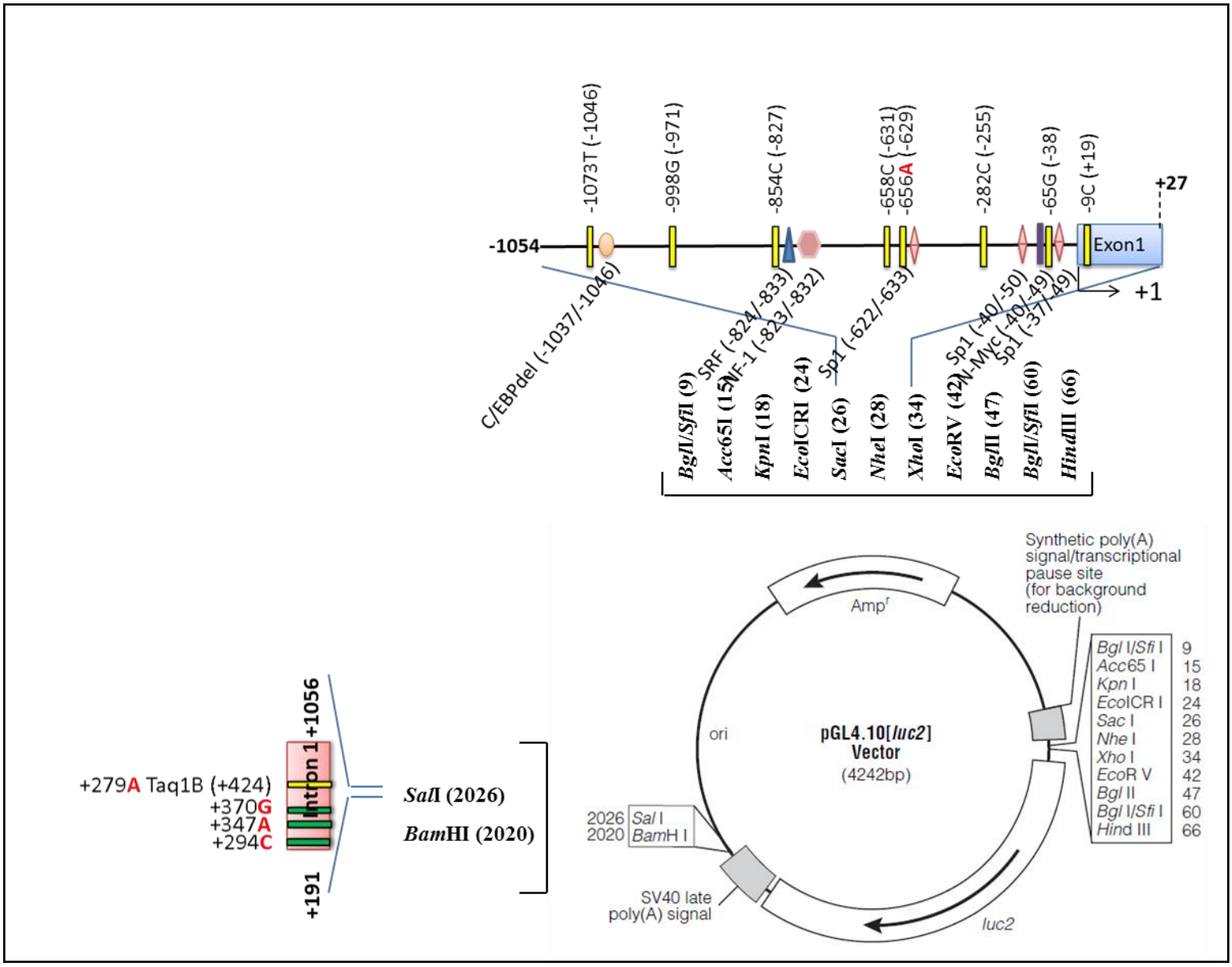

Figure 37. (-1054/+27) -656A CETP-luc-(+191/+1056) +279A INTRON 1 CETP. Human CETP gene fragment -1054/+27, bearing the less common allele at position -629, was inserted in pGL4.10 [luc2] vector using *Sac*I and *Xho*I restriction enzymes. Human CETP Intron 1 fragment +191/+1056 bearing the less common allele at position +424, was inserted in the same vector using *Sal*I restriction enzyme. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the positions polymorphisms are located in the DNA sequence of the CCHS, according to the +1 position. The three extra SNPs observed in the area spanning the first exon are shown (+294C, +347A and +370G). Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

1.6.2.4. Plasmid map for ABCA1 promoter and CETP first intron polymorphism

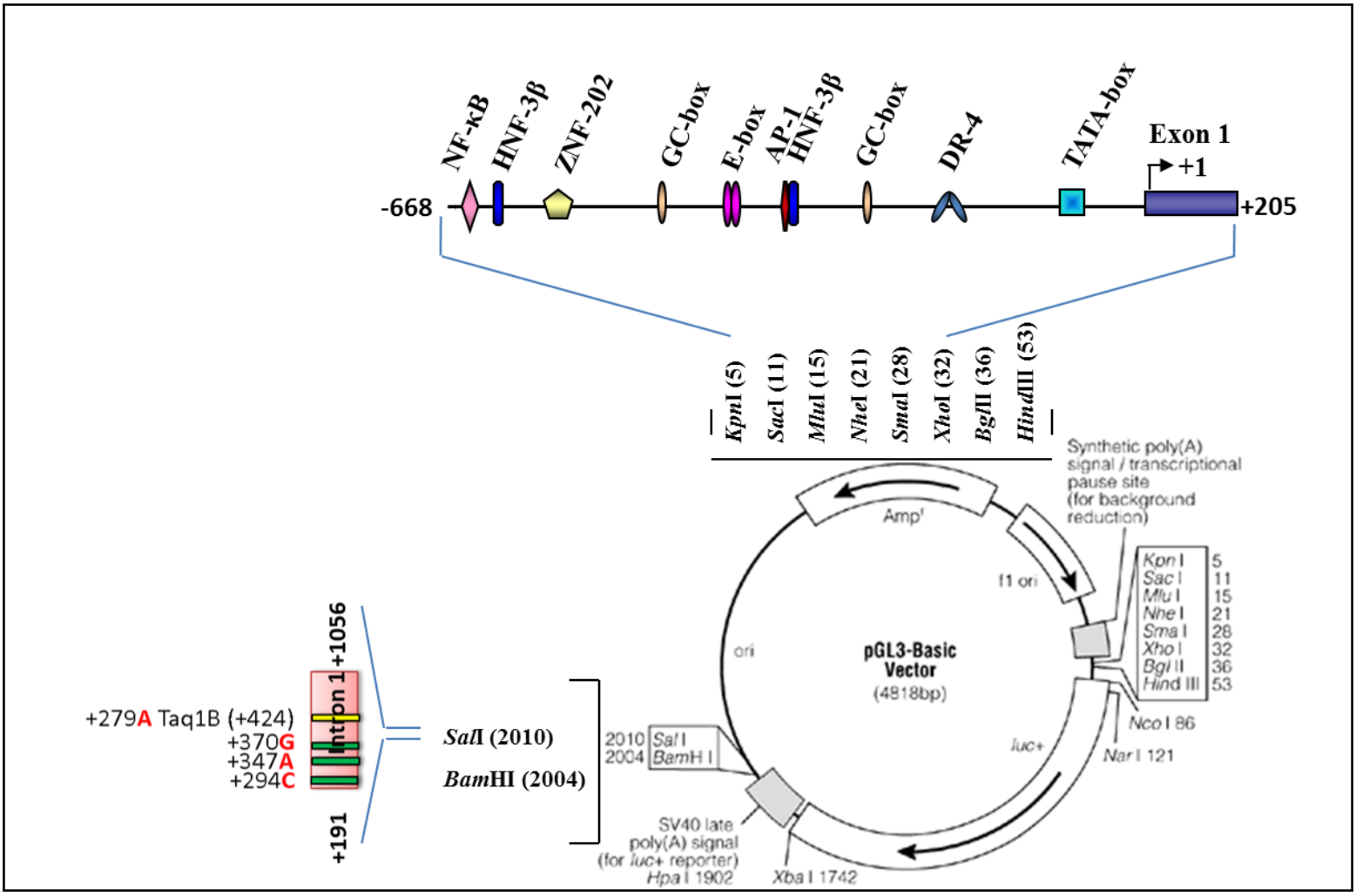

Figure 38. (-668/+205) ABCA1-luc-(+191/+1056) +279A INTRON 1 CETP. Human ABCA1 gene fragment -668/+205, was inserted in pGL3 Basic vector using *Kpn*I and *Xho*I restriction enzymes. Human CETP Intron 1 fragment +191/+1056 bearing the less common allele at position +424, was inserted in the same vector using *Sal*I restriction enzyme. The alleles present at each polymorphic position as shown. Number in the parenthesis demonstrate the position polymorphism is located in the DNA sequence of the CCHS, according to the +1 position. The three extra SNPs observed in the area spanning the first exon are shown (+294C, +347A and +370G). Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

1.6.2.5. Plasmid maps for apoE promoter polymorphisms

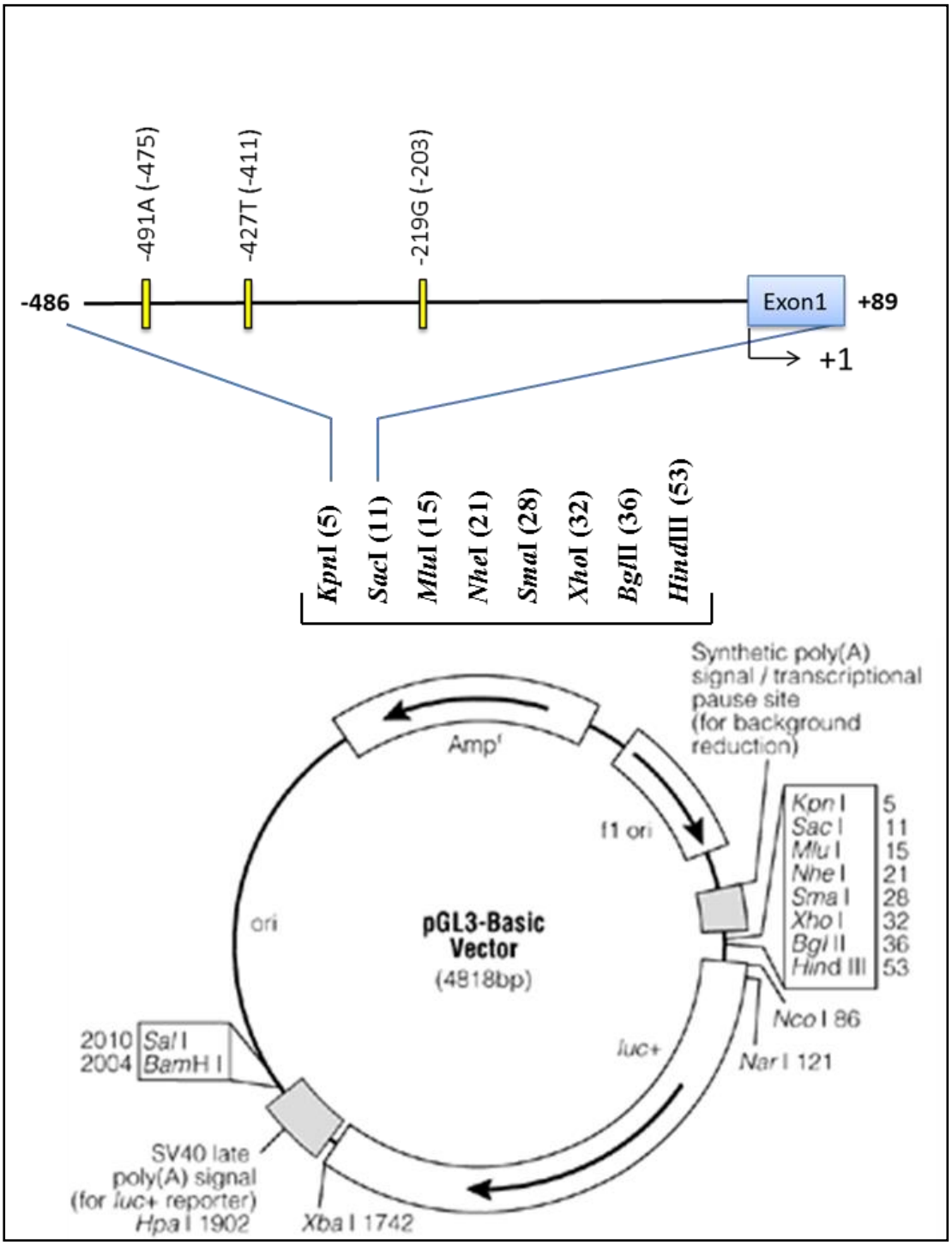

Figure 39. (-486/+89) apoE-luc. Human apoE gene fragment -486/+89, was inserted in pGL3 Basic vector using *KpnI* and *SacI* restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the position polymorphism is located in the DNA sequence of the CCHS, according to the +1 position.

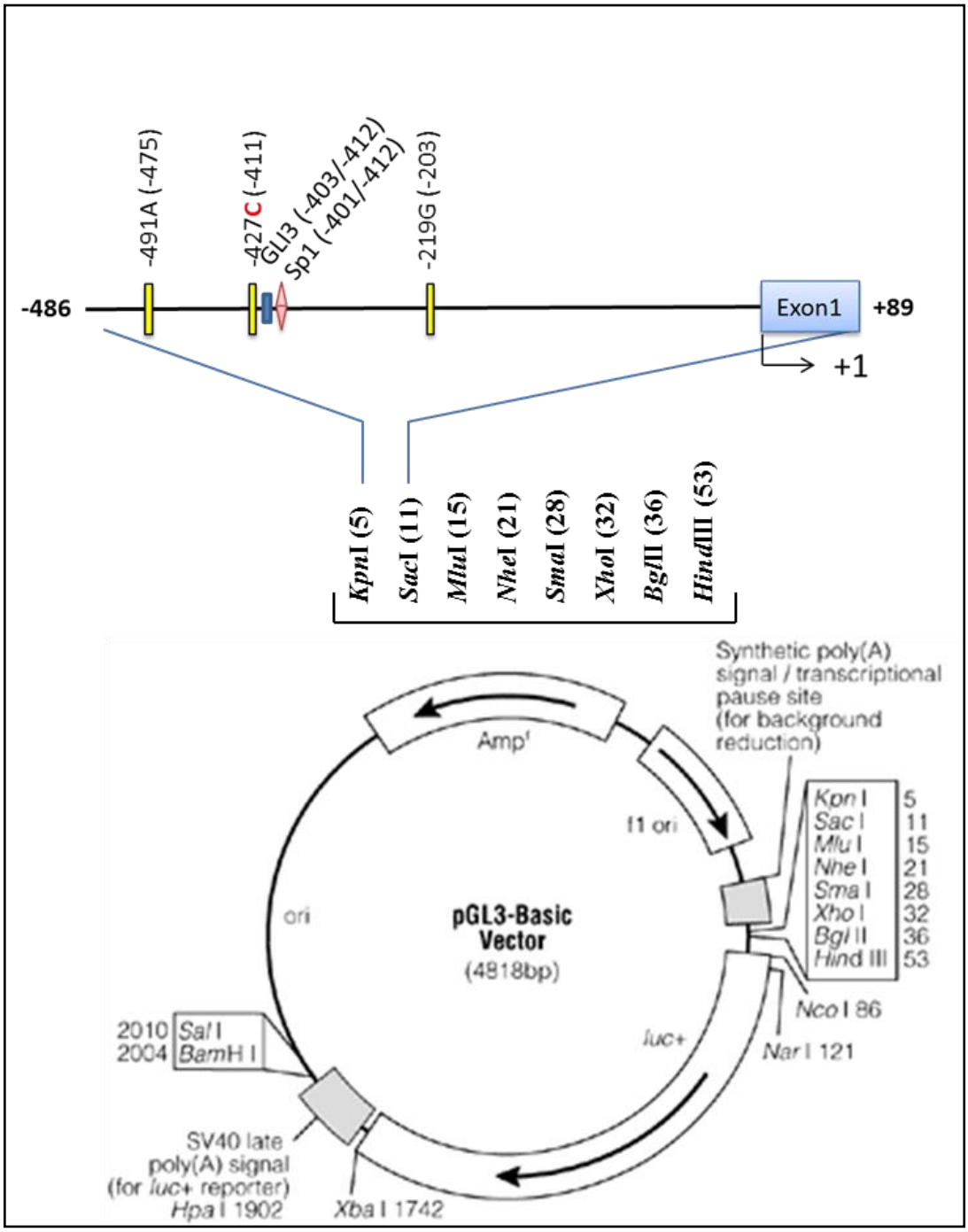

Figure 40. (-486/+89) -427C apoE-luc. Human apoE gene fragment -486/+89, bearing the less common allele at position -411, was inserted in pGL3 Basic vector using *Kpn*I and *Sac*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the position polymorphism is located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

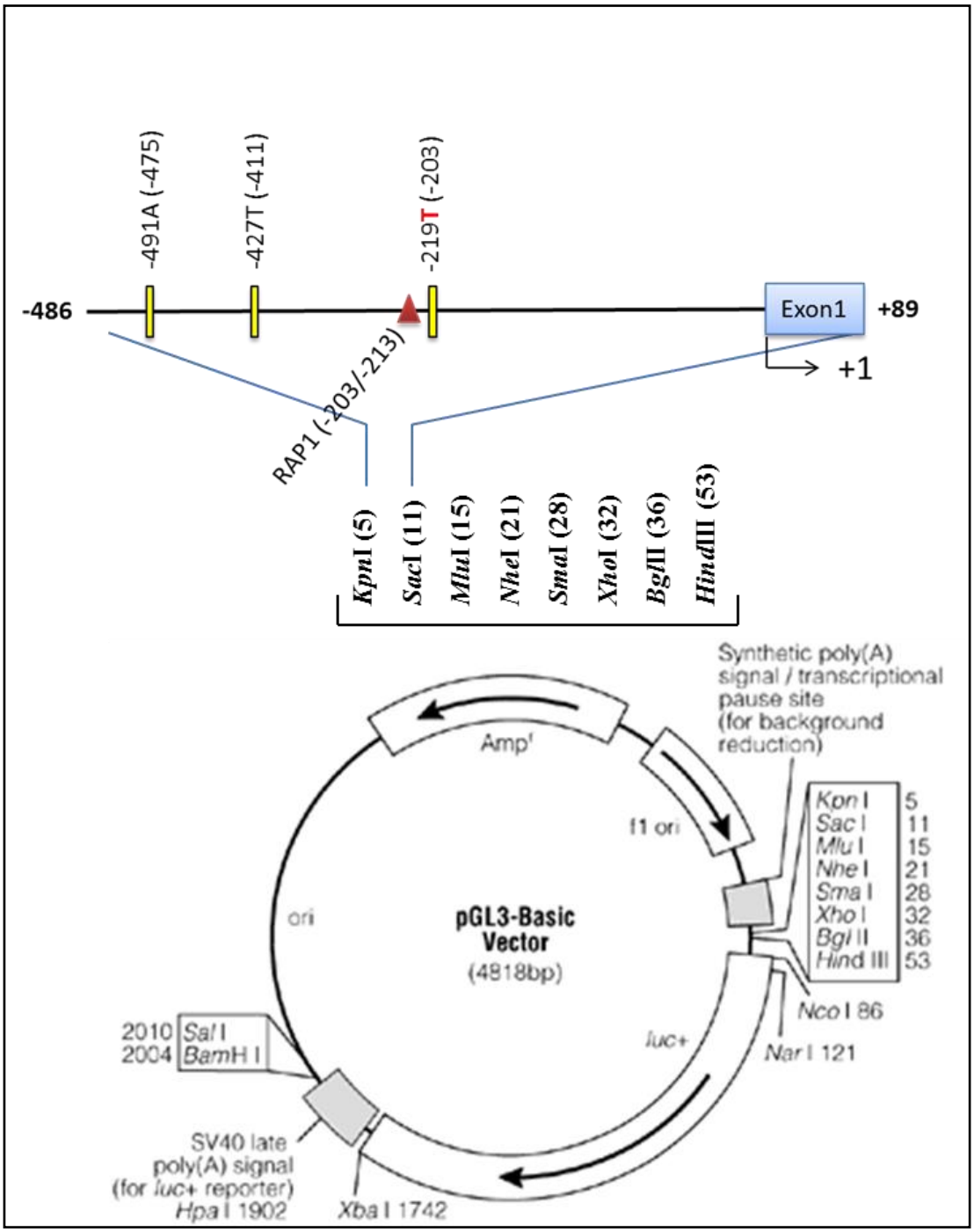

Figure 41. (-486/+89) -219T apoE-luc. Human apoE gene fragment -486/+89, bearing the less common allele at position -203, was inserted in pGL3 Basic vector using *Kpn*I and *Sac*I restriction enzymes. The alleles present at each polymorphic position as shown. Numbers in the parenthesis demonstrate the position polymorphism is located in the DNA sequence of the CCHS, according to the +1 position. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

1.6.3. TRANSFAC analysis of human CETP sequence (-1070/+1090) from the Copenhagen City Heart Study

Human CETP -1073T>C (position: 25)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
133	V$HNF3B_01	+	25	3.466.577.084.685.660	0.4841820988135506	29.823.949.858.721.100	29.823.949.858.721.100
624	V$E2F4DP2_01	+	25	-6.027.382.871.988.900	1.507.460.365.931.380	-75.348.432.379.202.800	75.348.432.379.202.800
1057	V$ESE1_Q3	+	25	33.339.469.610.193.300	5.596.981.366.853.120	-22.630.344.058.337.900	22.630.344.058.337.900
1218	V$RHOX11_01	+	25	2.250.449.043.450.660	0.028056622114213448	22.223.924.213.364.400	22.223.924.213.364.400
1485	V$MZF1_Q5	+	25	13.699.605.002.892.000	-19.519.675.945.981.500	33.219.280.948.873.600	33.219.280.948.873.600
1647	V$EHF_02	+	25	28.216.031.782.278.900	5.582.415.514.348.460	-27.608.123.361.205.700	27.608.123.361.205.700
1658	V$ESE1_01	+	25	33.528.549.685.078.900	5.285.740.772.649.360	-19.328.858.041.414.600	19.328.858.041.414.600
1813	V$EHF_03	+	25	32.030.786.591.752.300	5.261.972.348.228.800	-20.588.936.890.535.600	20.588.936.890.535.600
1825	V$ESE1_02	+	25	33.339.469.610.193.300	5.596.981.366.853.120	-22.630.344.058.337.900	22.630.344.058.337.900
3928	V$MAX_10	+	25	-1.747.470.739.395.750	11.302.317.974.907.500	-28.777.025.368.865.100	28.777.025.368.865.100
4415	V$ZNF85_02	+	25	12.277.541.961.396.200	-29.015.288.208.053.400	41.292.830.169.449.600	41.292.830.169.449.600
4871	V$PAX5_11	+	25	-13.708.043.327.976.400	0.2526323157381454	-16.234.366.485.357.900	16.234.366.485.357.900
5007	V$NF1C_Q6	+	25	4.174.961.170.099.540	24.308.000.745.291.300	17.441.610.955.704.000	17.441.610.955.704.000

Table 10. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism -1073T>C of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP -998G>A (position: 100)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
746	V$GR_Q6_01	+	100	23.263.386.422.054.000	4.156.754.319.723.280	-18.304.156.775.178.700	18.304.156.775.178.700
2018	V$SP1_03	+	100	22.994.093.661.681.200	-0.8160678512518142	31.154.772.174.199.300	31.154.772.174.199.300
2705	V$PEA3_Q6_02	+	100	-0.712291730962457	19.315.644.588.122.600	-26.438.561.897.747.200	26.438.561.897.747.200
3170	V$TFAP2A_02	+	100	-1.944.306.048.451.150	0.026207187048463143	-19.705.132.354.996.100	19.705.132.354.996.100
3176	V$TFAP2B_02	+	100	0.15880771368022795	215.077.039.461.481	-19.919.626.809.345.800	19.919.626.809.345.800
3189	V$TFAP2C_03	+	100	1.035.270.069.722.620	28.673.254.324.085.500	-18.320.553.626.859.200	18.320.553.626.859.200
3708	V$RHOXF1_02	+	100	0.7900255737960957	-1.132.748.914.378.070	19.227.744.881.741.700	19.227.744.881.741.700

Table 11. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism -998G>A of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP -854C>T (position: 244)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
342	V$PAX3_B	+	244	30.345.165.089.831.100	4.619.479.009.704.270	-15.849.625.007.211.500	15.849.625.007.211.500
999	V$DMRT7_01	+	244	0.3342172261325733	-29.137.102.873.110.100	32.479.275.134.435.800	32.479.275.134.435.800
1317	V$LHX4_01	+	244	-16.593.065.123.482.200	0.2282187583933626	-18.875.252.707.415.800	18.875.252.707.415.800
1502	V$ERG_01	+	244	0.3025152736085637	-1.504.839.648.449.040	18.073.549.220.576.000	18.073.549.220.576.000
2712	V$TBP_Q6_01	+	244	0.9801728441247826	26.682.288.378.100.400	-16.880.559.936.852.500	16.880.559.936.852.500
2714	V$TWIST_Q6	+	244	0.10998675867383101	-24.749.757.420.473.200	25.849.625.007.211.500	25.849.625.007.211.500
2903	REVERBALPHA_	+	244	-2.188.452.193.834.770	0.3965103068863834	-25.849.625.007.211.500	25.849.625.007.211.500
3677	V$PHOX2A_02	+	244	-3.014.205.868.163.010	0.7944113647992415	-38.086.172.329.622.500	38.086.172.329.622.500
3688	V$PROP1_04	+	244	-25.570.455.308.787.900	10.500.321.312.237.100	-36.070.776.621.025.100	36.070.776.621.025.100

Table 12. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism -854C>T of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP -658C>A (position: 440)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
154	V$LYF1_01	+	440	1.053.844.347.828.520	-5.795.753.178.754.380	16.334.196.657.039.600	16.334.196.657.039.600
915	V$YY1_Q6_02	+	440	26.297.447.951.305.100	0.6822122150246539	19.475.325.801.058.600	19.475.325.801.058.600
1072	V$CMAF_02	+	440	-1.593.014.389.029.930	0.325645413481062	-16.255.789.303.780.400	16.255.789.303.780.400
1457	V$MAFA_Q4	+	440	-1.463.399.560.984.780	18.280.982.588.593.000	-16.462.093.868.707.100	16.462.093.868.707.100
1541	V$CPBP_Q6	+	440	454.527.586.183.155	-14.057.864.355.123.200	18.603.140.216.954.800	18.603.140.216.954.800
1952	V$NKX2B_Q3	+	440	0.6019744578439783	-19.031.739.841.514.600	19.633.714.299.358.600	19.633.714.299.358.600
2017	V$SP1_03	+	440	14.196.010.505.853.300	-18.374.150.974.527.300	19.793.752.025.112.600	19.793.752.025.112.600
4034	V$NR3C1_08	+	440	0.5836744517704784	-1.634.931.879.657.840	16.932.993.248.348.900	16.932.993.248.348.900

Table 13. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism –658C>A of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP -656C>A (position: 442)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
1743	V$HNF4A_Q6_01	+	442	27.959.375.539.708.100	391.141.477.139.075	-11.154.772.174.199.300	11.154.772.174.199.300
2872	V$CP2_Q6	+	442	36.358.458.262.701.700	4.635.845.826.270.170	-10.000.000.000.000.000	10.000.000.000.000.000
3904	V$RAD21_02	+	442	6.424.440.341.564.750	2.251.247.626.288.530	41.731.927.152.762.200	41.731.927.152.762.200

Table 14. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism –656C>A of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP -282C>A (position: 816)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
15	V$CETS1P54_01	+	816	3.571.623.275.050.280	5.378.978.197.107.890	-18.073.549.220.576.000	18.073.549.220.576.000
277	V$XFD2_01	+	816	2.657.208.224.064.670	4.236.399.936.927.870	-15.791.917.128.632.000	15.791.917.128.632.000
995	V$DMRT5_01	+	816	4.844.256.535.341.080	7.651.611.457.398.690	-28.073.549.220.576.000	28.073.549.220.576.000
1665	V$ELF4_01	+	816	4.054.816.933.958.450	25.243.022.172.596.600	15.305.147.166.987.800	15.305.147.166.987.800
1674	V$ELK1_05	+	816	0.7299655865256915	3.314.928.087.246.840	-25.849.625.007.211.500	25.849.625.007.211.500
1682	V$ERF_01	+	816	4.029.083.064.435.060	6.054.618.156.542.190	-20.255.350.921.071.300	20.255.350.921.071.300
1695	V$ER71_01	+	816	-0.8950850591491684	13.734.037.767.767.300	-22.684.888.359.259.000	22.684.888.359.259.000
1697	V$CETS2_01	+	816	-16.443.181.809.339.700	0.9671165311483725	-26.114.347.120.823.400	26.114.347.120.823.400
1712	V$TEL1_01	+	816	6.394.271.978.925.240	4.836.276.525.804.350	15.579.954.531.208.800	15.579.954.531.208.800
1771	V$PET1_01	+	816	-13.019.409.510.373.100	1.335.488.969.577.970	-26.374.299.206.152.900	26.374.299.206.152.900
1777	V$GABPA_01	+	816	-11.773.186.687.705.700	1.460.111.251.844.710	-26.374.299.206.152.900	26.374.299.206.152.900
1795	V$SPI1_02	+	816	39.757.757.865.836.500	5.683.595.035.090.340	-17.078.192.485.066.800	17.078.192.485.066.800
1842	V$ERF_02	+	816	19.306.771.025.982.500	4.981.303.175.668.210	-30.506.260.730.699.600	30.506.260.730.699.600
1846	V$ERG_03	+	816	-3.107.626.589.910.550	0.41593536614645565	-35.235.619.560.570.100	35.235.619.560.570.100
1856	V$ER81_02	+	816	25.115.211.634.109.800	48.334.492.582.983.400	-23.219.280.948.873.600	23.219.280.948.873.600
1862	V$ER71_02	+	816	-0.8715541092009869	12.439.231.082.189.400	-21.154.772.174.199.300	21.154.772.174.199.300
1871	V$TEL1_02	+	816	6.394.271.978.925.240	4.836.276.525.804.350	15.579.954.531.208.800	15.579.954.531.208.800
1878	V$PET1_02	+	816	-13.241.826.182.506.000	114.712.310.067.498	-24.713.057.189.255.800	24.713.057.189.255.800
3008	V$ELK3_01	+	816	-23.361.976.418.426.500	0.5914795543800972	-29.276.771.962.227.500	29.276.771.962.227.500
3011	V$ELK4_02	+	816	-34.275.015.139.550.100	0.16084362616079523	-35.883.451.401.158.000	35.883.451.401.158.000
3018	V$ETS1_02	+	816	-2.045.223.003.428.900	0.799400303697934	-28.446.233.071.268.400	28.446.233.071.268.400
3021	V$ETV1_01	+	816	-1.289.405.667.771.250	19.874.651.521.307.700	-32.768.708.199.020.200	32.768.708.199.020.200
3027	V$ETV4_01	+	816	-16.671.963.842.548.900	12.735.011.554.949.200	-29.406.975.397.498.200	29.406.975.397.498.200
3030	V$ETV5_01	+	816	-1.865.293.175.250.950	0.4910506990013907	-23.563.438.742.523.400	23.563.438.742.523.400
3265	V$FOXC1_04	+	816	6.814.684.612.948.090	9.877.758.298.835.650	-30.630.736.858.875.600	30.630.736.858.875.600

Table 15. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism –282C>A of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP -65G>A (position: 1033)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
59	V$GATA2_01	+	1033	27.512.449.099.158.400	1.166.282.409.194.680	15.849.625.007.211.500	15.849.625.007.211.500
165	V$SRY_01	+	1033	22.824.998.309.557.700	-0.7175001690442279	30.000.000.000.000.000	30.000.000.000.000.000
629	V$POU1F1_Q6	+	1033	-2.072.547.967.221.400	0.23958054321112088	-23.121.285.104.325.200	23.121.285.104.325.200
1199	V$HOXA7_02	+	1033	0.579361470910678	-13.946.433.205.563.700	19.740.047.914.670.500	19.740.047.914.670.500
1262	V$IRXB3_01	+	1033	5.182.665.704.794.650	8.807.156.569.702.440	-36.244.908.649.077.900	36.244.908.649.077.900
1296	V$IRX2_01	+	1033	38.017.576.066.942.500	5.896.914.839.734.590	-20.951.572.330.403.400	20.951.572.330.403.400
1378	V$IRX5_01	+	1033	3.945.557.040.485.810	59.779.785.181.781.900	-20.324.214.776.923.800	20.324.214.776.923.800
1399	V$IRX3_02	+	1033	4.422.859.607.208.110	6.079.971.893.685.100	-16.571.122.864.769.900	16.571.122.864.769.900
2639	V$GATA4_Q5_01	+	1033	-2.298.565.601.570.080	13.452.905.882.046.300	-36.438.561.897.747.200	36.438.561.897.747.200
3110	V$POU2F3_04	+	1033	-1.709.772.078.643.140	17.698.374.219.317.800	-34.796.095.005.749.200	34.796.095.005.749.200
3115	V$POU3F1_03	+	1033	13.778.565.894.995.700	3.947.584.576.525.560	-25.697.279.870.259.900	25.697.279.870.259.900
3120	V$POU3F3_02	+	1033	0.854643953409615	35.274.020.586.257.800	-26.727.581.052.161.700	26.727.581.052.161.700
3123	V$POU3F3_03	+	1033	0.3248226690941429	36.049.305.882.868.700	-32.801.079.191.927.300	32.801.079.191.927.300
4541	V$ZNF555_02	+	1033	1.047.109.014.487.900	4.734.124.550.712.840	57.369.655.941.662.000	57.369.655.941.662.000

Table 16. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism –65G>A of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP -9C>T (position: 1089)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
1144	V$SOX_01	+	1089	13.892.889.336.645.900	-239.739.010.949.026	37.866.790.431.548.500	37.866.790.431.548.500
1417	V$TCF3_01	+	1089	-17.537.956.682.519.500	0.12067344966418603	-18.744.691.179.161.400	18.744.691.179.161.400
2124	V$SOX13_03	+	1089	-15.151.922.727.024.600	0.35927684521367564	-18.744.691.179.161.400	18.744.691.179.161.400
2130	V$SOX15_03	+	1089	-10.541.759.642.520.700	0.6303222100199921	-16.844.981.742.720.700	16.844.981.742.720.700
2143	V$SOX8_03	+	1089	0.469489940974508	21.475.618.460.871.400	-16.780.719.051.126.300	16.780.719.051.126.300
2320	V$GMEB1_04	+	1089	-0.6774165007798311	12.894.166.352.849.600	-19.668.331.360.648.000	19.668.331.360.648.000
2863	V$CFOSCJUN_Q5	+	1089	0.7325285762286228	-18.524.339.244.925.300	25.849.625.007.211.500	25.849.625.007.211.500
3057	V$IRF5_05	+	1089	0.9961276882050905	-10.124.343.252.983.300	20.085.620.135.034.200	20.085.620.135.034.200
3094	V$POU1F1_04	+	1089	32.118.157.696.230.100	16.800.952.901.788.300	15.317.204.794.441.700	15.317.204.794.441.700
3264	V$FOXC1_04	+	1089	-29.724.320.288.081.500	0.990419035325095	-39.628.510.641.332.400	39.628.510.641.332.400
3284	V$FOXJ2_04	+	1089	-25.132.560.930.166.300	0.3347408135383141	-28.479.969.065.549.400	28.479.969.065.549.400
3290	V$FOXJ3_07	+	1089	0.2540349297631357	34.776.835.070.700.000	-32.236.485.773.068.700	32.236.485.773.068.700
3350	V$BSX_03	+	1089	27.300.437.361.071.800	0.9749726265567863	17.550.711.095.504.000	17.550.711.095.504.000
3527	V$HOXC10_04	+	1089	39.677.000.465.681.300	234.646.215.945.543	16.212.378.871.127.000	16.212.378.871.127.000
3654	V$NKX61_07	+	1089	0.5074618024559414	38.595.518.678.535.500	-33.520.900.653.976.100	33.520.900.653.976.100
3660	V$OTX1_04	+	1089	0.6745496817771356	-18.259.556.685.916.200	25.005.053.503.687.600	25.005.053.503.687.600
4748	V$POU2F1_Q4_01	+	1089	22.333.263.904.995.900	0.06340138905728587	21.699.250.014.423.100	21.699.250.014.423.100

Table 17. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism –9C>T of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP +279G>A Taq1B (position: 1494)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
70	V$MZF1_02	+	1494	0.9056493303228528	-0.901705591734751	18.073.549.220.576.000	18.073.549.220.576.000
956	V$MOVOB_01	+	1494	11.573.814.144.652.900	-14.275.810.862.558.600	25.849.625.007.211.500	25.849.625.007.211.500
2062	V$GLIS2_03	+	1494	0.1896090980504278	18.270.390.186.657.100	-16.374.299.206.152.900	16.374.299.206.152.900
2262	V$EHF_07	+	1494	-0.41522292328289717	1.648.907.414.136.810	-20.641.303.374.197.100	20.641.303.374.197.100
4756	V$POU4F3_Q5	+	1494	-2.083.248.824.464.590	0.22716068595449634	-23.104.095.104.190.900	23.104.095.104.190.900

Table 18. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism +279G>A TaqIB of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP Intron 1 (+191/+1056) (position: 1261-2126)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
574	V$LXR_Q3	+	1477	-0.4917198221516732	2.305.747.944.994.560	-27.974.677.671.462.400	27.974.677.671.462.400
2176	V$ZBTB3_03	+	1478	4.406.825.338.024.390	27.369.739.397.167.200	16.698.513.983.076.600	16.698.513.983.076.600
4872	V$PAX5_11	+	1478	0.065227882428211 54	-1.485.125.581.362.800	15.503.534.637.910.100	15.503.534.637.910.100
2319	V$GMEB1_04	+	1479	-10.190.002.780.779.300	0.9478328579868643	-19.668.331.360.648.000	19.668.331.360.648.000
2544	V$TCFAP2B_04	+	1480	0.9730339686900575	25.090.868.689.302.600	-15.360.529.002.402.000	15.360.529.002.402.000
2577	V$ZFP281_05	+	1480	-33.344.658.505.253.800	0.6197304598614939	-39.541.963.103.868.700	39.541.963.103.868.700
2400	V$PLAGL1_04	+	1482	280.755.682.494.871	4.682.025.942.864.850	-18.744.691.179.161.400	18.744.691.179.161.400
4004	V$NANOG_05	+	1482	-0.839638717039815	21.901.086.263.542.300	-30.297.473.433.940.500	30.297.473.433.940.500
1256	V$SIX4_01	+	1483	-27.286.041.545.515.700	2.241.022.196.404.910	-49.696.263.509.564.800	49.696.263.509.564.800
2081	V$IRF6_03	+	1483	0.6380665795523686	-32.847.655.599.251.700	39.228.321.394.775.400	39.228.321.394.775.400
2100	V$NR2F2_03	+	1483	-2.507.393.819.859.820	12.054.210.388.432.400	-37.128.148.587.030.600	37.128.148.587.030.600
2272	V$ESRRA_04	+	1483	0.1087557821192537	38.061.789.849.787.200	-36.974.232.028.594.600	36.974.232.028.594.600
2978	V$ZNF524_02	+	1483	13.671.603.080.752.200	11.031.248.579.689.000	26.403.545.010.631.700	26.403.545.010.631.700
458	V$FAC1_01	+	1484	1.450.956.395.480.900	-2.455.934.200.127.600	39.068.905.956.085.100	39.068.905.956.085.100
2226	V$ASCL2_04	+	1484	25.653.363.423.823.400	-0.7972337370023654	33.625.700.793.847.000	33.625.700.793.847.000
2466	V$SOX17_04	+	1484	-2.071.232.699.005.190	0.5917323137172374	-26.629.650.127.224.200	26.629.650.127.224.200
2737	V$PR_Q6	+	1484	-7.102.969.731.960.530	0.2032119071698844 6	-73.061.816.391.304.100	73.061.816.391.304.100
2968	V$ZBTB7C_01	+	1484	4.738.857.730.278.420	1.528.695.597.727.390	32.101.621.325.510.300	32.101.621.325.510.300
719	V$PAX_Q6	+	1485	24.586.040.650.957.100	636.549.466.070.423	-39.068.905.956.085.100	39.068.905.956.085.100
1998	V$PPARGRXRA_01	+	1485	-0.5694493238307157	18.102.768.596.526.400	-23.797.261.834.833.600	23.797.261.834.833.600
2085	V$ISGF3G_03	+	1486	6.665.874.696.836.890	3.893.285.192.939.960	27.725.895.038.969.200	27.725.895.038.969.200
4458	V$ZNF124_01	+	1486	0.3017135122216259	-34.609.512.896.366.300	37.626.648.018.582.500	37.626.648.018.582.500
4490	V$ZNF836_03	+	1486	7.612.897.367.800.920	5.838.744.252.444.980	17.741.531.153.559.400	17.741.531.153.559.400
4590	V$ZNF334_01	+	1486	51.089.942.967.940.300	13.043.265.358.444.900	38.046.677.609.495.300	38.046.677.609.495.300
57	V$GATA2_01	+	1487	0.28459153136189874	-143.161.550.263.751	17.162.070.339.994.000	17.162.070.339.994.000
1377	V$IRX5_01	+	1487	-3.733.889.456.027.410	2.684.209.368.140.880	-64.180.988.241.683.000	64.180.988.241.683.000
2268	V$ELF4_04	+	1487	-16.559.420.823.673.500	0.34405791763264415	-19.999.999.999.999.900	19.999.999.999.999.900
3749	V$NR2C2_01	+	1487	-23.569.940.464.196.700	0.35836940624302294	-27.153.634.526.626.900	27.153.634.526.626.900
4116	V$ZNF688_01	+	1487	0.7609455784745105	-0.7405319367435647	15.014.775.152.180.700	15.014.775.152.180.700
447	V$ZIC2_01	+	1488	0.12995400885374309	-26.774.009.132.038.600	28.073.549.220.576.000	28.073.549.220.576.000
542	V$STAT5A_04	+	1488	27.347.868.526.884.200	11.498.243.519.672.600	15.849.625.007.211.500	15.849.625.007.211.500
604	V$ZF5_01	+	1488	36.899.860.736.281.000	0.9895463554870131	27.004.397.181.410.900	27.004.397.181.410.900
806	V$T3R_Q6	+	1488	0.29321667872113244	2.556.251.084.554.920	-22.630.344.058.337.900	22.630.344.058.337.900
895	V$HNF4_Q6_01	+	1488	-15.942.003.804.494.700	0.6008818120618984	-21.950.821.925.113.600	21.950.821.925.113.600
1180	V$LHX2_01	+	1488	-17.944.789.607.136.400	1.659.696.932.472.150	-34.541.758.931.858.000	34.541.758.931.858.000
1251	V$GBX1_01	+	1488	-175.935.872.979.622	16.948.171.633.895.800	-34.541.758.931.858.000	34.541.758.931.858.000
1522	V$BDP1_01	+	1488	39.230.389.750.278.200	118.723.731.399.525	27.358.016.610.325.700	27.358.016.610.325.700
2331	V$HOXA3_07	+	1488	0.5841154479778631	-34.158.845.520.221.300	39.999.999.999.999.900	39.999.999.999.999.900
2369	V$MAFK_04	+	1488	-16.827.593.827.565.400	0.9566509019869878	-26.394.102.847.435.300	26.394.102.847.435.300
2434	V$SFPI1_04	+	1488	16.889.431.300.871.700	-0.5037019478552258	21.926.450.779.423.900	21.926.450.779.423.900
2511	V$SOX8_04	+	1488	-1.230.351.095.590.360	2.692.481.043.887.170	-39.228.321.394.775.400	39.228.321.394.775.400
2537	V$TCF1_07	+	1488	23.688.657.487.079.200	0.5845944397633621	17.842.713.089.445.600	17.842.713.089.445.600
3758	V$NR2F1_04	+	1488	-0.5303797900150374	38.557.565.140.714.700	-43.861.363.040.865.000	43.861.363.040.865.000
3989	V$ETS1_06	+	1488	12.650.462.656.636.500	-4.421.454.261.519.560	56.865.005.271.832.100	56.865.005.271.832.100
4028	V$HDAC2_04	+	1488	0.787297051628391	-29.623.453.353.372.300	37.496.423.869.656.200	37.496.423.869.656.200
4077	V$E2F4_07	+	1488	11.487.529.275.868.900	-0.45545579702960204	16.042.087.246.164.900	16.042.087.246.164.900
4178	V$SALL1_02	+	1488	4.439.138.316.392.080	2.937.285.941.308.730	15.018.523.750.833.500	15.018.523.750.833.500
4315	V$ZNF66_01	+	1488	15.725.339.678.270.800	4.494.239.811.046.910	-29.217.058.432.198.200	29.217.058.432.198.200
4529	V$ZFP770_01	+	1488	5.128.790.079.734.610	34.619.948.576.244.000	16.667.952.221.102.000	16.667.952.221.102.000
4588	V$ZFP287_02	+	1488	-22.015.912.902.146.200	0.22315425872353067	-24.247.455.489.381.600	24.247.455.489.381.600
4651	V$SALL1_04	+	1488	4.341.409.670.410.040	28.402.155.273.814.800	15.011.941.430.285.500	15.011.941.430.285.500

Table 19. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for the cloned region (+191/+1056) of the first intron of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070.

Human CETP Intron 1 (+191/+1056) (position: 1261-2126)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
430	V$VDR_Q3	+	1489	0.32663590515386753	3.326.635.905.153.860	-29.999.999.999.999.900	29.999.999.999.999.900
1324	V$HNF1B_01	+	1489	-2.214.767.775.373.620	0.3701947253475266	-25.849.625.007.211.500	25.849.625.007.211.500
1740	V$HNF4A_Q6_01	+	1489	-0.6786928944360283	21.558.834.963.566	-28.345.763.907.926.200	28.345.763.907.926.200
2010	V$RXRRAR_01	+	1489	-34.771.117.530.925.900	0.9508969966481899	-44.280.087.497.407.800	44.280.087.497.407.800
2247	V$BCL6B_04	+	1489	13.556.723.162.998.600	-2.158.900.856.529.890	35.145.731.728.297.500	35.145.731.728.297.500
3761	V$RARA_07	+	1489	-57.869.880.628.286.200	0.9006339245541479	-66.876.219.873.827.700	66.876.219.873.827.700
4106	V$SP1_08	+	1489	3.924.362.563.055.490	20.202.966.512.104.200	19.040.659.118.450.700	19.040.659.118.450.700
632	V$P53DECAMER_Q2	+	1490	-2.451.768.949.208.560	0.7961585642350175	-32.479.275.134.435.800	32.479.275.134.435.800
1563	V$DEC2_Q2	+	1490	-21.635.959.140.932.100	0.6326422959670015	-27.962.382.100.602.100	27.962.382.100.602.100
2389	V$OSR1_04	+	1490	24.060.414.308.175.800	-10.533.901.878.197.100	34.594.316.186.372.900	34.594.316.186.372.900
2572	V$ZFP187_04	+	1490	2.464.020.461.858.940	-0.36040797355760035	28.244.284.354.165.400	28.244.284.354.165.400
2910	V$TTF1_Q5_01	+	1490	2.847.848.880.231.020	63.332.757.074.012.600	-34.854.268.271.702.400	34.854.268.271.702.400
3484	V$HMX2_02	+	1490	-0.9698875818054565	1.656.465.988.568.980	-26.263.535.703.744.400	26.263.535.703.744.400
3489	V$HMX3_03	+	1490	-0.8318934046098493	17.879.406.098.151.800	-26.198.340.144.250.300	26.198.340.144.250.300
4048	V$ESR1_05	+	1490	2.006.052.603.486.420	42.925.753.622.425.900	-22.865.227.587.561.600	22.865.227.587.561.600
4321	V$ZNF625_01	+	1490	6.034.544.621.467.650	2.493.489.382.133.350	35.410.552.393.342.900	35.410.552.393.342.900
4692	V$DBP_Q6_02	+	1490	-22.691.724.804.249.200	0.43126723771616904	-27.004.397.181.410.900	27.004.397.181.410.900
137	V$TST1_01	+	1491	0.7302412507761247	-12.632.649.871.383.700	19.935.062.379.144.900	19.935.062.379.144.900
147	V$OCT1_03	+	1491	11.880.999.485.535.200	2.995.454.870.611.130	-18.073.549.220.576.000	18.073.549.220.576.000
2381	V$MYF6_04	+	1491	1.000.412.021.869.680	-245.901.959.676.761	34.594.316.186.372.900	34.594.316.186.372.900
3633	V$MIXL1_01	+	1491	-2.709.395.066.696.260	0.1758912355688477	-28.852.863.022.651.100	28.852.863.022.651.100
4198	V$ALF_01	+	1491	0.9041296655881349	4.792.872.914.486.390	-38.887.432.488.982.500	38.887.432.488.982.500
83	V$PAX2_01	+	1492	-0.3268582444177236	14.804.966.776.398.700	-18.073.549.220.576.000	18.073.549.220.576.000
198	V$MYB_Q6	+	1492	4.073.245.883.623.540	60.732.458.836.235.400	-20.000.000.000.000.000	20.000.000.000.000.000
638	V$DR1_Q3	+	1492	-17.702.361.168.419.000	0.622081305936851	-23.923.174.227.787.600	23.923.174.227.787.600
2545	V$TCFAP2C_04	+	1492	0.08577965253781694	-317.056.010.072.197	32.563.397.532.597.800	32.563.397.532.597.800
2862	V$CFOSCJUN_Q5	+	1492	14.936.177.547.109.900	-26.356.652.622.339.700	41.292.830.169.449.600	41.292.830.169.449.600
2880	V$FRA1_Q6_01	+	1492	24.158.266.414.226.200	-16.996.505.759.973.100	41.154.772.174.199.300	41.154.772.174.199.300
3352	V$BARHL1_03	+	1492	0.4960468124949148	21.209.314.797.910.900	-16.248.846.672.961.700	16.248.846.672.961.700
3373	V$DLX2_03	+	1492	-27.733.376.243.585.800	0.26398517280455214	-30.373.227.971.631.300	30.373.227.971.631.300
3429	V$ESX1_03	+	1492	-0.5912638442281011	20.199.485.556.508.800	-26.112.123.998.789.800	26.112.123.998.789.800
3433	V$EVX1_03	+	1492	-0.0327815604156846	21.197.577.512.257.900	-21.525.393.116.414.700	21.525.393.116.414.700
3450	V$GSC2_01	+	1492	3.270.003.434.375.150	10.708.006.404.105.900	21.992.027.939.645.500	21.992.027.939.645.500
3456	V$GSC_03	+	1492	14.747.557.163.707.800	-0.9933614868680365	24.681.172.032.388.200	24.681.172.032.388.200
3517	V$HOXB3_03	+	1492	22.847.338.330.392.500	38.145.134.184.911.000	-15.297.795.854.518.400	15.297.795.854.518.400
3658	V$NOTO_01	+	1492	-33.767.304.940.131.000	0.25387176823271185	-36.306.022.622.458.100	36.306.022.622.458.100
4148	V$ZNF663_01	+	1492	16.949.169.365.317.100	-5.631.687.996.084.290	73.266.049.326.160.100	73.266.049.326.160.100
4714	V$LHX2_Q4	+	1492	19.985.363.996.999.600	479.600.416.684.621	-27.974.677.671.462.400	27.974.677.671.462.400
4950	V$DLX3_Q6	+	1492	0.27702675279579425	2.277.026.752.795.790	-19.999.999.999.999.900	19.999.999.999.999.900
4998	V$MAZ_Q5	+	1492	34.782.181.094.653.900	0.630221202910443	28.479.969.065.549.400	28.479.969.065.549.400
399	V$HOXA3_01	+	1493	30.803.032.031.978.700	5.402.231.298.085.230	-23.219.280.948.873.600	23.219.280.948.873.600
1480	V$EAR2_Q2	+	1493	-0.5425939236784247	1.038.036.156.246.690	-15.806.300.799.251.200	15.806.300.799.251.200
3124	V$POU3F3_03	+	1493	-0.6679643251154237	13.415.584.490.030.800	-20.095.227.741.185.000	20.095.227.741.185.000
3349	V$BSX_03	+	1493	20.198.393.194.934.400	4.759.603.547.012.320	-27.397.642.275.188.800	27.397.642.275.188.800
3377	V$DLX3_04	+	1493	-16.237.992.385.046.600	0.1785200936556427	-18.023.193.321.603.000	18.023.193.321.603.000
3403	V$DLX1_04	+	1493	-18.883.097.798.895.900	0.3947990661072437	-22.831.088.459.968.300	22.831.088.459.968.300
3736	V$VAX2_03	+	1493	-2.614.682.617.299.510	0.058894270595205855	-26.735.768.878.947.200	26.735.768.878.947.200
3746	V$VSX2_01	+	1493	10.008.017.264.197.900	2.660.214.176.300.830	-16.594.124.498.810.300	16.594.124.498.810.300
70	V$MZF1_02	+	1494	0.9056493303228528	-0.901705591734751	18.073.549.220.576.000	18.073.549.220.576.000
956	V$MOVOB_01	+	1494	11.573.814.144.652.900	-14.275.810.862.558.600	25.849.625.007.211.500	25.849.625.007.211.500
2062	V$GLIS2_03	+	1494	0.1896090980504278	18.270.390.186.657.100	-16.374.299.206.152.900	16.374.299.206.152.900
2262	V$EHF_07	+	1494	-0.41522292328289717	1.648.907.414.136.810	-20.641.303.374.197.100	20.641.303.374.197.100
4756	V$POU4F3_Q5	+	1494	-2.083.248.824.464.590	0.22716068595449634	-23.104.095.104.190.900	23.104.095.104.190.900

Table 20. TRANSFAC analysis of human CETP sequence -1070/+1090 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for the cloned region (+191/+1056) of the first intron of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -1070/+1090, if we count as 1 position -1070. This table is a continuity of the previous table.

1.6.4. TRANSFAC analysis of human apoE sequence (-560/+160) from the Copenhagen City Heart Study

Human apoE -491A>T (position: 86)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
336	V$COUPTF_Q6	+	86	4.931.526.809.035.890	8.101.451.810.478.200	-31.699.250.014.423.100	31.699.250.014.423.100
381	V$ETS2_Q6	+	86	5.764.650.123.416.170	8.325.750.104.591.860	-25.610.999.811.756.900	25.610.999.811.756.900
504	V$MAFA_Q4	+	86	-13.147.525.230.448.700	18.423.683.558.259.400	-31.571.208.788.708.100	31.571.208.788.708.100
551	V$ATF1_Q6_01	+	86	-0.05801404300926738	34.014.175.756.280.300	-34.594.316.186.372.900	34.594.316.186.372.900
618	V$SPI1_Q5	+	86	19.360.999.402.814.400	4.199.134.346.115.240	-22.630.344.058.337.900	22.630.344.058.337.900
998	V$CMAF_Q5	+	86	10.520.786.408.558.200	30.520.786.408.558.200	-19.999.999.999.999.900	19.999.999.999.999.900
1065	V$SNAI2_01	+	86	14.993.434.649.453.900	-0.4450968689795225	19.444.403.339.249.100	19.444.403.339.249.100
1175	V$FOXO3_05	+	86	-19.264.968.292.380.600	15.865.727.530.013.000	-35.130.695.822.393.700	35.130.695.822.393.700
1515	V$ZNF826_01	+	86	-20.836.556.605.658.800	18.968.919.767.679.300	-39.805.476.373.338.200	39.805.476.373.338.200
1585	V$ZNF627_02	+	86	7.743.907.214.191.770	5.926.771.271.341.580	18.171.359.428.501.800	18.171.359.428.501.800
1780	V$P53_Q3	+	86	0.23812978700359178	-19.545.152.909.388.000	21.926.450.779.423.900	21.926.450.779.423.900

Table 21. TRANSFAC analysis of human apoE sequence -560/+160 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism -491A>T of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -560/+160, if we count as 1 position -560.

Human apoE -427T>C (position: 150)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
29	V$MZF1_02	+	150	5.211.242.487.705.280	2.889.314.392.817.920	23.219.280.948.873.600	23.219.280.948.873.600
276	V$SP1_Q2_01	+	150	8.736.349.187.069.010	11.185.656.588.432.600	-24.493.074.013.635.800	24.493.074.013.635.800
496	V$AML1_Q4	+	150	15.028.092.003.993.800	4.310.164.122.456.980	-28.073.549.220.576.000	28.073.549.220.576.000
527	V$PLAG1_01	+	150	18.343.273.740.130.200	4.523.187.128.546.860	-26.888.597.545.338.300	26.888.597.545.338.300
686	V$NKX2B_Q3	+	150	-0.8258701291998665	23.299.781.263.911.200	-31.558.482.555.909.900	31.558.482.555.909.900
815	V$GCM1_04	+	150	403.635.766.870.542	5.999.831.792.680.300	-19.634.741.239.748.800	19.634.741.239.748.800
955	V$KLF3_Q3	+	150	0.2002485340832485	290.068.825.222.434	-27.004.397.181.410.900	27.004.397.181.410.900
1046	V$REST_Q5	+	150	21.561.591.826.176.000	0.23862134280958203	19.175.378.398.080.200	19.175.378.398.080.200
1502	V$SP1_08	+	150	10.747.490.168.154.800	12.628.055.279.620.700	-18.805.651.114.658.700	18.805.651.114.658.700
1688	V$ZFP770_03	+	150	7.929.784.330.544.020	10.355.335.056.505.500	-24.255.507.259.615.100	24.255.507.259.615.100

Table 22. TRANSFAC analysis of human apoE sequence -560/+160 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism -427T>C of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -560/+160, if we count as 1 position -560.

Human apoE -219G>T (position: 358)							
Site.num	Matrix	Strand	Position	WT.score	Mut.score	WT-Mut	Diff.
116	V$GEN_INI_B	+	358	0.8903299688500956	-2.279.595.032.592.210	31.699.250.014.423.100	31.699.250.014.423.100
196	V$STAT1_03	+	358	21.879.478.151.388.100	0.380592893081211	18.073.549.220.576.000	18.073.549.220.576.000
384	V$ESE1_Q3	+	358	0.8556267232023226	2.633.234.301.865.870	-17.776.075.786.635.500	17.776.075.786.635.500
387	V$MAFB_01	+	358	33.808.385.871.646.100	-0.31960113097647724	37.004.397.181.410.900	37.004.397.181.410.900
505	V$CRX_Q4_01	+	358	8.317.656.155.972.420	4.459.675.160.844.850	38.579.809.951.275.700	38.579.809.951.275.700
641	V$ESE1_02	+	358	0.8556267232023226	2.633.234.301.865.870	-17.776.075.786.635.500	17.776.075.786.635.500
1400	V$FOS_01	+	358	14.614.332.701.656.900	29.808.074.292.592.700	-15.193.741.590.935.700	15.193.741.590.935.700

Table 23. TRANSFAC analysis of human apoE sequence -560/+160 using the latest version 2014.4. In the table, score for the most common allele (WT.score), score for the less common allele (Mut.score), WT score minus Mut score (WT-Mut), as well as the absolute value of WT score minus Mut score (Diff.) are indicated for polymorphism -219G>T of the CCHS. The position number in parenthesis, demonstrates the position the SNP is located in the sequence -560/+160, if we count as 1 position -560.

CHAPTER 2

Transcriptional regulation of mouse Carboxylesterase 3 beta (Ces3b) by Hepatocyte Nuclear Factor 4 (HNF-4).

2.1. INTRODUCTION

It has been well established over the years that in most members of the superfamily of ligand-dependent nuclear receptors, dimerization plays an important role for their ability to function. This family of soluble, nuclear receptors contains over 150 members, including the steroid, thyroid hormone, vitamin D and vitamin A receptors, as well as an even larger number of orphan receptors (Jiang *et al.,* 1997b).

Nuclear receptors (NRs) are ligand-dependent transcription factors (TFs) which play important roles in almost every aspect of human physiology (Laudet *et al.*, 2004). The NR superfamily is characterized by two highly conserved domains: the DNA binding domain (DBD), located in the N-terminal half of the protein that consists of approximately 60-90 amino acids and forms two zinc finger modules that are followed by a C-terminal extension; and the ligand binding domain (LBD), which is found in the C-terminal half of the protein and consists of approximately 200 amino acids. The latter domain performs a variety of functions including ligand binding, transactivation, as well as protein dimerization via heptad repeats of hydrophobic residues (Kurokawa *et al.*, 1993; Zechel *et al.*, 1994a; Zechel *et al.*, 1994b). NRs are closely linked to human disease; they are important drug targets and play a significant role in the regulation of expression of genes which are responsible for drug metabolism (Overington *et al.*, 2006; Chen, 2008; Xie, 2008). NRs regulate gene expression by binding to specific DNA response elements in the regulatory regions of target genes. Several studies have demonstrated the presence of DNA binding motifs of NRs. Such motifs consist of distinct half sites with the sequence AGAACA(in the case of the steroid receptors) or AGGTCA (in the vase of the non-speroid receptors) in various configurations such as direct repeats (DRs), inverted repeats (IRs), everted repeats (ERs) and non-repeats (nRs) with variable spacing between the repeats (Evans, 1988; Umesono *et al.*, 1989; Umesono *et al.*, 1991; Cotnoir *et al.*, 2001). DRs with an intermediate spacing of a single nucleotide are designated as DR1. There are three well-characterized subfamilies of receptors that bind DR1s as homodimers: RXRα (NR2B1), COUPTF2 (NR2F2) and HNF-4α (NR2A1) (Fang *et al.*, 2012).

2.1.1. Hepatocyte Nuclear Factor 4 alpha (HNF-4α)

Hepatocyte nuclear factor 4 alpha (HNF-4α, NR2A1) comprises a highly conserved member of the NR superfamily of ligand-dependent transcription factors (Sladek *et al.*, 1990) (Figure 42) and appears to be essential both in early as well as in adult development (Chen *et al.*, 1994; Hayhurst *et al.*, 2001). HNF-4α is mainly expressed in the adult liver, intestine, pancreas, kidney and colon and to a lesser degree in the visceral endoderm and the stomach.

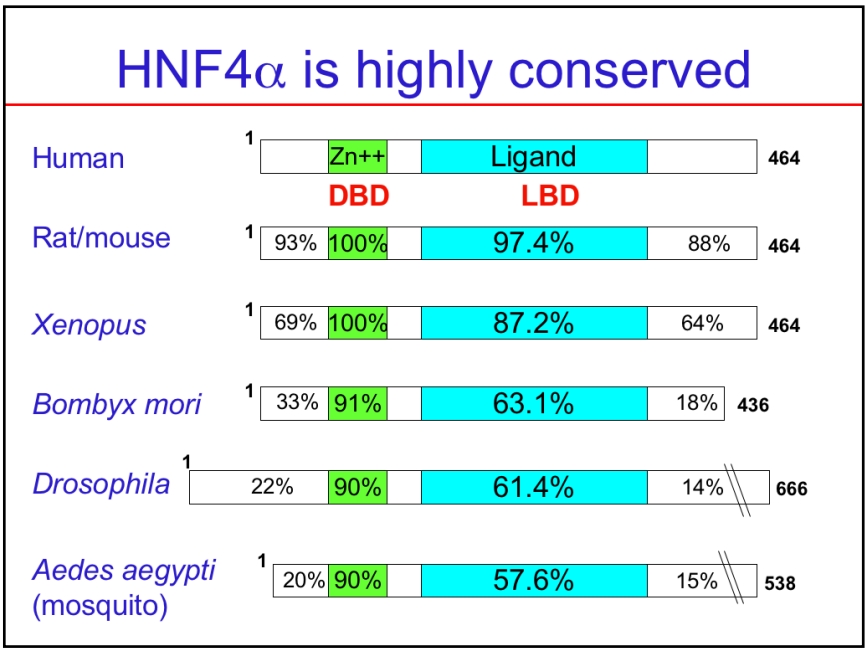

Figure 42. Conservation of HNF-4α during evolution. The figure depicts the amino acid identity (%) among different functional domains of the protein among various organisms. Numbers refer to amino acid residues. DBD, DNA Binding Domain. LBD, Ligand Binding Domain (Bolotin *et al.*, 2010 at http://www.cisreg.ca/tfe).

HNF-4α is a nuclear protein which binds DNA exclusively as a homodimer (Jiang *et al.*, 1995) and activates the expression of a wide variety of genes, especially those involved in intermediary metabolism, including fatty acid, glucose, xenobiotic, cholesterol and drug metabolism (Odom *et al.*, 2004; Waxman *et al.*, 2009), as well as others involved in blood coagulation, urea biosynthesis, hepatitis B infections, and liver differentiation (Jiang et al., 1997b) (Figure 43).

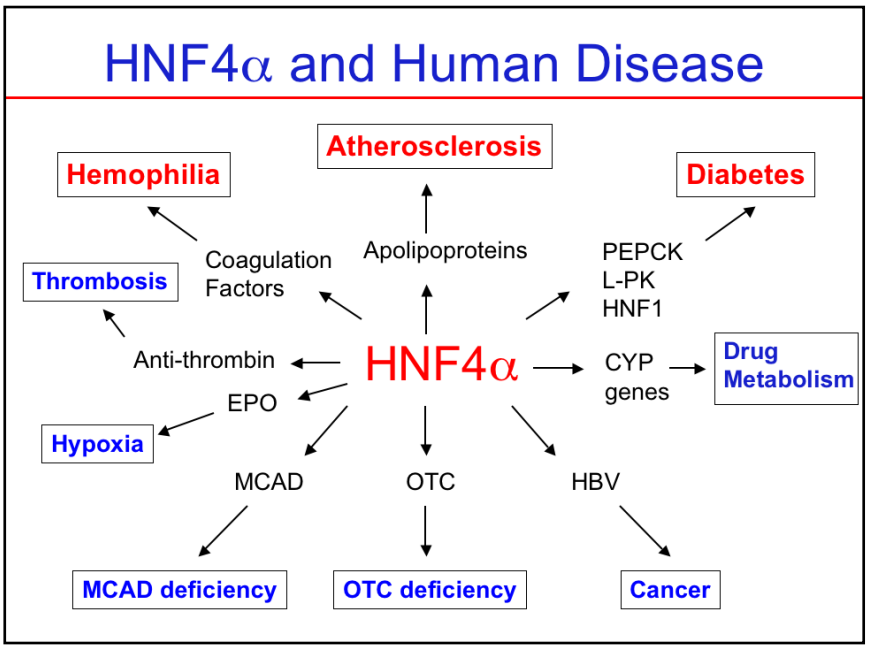

Figure 43. Implication of HNF-4α to human disease. HNF-4α is linked to several human diseases via mutations in its binding sites (diabetes, hemophilia) or in its gene (diabetes). An indirect linkage towards many human diseases is also known via regulation of its target genes (Bolotin *et al.*, 2010 at http://www.cisreg.ca/tfe).

HNF-4α is encoded by a single gene, which spans ~74 kb, and located on the long arm of chromosome 20 in humans (Kritis *et al.*, 1996). It is transcriptionally regulated by two distinct promoters, which are physically separated by more than 45 kb (Thomas *et al.*, 2001). The isoforms produced by the activity of the closer promoter are designated as P1, whereas those isoforms which are a consequence of the second-distal promoter are designated as P2 (Figure 44). The proximal P1 promoter is active in the adult liver, intestine and kidney, where most of the xenobiotic/drug metabolism takes place (Hwang-Verslues et al., 2010). The P1 isoforms (α1, α2, α3, α4, α5 and α6) are generated by alternative splicing of the pre-mRNA of HNF-4α, while the P2 isoforms (α7, α8 and α9) are generated through the same process (Drewes *et al.*, 1996; Kritis *et al.*, 1996; Furuta *et al.*, 1997; Nakhei *et al.*, 1998; Thomas *et al.*, 2001; Hansen *et al.*, 2002). The functional relevance of such a variety of HNF-4α isoforms still remains to be determined. Even though these classes of isoforms share almost 90% of homology in their protein structures, an important difference has been observed among their respective N-terminal domains (Figure 44). Specifically, the P2 isoforms are shorter compared to P1, and more importantly they lack a transactivation domain called activating function (AF)-1. Such differences in structure imply that P1 and P2 isoforms could harbor distinct roles, through different interaction with specific cofactors and by different regulation in specific contexts of physiological importance (Babeu *et al.*, 2014). HNF-4α expression is restricted to the epithelial compartment both in intestine and in liver. However adult hepatocytes express P1 isoforms only, whereas intestinal epithelial cells express both P1 and P2 isoforms (Torres-Padilla *et al.*, 2001; Tanaka *et al.*, 2006). It has been shown that HNF-4α expression varies according to the environmental context. Studies have demonstrated that the expression profile of P1 and P2 isoforms is modified in many cancers such as hepatocellular carcinoma in which the expression of the P1 isoforms is inhibited and o the P2 isoforms is activated (Tanaka *et al.*, 2006). Modulation of HNF-4α isoforms expression results from a complex regulatory circuit, involving transcriptional, post-transcriptional, as well as post-translational mechanisms (Babeu *et al.*, 2014).

HNF-4α has long been considered as an orphan receptor, however recent studies demonstrated that mammalian HNF-4α binds a single fatty acid (linoleic acid, LA, C18:2ω6) in a reversible fashion (Yuan *et al.*, 2009; Hwang-Verslues *et al.*, 2010). This finding implies that HNF-4α may be a potential drug target, as many of the other NRs are (Hwang-Verslues *et al.*, 2010).

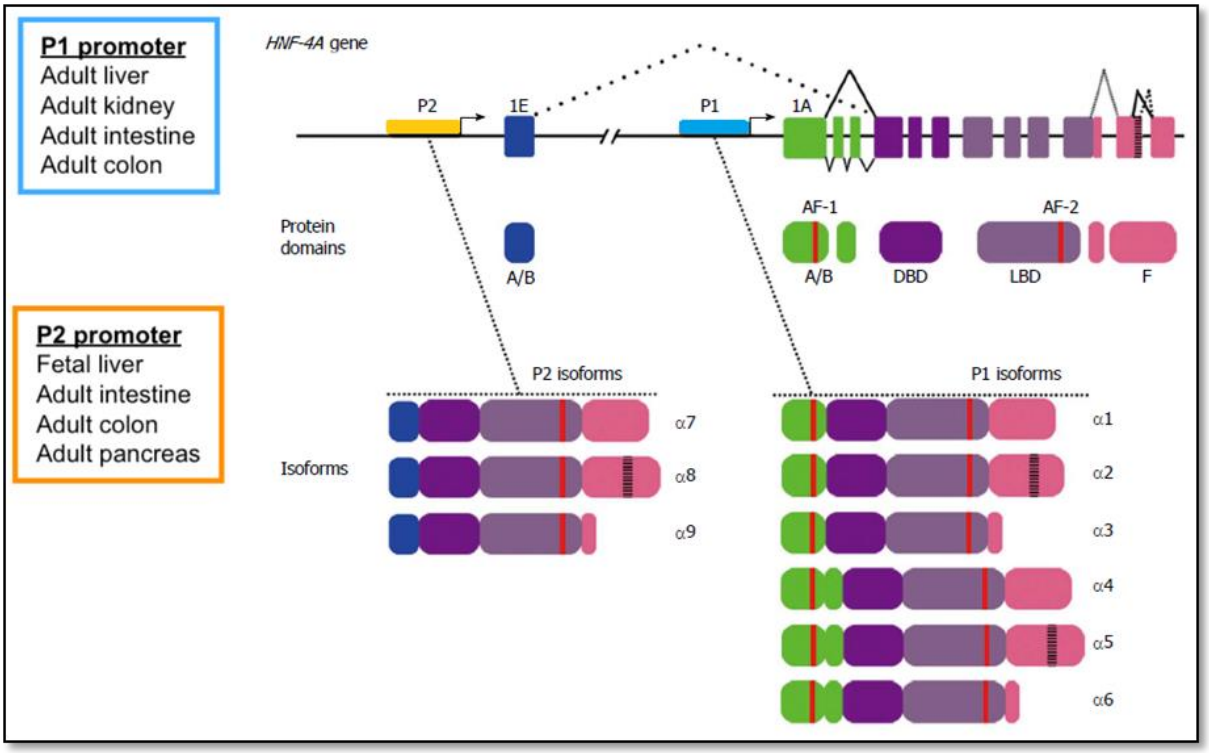

Figure 44. HNF-4α P1 and P2 isoforms classes originate from alternative promoters and splicing. HNF-4α contains two distinct promoters (designated as P1 and P2), which drive the expression of nine known isoforms (α1 to α9). Transcription through P1 promoter allows the inclusion of the exon 1A coding for the N-terminal domain of HNF-4α. P1 isoforms class bears a N-terminal region containing the cofactor interacting domain (AF-1). Transcription through P2 promoter allows the inclusion of the exon 1E but the exclusion of the exon 1A. P2 isoforms class bears a smaller N-terminal domain comparing to P1 isoforms and does not contain the AF-1 region. The regulating F domain is modified by alternative splicing of the last exons of HNF-4α, while alternative splicing of exon 14 modifies only A/B domain of the P1 isoforms. DBD, DNA Binding Domain. LBD, Ligand Binding Domain. AF-1, Activating Function-1. AF-2, Activating Function-2 (Adjustment from Bolotin *et al.,* 2010 at http://www.cisreg.ca/tfe and Babeu *et al.*, 2014).

2.1.1.1. Structure

As mentioned above, all members of the NR superfamily, such as HNF-4α, are composed of two conserved domains, the DBD and the LBD. More specifically, the DBD consists of two zinc fingers comprised of 4 cysteines each, followed by a hinge region. In 2012 Fang *et al.* identified a pair of residues in the HNF-4 DBD, which are responsible for binding of HNF-4 to a specific motif called HNF-4 specific binding motif (H4-SBM). These residues are Asp69 and Arg76, the former seems to have the greater effect on HNF-4-specific binding and is located in the P box. On the other hand, the latter (Arg76) is located in the helix contacting the DNA. The two residues interact with side chains of nucleotides at position 3 and position 4 in both half sites of the DR1 motif. It must be mentioned that HNF-4 is the only human NR with the combination of Asp69 and Arg76 in all HNF-4 DBDs (Fang *et al.*, 2012). The LBD consists of 12 alpha helices, creating a hydrophobic pocket. The N-terminal region to the DBD contains an activation function (AF-1) (Green *et al.*, 1998), which can activate transcription on its own in the isoforms driven by the P1 promoter, whereas P2-driven isoforms lack an AF-1 function (Torres-Padilla *et al.*, 2002). Moreover, an unusually long F domain of ~90 amino acids is found in the C-terminal region of the

protein (Figure 45), which contains a repressor function (Hadzopoulou-Cladaras *et al.*, 1997; Sladek *et al.*, 1999).

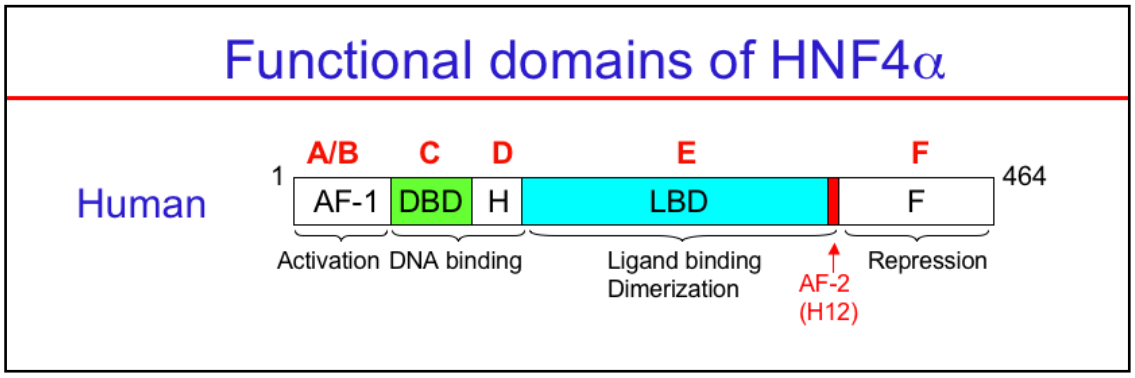

Figure 45. Domain structure of HNF-4α. Classical domains, A/B, C, D, E and F are demonstrated in red. The AF-1 region is known to interact with many co-activators (p300, CBP, GRIP1, SRC1), general transcription factors (TFIIB, TBP, TAFs, PC4, ADA2), mediator components and transcriptional activators (Smad3/4). The zinc finger (Zn++) plus hinge (H) region is sufficient for DNA binding, however the LBD provides the major dimerization motifs in helices 9 and 10. AF-2 is located in helix 12 and is obligatory for transactivation and interaction with various co-activators. The F domain represses transcription (Bolotin *et al.*, 2010 at http://www.cisreg.ca/tfe).

HNF-4α protein is covalently modified in many ways. Phosphorylation is the most prominent among the post translational modifications, with more than a dozen Ser/Thr sites being modified, although only few of those sites have been assigned to specific kinases, such as AMPK (Hong *et al.*, 2003), protein kinase A (Jiang *et al.*, 1997c), protein kinase C (Sun *et al.*, 2007) and p38 kinase (Guo *et al.*, 2006). Apart from Ser/Thr, phosphorylation of tyrosine residues in HNF-4 has also been observed (Ktistaki *et al.*, 1995), as well as acetylation by CBP (Soutoglou *et al.*, 2000) and methylation on arginines (Barrero *et al.*, 2006). All these modifications have different effects on the function of the HNF-4α protein, such as alterations in protein dimerization, DNA binding, transactivation and intracellular localization.

2.1.1.2. Expression

HNF-4α1 is expressed in kidney (proximal tubules), small intestine, liver (hepatocytes), pancreas (beta cells) and colon, however hepatocytes show the highest levels of expression (Sladek *et al.*, 1990). HNF-4α2 is the predominant isoform in adult kidney and liver. HNF-4α7/8 is not present in the adult kidney or liver but appears to be the predominant form in the pancreatic beta cells. In addition, it is expressed in the colon and small intestine, along with HNF-4α1/2. The mRNA levels of HNF-4α in the liver are regulated by many physiological as well as pathological conditions, such as hyperinsulinemia which decreases HNF-4α levels and fasting which increases HNF-4α levels, via the SREBP transcription factors (Xie *et al.*, 2009). HNF-4α has also been observed in the mammary epithelium (Ishikawa *et al.*, 2008) and ovarian cancer (Sugai *et al.*, 2008), the stomach (Tanaka *et al.*, 2006; Harries *et al.*, 2008) as well as in non-beta cell compartments of the pancreas

(Nammo *et al.*, 2008), however its role in those tissues has not been clarified. Moreover, HNF-4α is expressed in several human and rodent cell lines of colonic, hepatic and beta cell origin, although the levels of its expression varies among these cell lines. It has been demonstrated that transformed cell lines of kidney origin, such as HEK293T, Cos-1 and Cos-7 lack HNF-4α expression (Tanaka *et al.*, 2006). Cell lines known for their HNF-4α expression are HepG2 which express mainly HNF-4α1/2 and to a lesser degree HNF-4α7/8, Caco2 which express mainly HNF-4α7/8 and to a lesser extent HNF-4α1/2, Hep3b, INS-1 and others (Tanaka *et al.*, 2006; Huang *et al.*, 2008).

2.1.1.3. Hepatocyte Nuclear Factor 4 alpha (HNF-4α) Binding Sites

Several studies over the years have demonstrated the binding of HNF-4α to direct repeats with a spacing of one nucleotide (DR1). As mentioned above, HNF-4α binds to half sites having the sequence AGGTCA and typically one of the half sites appears to be more GC rich. However there are plenty of reports in the literature about binding of HNF-4α to different variations of the DR1. In most cases, the spacer nucleotide seems to be an adenine (A) (Bolotin et al., 2010 at http://www.cisreg.ca/tfe). Jiang *et al.* in 1997 demonstrated that HNF-4α can also bind to direct repeats with a spacing of two nucleotides (DR2), but with a lower affinity compared to DR1s. In addition, the same group revealed that HNF-4α demonstrated low levels binding activity on single half-sites (Jiang *et al.*, 1997b). In 2012 Fang *et al.* demonstrated for the first time an HNF-4-specific binding motif (H4-SBM), xxxxCAAAGTCCA, as well as a previously unrecognized polarity in the classical DR1 motif (AGGTCAxAGGTCA) for HNF-4α (Figure 46), RXRα and COUPTF2 homodimers. Through ChIP-seq analysis it was shown that the H4-SBM binds HNF-4α but not other NRs in *vivo*, however farnesoid X receptor α (FXRα), pregnane X receptor (PXR) and Rev-Erbα seem to bind to sites that are adjacent to H4-SBMs.

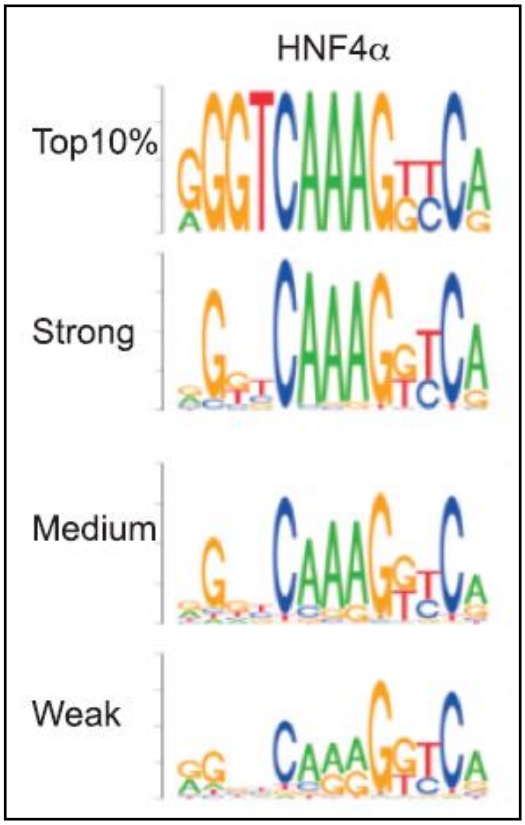

Figure 46. DNA motifs for the top 10% of binders, as well as top 33% (strong), middle 33% (medium) and bottom 33% (weak) of binders according to Protein Binding Microarrays (PBMs) (Fang *et al.*, 2012)

2.1.1.4. Hepatocyte Nuclear Factor 4 (HNF-4) Liver Knock Out Mice

Gene knock-out studies of HNF-4, either in the fetus or adult, have shown that HNF-4 regulates a great number of genes involved in most aspects of mature hepatocyte functions, such as xenobiotic detoxification, energy metabolism, bile acid

synthesis and serum protein production (Watt *et al.*, 2003). In 1994 Chen *et al.* disrupted the HNF-4 gene in mice, in order to address its function during development. HNF-4 null embryos failed to complete gastrulation, since HNF-4 was required for expression of several factors necessary for gastrulation support within the embryo (Chen *et al.*, 1994). Even though HNF-4 null embryos died prior to the onset of liver development, the consequences observed on gene expression upon loss of the HNF-4 function, provided an insight into potential targets for HNF-4-mediated transcription in hepatocytes (Watt *et al.*, 2003).

In 2001 Hayhurst *et al.* used the Cre-*lox*P system to produce adult hepatocytes lacking HNF-4, so as to address the function of HNF-4 in adult livers. Mice bearing HNF-4 null hepatocytes (H4LivKO) exhibited disrupted lipid homeostasis manifested as a marked increase in intracellular hepatocyte lipid levels and a reduction in triglyceride levels and serum cholesterol. The presence of high serum bile acid levels in these mice implied defects in the transport of bile acids (Hayhurst *et al.*, 2001). The aforementioned phenotypes were linked to a disruption in the expression of a number of genes known to encode products with specific roles in these processes, such as ApoB and microsomal triglyceride transfer protein (MTTP), which both encode for proteins involved in the export of very low-density lipoproteins from the liver. Expression of these proteins was significantly reduced in H4LivKO hepatocytes, however an increase in scavenger receptor B1 (SR-B1) expression, which is involved in cholesterol uptake from high-density lipoprotein (HDL), was observed. Increased expression of SR-B1 was assumed to cause a net accumulation of lipids in H4LivKO hepatocytes. Disrupted expression was also observed in other genes such as apolipoproteins C2 (Apoc2), C3 (Apoc3), HNF-1α, liver fatty acid binding protein 1 (Fabp1) and others (Li *et al.*, 2000).

2.1.1.5. Hepatocyte Nuclear Factor 4 (HNF-4) in Human Diseases

Several mutations in the HNF-4α gene have been identified over the years leading to a variety of phenotypes according to the type of the mutation. No homozygous mutations have been observed in human HNF-4α, and this is consistent with the embryonic lethality observed in the mouse (Chen *et al.*, 1994). Infants heterozygous for HNF-4α exhibited macrosomia (high weight), hypoglycemia at birth and hyperinsulinemia (Pearson *et al.*, 2007). Many single nucleotide polymorphisms (SNPs) have been mapped in the coding region of HNF-4α, as well as in the P2 but not in the P1 promoter, which are associated with diabetes (Ellard *et al.*, 2006). Moreover, it is well established that mutations in the HNF-4α gene lead to Maturity Onset Diabete of the Young 1 (MODY1). MODY1 is an autosomal dominant disease, which is inherited among generations, and is characterized by and early onset of non-insulin dependent diabetes mellitus (NIDDM). From a clinical point of view, this diabetes is manifested by defective insulin secretion of the pancreatic β-cells and is characterized by the occurrence of diabetes in at least two generations, bearing at

least one affected family member under the age of 25 years old. Studies suggest that people carrying one mutant allele of the HNF-4α are born with completely normal biochemical and physiological functions of the β-cells of the pancreas and that diabetes will occur sometime during adolescence (Ruffel 2001). The in-frame insertion V328ins is found in helix 10 of the LBD and is believed to alter this structural element. Moreover, frameshift mutations, such as K99fsdel and F75fsdel, result in truncated proteins, which lack part of the zinc finger domain that is essential for the binding of DNA. Nonsense mutations, such as Q268X and R154X, retain the DNA binding part, however they lack a significant portion of the potential ligand binding domain. All in all, such type of mutations lead to truncated proteins that have lost essential domains of HNF-4α, thus forming a defective protein (Ryffel, 2001).

2.1.2. Mammalian Carboxylesterases

In 1953, Aldridge classified esterase enzymes in rat, rabbit and horse serum based upon the nature of their interaction with organophosphates (Aldridge, 1953). Esterases that were unaffected by organophosphates and degraded the compounds were classified as A-esterases, whereas esterases that were inhibited by organophosphates were classified as B-esterases. On 1957 Bergmann et al. revealed the presence of a third group of esterases (designated as C-esterases), which did not interact with organophosphates at all (Bergmann *et* al., 1957). Using this classification scheme, the superfamily of carboxylesterase (CES) enzymes belong to the B-esterase group (Staudinger *et al*., 2010).

Mammalian carboxylesterases (CES, EC 3.1.1.1) comprise a family of genes, whose products are localized in the endoplasmic reticulum (ER) of many tissues. These enzymes can efficiently catalyze the hydrolysis of a variety of amide- and ester-containing chemicals, as well as drugs (including prodrugs), to the respective free acids. They are involved in the metabolic activation or detoxification of various drugs, carcinogens, and environmental toxicants. Carboxylesterases also catalyze the hydrolysis of endogenous compounds, such as long- and short-chain acyl-glycerols, long-chain acyl-coenzyme A (CoA), and long-chain acyl-carnitine esters (Hosokawa *et al.*, 2007). Apart from catalyzing hydrolytic and transesterification reactions with anticancer prodrugs, xenobiotics and narcotics (Satoh and Hosokawa 1998, 2006; Satoh *et al.*, 2002; Ohtsuka *et al.*, 2003; Redinbo and Potter 2005), the conversion of lung alveolar surfactant (Ruppert *et al.*, 2006), and several metabolic reactions (Tsujita and Okuda 1993; Becker *et al.*, 1994; Ghosh 2000; Diczfalusy *et al.*, 2001; Hosokawa *et al.*, 2007), they can also assist with the assembly of low-density lipoprotein particles in the liver (Wang *et al.*, 2007).

Five families of mammalian carboxylesterases have been described, including CES1, which is the major liver enzyme (Munger *et al.*, 1991; Shibita *et al.*, 1993; Ghosh 2000; Holmes *et al.*, 2009a), CES2, which comprises the major intestinal

enzyme (Langmann *et al.*, 1997; Schewer *et al.*, 1997; Holmes *et al.*, 2009a), CES3, which is expressed in several tissues such as brain, colon and liver (Sanghani *et al.*, 2004; Holmes *et al.*, 2010a), CES5 (also known as CES7 or cauxin), which is a major urinary protein of the domestic cat but is also present in human tissues (Miyazaki *et al.*, 2003, 2006; Holmes *et al.*, 2008a; Zhang *et al.*, 2009) and finally CES6 (currently designated CES4), which is a predicted CES-like enzyme in brain (Holmes *et al.*, 2009a; Williams *et al.*, 2010).

Structures for both human and animal CES genes have been reported, including human CES1 and CES2 genes (Becker *et al.*, 1994; Langmann *et al.*, 1997; Ghosh 2000; Marsh *et al.*, 2004) and rodent CES1-"like" and CES2-"like" genes (Ghosh *et al.*, 1995; Dolinsky *et al.*, 2001; Hosokawa *et al.*, 2007). Predicted gene structures have also been described for the human CES3, CES5 and CES6 genes, which are localized along with CES1 and CES2 in two contiguous CES gene clusters on human chromosome 16 (Holmes *et al.*, 2008a, 2009a,b, 2010a). Moreover, a CES1-like pseudogene (currently designated CES4) is located with the CES1-CES5 gene cluster (Yan *et al.*, 1999). Mammalian CES genes usually contain 12-14 exons of DNA encoding CES enzyme sequences which may be shuffled during mRNA synthesis, and in that way generating several CES transcripts and enzymes encoded by each of the CES genes (Thierry-Mieg 2006). CES promoters share several common binding sites for transcription factors among the same CES families, suggesting that orthologous CES genes have conserved transcriptional regulatory patterns during evolution. Potential binding sites of CES promoters for transcription factors include Sp1, Sp3, C/EBP, USF1, NF-1, NFkB, PPAR, HNF-1, HNF-3, HNF-4, GR and SREBP binding sites (Hosokawa *et al*., 2007).

2.1.2.1. Human Carboxylesterases (CES) Genes

As mentioned before, there are two contiguous gene clusters, where human CES genes are located in. Human CES1P1 (a CES1-like pseudogene), CES1 as well as CES5A are located in cluster 1 on chromosome 16, while CES2, CES3 and CES4A are in a separate cluster, cluster 2, on the same chromosome. Cluster 1 genes (CES1 and CES5A) are transcribed on the negative strand, whereas genes in cluster 2 (CES2, CES3 and CES4A) are transcribed on the positive strand. In addition, CES1 and CES4A genes contain 14 exons, CES3 and CES5A 13 exons, while CES2 has 12 exons. CES gene and transcript sequences vary in size ranging from 11 kb for CES2 to 79 kb for CES5A and exhibit distinct structures in each case. What is more, several isoforms have been generated *in vivo* for each of the human CES genes, which bear different structures as a result of transcriptional events, including alternative splicing or retention of introns, truncation of the 5' ends, differential presence or absence of exons, or overlapping exons with different boundaries (Holmes *et al.*, 2010b). Studies on human CES1 have established at least two major isoform transcripts, designated as CES1A1 and CES1A2. These isoforms encode sequences that differ by

only four amino acid residues within the N-terminal region (exon 1) (Tanimoto *et al*., 2007).

Examination of the amino acid sequences for the human CES subunits revealed that CES1, CES2, CES3, CES4A and CES5A contain 567, 559, 571, 561 and 575 residues, respectively. Different studies on human CES1 have identified key residues, contributing to the catalytic, oligomeric, subcellular localization as well as to the regulatory functions of this enzyme. More specifically, the catalytic triad for the active site has been identified (Ser221; Glu354; His468) (Cygler *et al.*, 1993), microsomal targeting sequences including the C-terminal endoplasmic reticulum (ER) retention sequence (His-Ile-Glu-Leu) (Robbi and Beaufay 1983) and the hydrophobic N-terminus signal peptide (Heijne 1983; Zhen *et al.*, 1995; Potter *et al.*, 1998), disulfide bond-forming residues (Cys87/Cys116 and Cys274/Cys285) (Lockridge *et al.*, 1987), and ligand-binding sites including the "gate" (Phe550) which may facilitate product release following catalysis) , the "side door" (Val424-Met425-Phe426) and the "Z-site" (Gly356) residues, where substrates, fatty acids and cholesterol analogs, respectively, are bound (Bencharit *et al.*, 2003, 2006; Fleming *et al.*, 2005). Similar residues were observed for each of the human CES subunit families both for the disulfide bond-forming residues and for the active site triad. However, changes were noticed for some key residues for CES1 subunits, including the "gate" and the "side door" of the active site, with family-specific residues or sequences in each case. The "Z-site" has been maintained for human CES2 and CES5A sequences, however, it was substituted for CES3 (Ser) and for CES4A (Asn). Major changes were observed at the hydrophobic N-terminal sequence for human CES sequences, although the signal peptide property has been retained. Regarding the C-terminal tetrapeptide sequences of the human CES, there has also been observed some changes, however CES2 (HTEL) and CES3 (QEDL) resemble in sequence with human CES1 (HIEL), which has been found to be important in the localization of human CES1 within endoplasmic reticulum membranes (Robbi and Beaufay, 1983). Bencharit *et al.* in 2003 and 2006 have compared predicted secondary structures for human CES2 (Holmes *et al.*, 2009b), CES3 (Holmes *et al.*, 2010a) CES4A (Holmes *et al.*, 2009a) and CES5A (Holmes *et al.*, 2008a), and revealed a similar a-helix β-sheet structure with human CES1.

2.1.2.1.1. *Human CES3 Gene*

Human CES3 gene was cloned and characterized for the first time by Sanghani *et al*. as a novel enzyme involved in the metabolism of xenobiotics (Sanghani *et al*. 2004), which are known to be substrates for CES1. Through Northern Blot experiments, it was proved that CES3 is expressed in liver, colon and small intestine (Figure 47), whereas in another set of experiments the carboxylesterase activity of CES3 enzyme to metabolize irinotecan (CPT-11) to its active form was established (Sanghani *et al*. 2004).

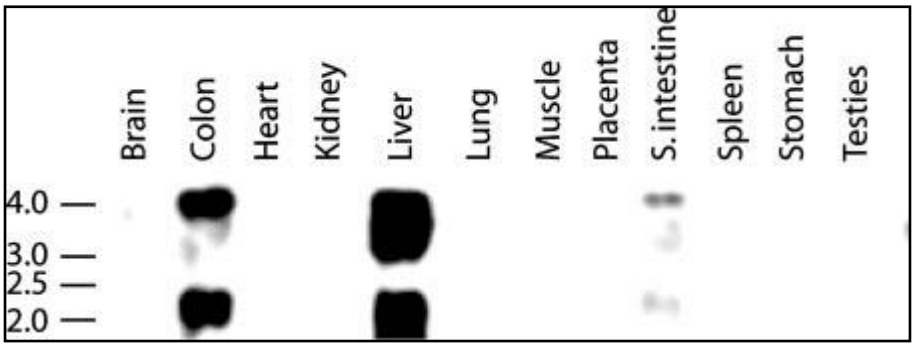

Figure 47. Human multi-tissue Northern blot probed for CES3. A CES3-specific probe was generated by PCR across 919-1234 nucleotides of CES3 gene. ^{32}P-dCTP was used for random primed labeling of the probe. The two major bands observed at 2 and 4 kb, may arise from multiple transcriptional start sites (Sanghani *et al.* 2004).

Moreover, Sanghani *et al*. at 2009 demonstrated that CES3 contains conserved structural and catalytic residues, a fact that leads to broad yet similar substrate specificities with other human CES genes. Holmes *et al*. compared human CES1 and CES3 genes and came to the conclusion that the predicted three-dimensional structure of CES3 is similar to that of CES1, which suggests similar enzymatic activities (Holmes *et al*., 2010a). Human CES3 mRNAs encode isoforms which are predicted to be either secreted or extracellular (CES3a-c), cytosolic (CES3d) or extracellular (CES3e-u), and are considered to play distinct roles in the metabolism of carboxyl-ester in the body. Comparison of the predicted cellular locations of mammalian CES3 revealed, that it is likely to be extracellular (including the cell wall) in primate (human, chimp and orangutan) and mouse tissues, whereas bovine and horse CES3 are predicted to be localized in the endoplasmic reticulum. The extracellular locations predicted for mouse and human CES3 could reflect distinct roles for this enzyme in drug metabolism among peripheral tissues of the body (Holmes *et al.*, 2010a).

As mentioned before, the CES family of enzymes was originally characterized as enzymes involved in xenobiotic metabolism, however it was shown afterwards, that they play significant role in metabolism of endogenous lipids. Zhao *et al*. at 2012 identified a new variant of macrophage CES3 that encodes for a bona fide lipid hydrolase. Such enzymatic activity was proved by performing transient transfections experiments in COS-7 cells with a CES3 expression vector (pCMV-CES3), which revealed an increase in cholesteryl esters (CE) as well as triolein hydrolytic activity. Moreover, the same group revealed a compensatory mechanism when silencing CES1, where an increase in CES3 expression occurred, resulting finally in a lack of phenotypic change (Zhao *et al*., 2012). It must be mentioned, that gene compensation is an established mechanism where loss of one gene results in an absence of change in phenotype, and such mechanism is specifically predicted in enzymes bearing overlapping substrate specificities and belonging to gene families (Gonzalez, 1998), as is the case for CES1 and CES3, both of which belong to the CES family of enzymes (Zhao *et al*., 2012).

2.1.2.2. Mouse Carboxylesterases (Ces) Genes

At least 20 mouse Ces genes are recognized on the Mouse Genome Database (http:/www.informatics.jax.org/) and located on mouse chromosome 8. Eight Ces1-like genes are located in tandem within a 360-kb segment on chromosome 8 (between coordinates 95,544,116 and 95,903,624) of mouse (Jones *et al.*, 2013), bearing an average gene size of 28 kb. The nomenclature of these genes is according to the order of their locations on the mouse genome (Ces1a, Ces1b, Ces1c, Ces1d, Ces1e, Ces1f, Ces1g and Ces1h). The Ces1-like gene cluster is also located close to the mouse Ces5a gene, a fact that is comparable to the CES1P1-CES1-CES5A cluster on human chromosome 16. Each of the aforementioned genes contain 13 or 14 exons predicted for transcription on the negative strand and with encoded CES subunits displaying distinct wet similar amino acid sequences (554-567 residues). The Ces1 family has been reported to be expressed predominantly in liver (Holmes *et al.*, 2010b). Mouse Ces1-like genes have been investigated over the years, such as Ces1c (previously called Es1) which not only encodes for a major mouse plasma esterase bearing 554 amino acid residues but also exhibits lung surfactant convertase activity (Genetta *et al.*, 1988; Krishnasamy *et al.*, 1998), Ces1d (previously designated as Ces3) encoding a mouse liver enzyme with 565 residues and exhibiting triacylglycerol hydrolase (TGH) activity (Dolinsky *et al.*, 2001), Ces1e (previously called Es22 or egasyn) which encodes a liver CES bearing 562 amino acid residues and exhibits β-glucuronidase-binding properties (Ovnic *et al.*, 1991), and Ces1g (previously called Ces1) which encodes a liver CES with 565 residues and exhibits lipid metabolizing activity (Holmes *et al.*, 2010b). However, in 2013 Jones *et al.* demonstrated that Ces1d mRNA was detectable in liver as mentioned before, but was far more abundant in brown (BAT) and white adipose tissues (WAT) (Jones *et al.*, 2013). Moreover, the same group revealed, that Ces1f mRNA levels were greatest in kidney comparing to other tissues examined, and in addition, that even though Ces1h mRNA levels were the highest in mouse lung comparing to other tissues, quantification of the PCR product was nearly at the limit of detection (Jones *et al.*, 2013).

Regarding the Ces2-like genes, eight have been observed in a second gene cluster of 286 kb on mouse chromosome 8, in which there is an average gene size of approximately 8 kb (Table 24). Ces2-like genes were named according to their sequence of position on the genome of mouse (designated as Ces2a, Ces2b, Ces2c, Ces2e, Ces2f, Ces2g and Ces2h and a pseudogene designated as Ces2d-ps). Previous studies have described three of these mouse Ces2-like genes, Ces2c (previously known as Ces2) which encodes an inducible liver acyl-carnitine hydrolase enzyme with 561 residues (Furihata *et al*., 2003), Ces2e (formerly known as Ces5) which encodes an intestinal and liver enzyme bearing 560 amino acid residues and finally Ces2a (also known as Ces6) which encodes a colon and liver enzyme with 558 residues (Holmes *et al*., 2010b). In 2013 Jones *et al.* demonstrated that almost all

the Ces2 family members were mostly expressed in segments of the small intestine, however it was shown that there is relatively high levels of mRNA expression in other organs as well, including liver (Ces2a and Ces2e), spleen (Ces2g) and kidney (Ces2c). Regarding Ces2f mRNA levels, those were extremely low in all tissues tested, including liver, colon, brain, WAT, BAT, kidney and spleen (Jones *et al*., 2013). The Ces2-like cluster is located alongside two Ces3-like mouse genes (Ces3a and Ces3b) and a Ces4a gene. This gene cluster is comparable to the CES2-CES3-CES4A human gene cluster on chromosome 16 (Holmes *et al*., 2010b).

There are two Ces3-like mouse genes. The Ces3a gene (previously known as mouse esterase 31 or Est31) is expressed strongly in livers of male mice and encodes a 554-residue CES3-like subunit (Aida *et al*., 1993). Regarding the Ces3b gene (previously designated as Es31L or EG13909) not much is known, however it is expressed in liver and encodes a 568-residue subunit (Holmes *et al*., 2010b) (Table 24). The liver mRNA expression of both Ces3a and Ces3b genes was also demonstrated through real-time PCR experiments, by Jones *et al*., in 2013.

The Ces4a gene (previously known as EST8 or Ces8) encodes and enzyme predicted for secretion in epidermal cells, which bears 563 amino acid residues. Ces4a shows 72% identity with human CES4A (Holmes *et al*., 2010b). It was demonstrated that Ces4a mRNA levels were highest in liver comparing to other tissues examined (Jones *et al*., 2013).

Regarding Ces5a with is the only member of the Ces5 family, Jones *et al*. demonstrated that its expression is evident in brain, WAT and the glucagonoma alpha-cell line (aTC1) through real-time PCR experiments (Jones *et al*., 2013).

Alignments of the human CES subunits revealed between 39 and 46% sequence identities, which imply that these are products of separate, however related gene families. Moreover, sequence alignments of human CES1 and CES2 with mouse CES1-like and CES2-like subunits exhibited 66-78% identities for human and mouse CES1-like subunits and 64-72% for human and mouse CES2-like subunits, suggesting that these are members of the same mammalian CES families. Comparisons between human CES3, CES4A and CES5A with the corresponding mouse CES homolog sequences, revealed 65, 72 and 69% identities, respectively. Such identities support the designation that these CES genes are members of the same family, in each case (Holmes *et al.*, 2010b).

Mouse CES gene (proposed)	Chr 8 coordinates	Gene size (bp)	Exons Strand[a]	Subunit MW	Amino acids	GenBank ID	MGI ID_YZ	Current MGI symbol_YZ	Current gene symbols	NCBI transcript	Vega ID	Ensembl ID	UNIPROT ID	Tissue expression (relative)[b]
Ces1a	95,544,116–95,572,091	27,979	14 −ve	61,744	563	BC089371	MGI:3648919	Gm4976	*EG244595*	NM_001013764	None	ENSMUSG 00000071047	Q5FWH4	Fetal liver [0.08]
Ces1b	95,580,789–95,603,815	23,027	13 −ve	62,197	567	*NM_001081372	MGI:3779470	Gm5158	*CesN*	NM_001081372	None	ENSMUSG 00000078964		Liver [×2.0]
Ces1c	95,622,914–95,655,182	32,268	13 −ve	61,172	554	BC028907	MGI:95420	Es1	*Es1, Ces-N*	NM_007954	ENSMUSG 00000024453	ENSMUSG 00000057400	P23953	Liver [×2.0]
Ces1d	95,690,157–95,721,618	31,462	14 −ve	61,788	565	BC019198	MGI:2148202	Ces3	*Ces3*	NM_053200	ENSMUSG 00000024539	ENSMUSG 00000056973	Q8VCT4	Tongue, liver [−2.2]
Ces1e	95,725,306–95,753,320	28,015	14 −ve	61,582	562	BC019208	MGI:95432	Es22	*Es22*	NM_133660	ENSMUSG 00000024532	ENSMUSG 00000061959	Q64176	Liver, kidney [0.4]
Ces1f	95,780,331–95,803,599	23,269	14 −ve	61,698	561	BC013479	MGI:234564	AU018778	*CesML1, TGH-2*	NM_144930	ENSMUSG 00000024519	ENSMUSG 00000031725	Q91WU0	Tongue, kidney [2.6]
Ces1g	95,826,807–95,861,053	34,247	14 −ve	62,680	565	BC021150	MGI:88378	Ces1	*Ces1*	NM_021456	ENSMUSG 00000024535	ENSMUSG 00000057074	Q3UW56	Tongue, kidney [2.6]
Ces1h	95,875,926–95,903,624	27,699	14 −ve	62,087	562	AK009689	MGI:75704	2310039D24Rik	*AK009689*	XM_134476	ENSMUSG 00000033579	ENSMUSG 00000074156		Tongue, kidney [2.6]
Ces2a	107,257,972–107,265,313	7,342	12 +ve	61,940	558	BC024491	MGI:2142491	Ces6	*Ces6*	NM_133960	OTTMUSG 00000027410	ENSMUSG 00000055730	Q8QZR3	Liver, colon [×1.0]
Ces2b	107,355,572–107,362,353	6,782	12 +ve	61,927	556	BC015286	MGI:2448547	BC015286	*BC015286*	NM_198172	OTTMUSG 00000027467	ENSMUSG 00000050097	Q5PDB7	Kidney, colon [0.1]
Ces2c	107,371,033–107,378,161	7,129	12 +ve	62,470	561	BC031170	MGI:2389505	Ces2	*Ces2*	NM_145603	OTTMUSG 00000027466	ENSMUSG 00000061825	Q91WG0	Kidney, colon [1.2]
Ces2d-ps	107,391,388–107,397,764	3,762	6 +ve			BC034182	MGI:3704319	Gm9756		XR_002069	None	ENSMUSG 00000031884		Pseudogene
Ces2e	107,450,221–107,457,611	7,391	12 +ve	62,735	560	BC055062	MGI:2443170	Ces5	*Ces5*	NM_172759	None	ENSMUSG 00000031886	Q8BK48	Liver, intestine [0.6]
Ces2f	107,471,256–107,479,862	7,335	12 +ve	62,707	561	BC117742	MGI:1919153	2310038E17Rik		NM_001079865	None	ENSMUSG 00000062826	Q08ED5	Tongue, thymus [0.2]
Ces2g	107,485,688–107,492,328	6,771	10 +ve	52,731	478	BC027185	MGI:1919611	2210023G06Rik		NM_197999	None	ENSMUSG 00000031877		Kidney, stomach [0.7]
Ces2h	107,524,753–107,544,307	19,554					MGI:3648740	Gm5744		XM_488149	None	None		Not available
Ces3a	107,572,572–107,582,000	21,512	13 +ve	61,510	554	AK138932	MGI:102773	Es31	*Es31*	NM_198672	None	ENSMUSG 00000069922	Q63880	Liver, aorta [1.1]
Ces3b	107,607,670–107,617,468	9,799	14 +ve	63,007	568	BC019047	Gm4738	Es31L	*Es31L*	NM_144511	None	ENSMUSG 00000062181		Liver [0.5]
Ces4a	107,655,852–107,673,417	17,566	14 +ve	62,123	563	BC026374	BC026374	Ces8	*Ces8*	NM_146213	OTTMUSG 00000027469	ENSMUSG 00000060560		Skin [0.1]
Ces5a	96,038,095–96,059,607	21,512	13 +ve	64,167	575	AB186393	MGI:1915185	Ces7	*Ces7*	NM_001003951	None	ENSMUSG 00000058019	Q8R0W5	Prostate [0.03]

Table 24. Mouse Ces genes and subunits (Holmes *et al.* 2010b)

RefSeq, GenBank, UNIPROT, MGI, Vega, and Ensembl IDs provide the sources for the gene and protein sequences. Gene sizes are given as base pairs of nucleotides

http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/

ps pseudogene (Ces2d-ps)

[a] +ve and −ve = transcription strand

[b] The relative gene expression level for mouse Ces genes in comparison with the expression of an average mouse gene is given in brackets

2.1.2.2.1. *Mouse Ces3b Gene*

As mentioned before, not much has been known over the years for mouse Ces3b gene, apart from the fact that it is mainly expressed in the liver (Jones *et al.*, 2013) (Figure 48) and encodes a 568-residue subunit (Holmes *et al*., 2010b). Ces3b (Ces31L or Ces3L or Es31L or EG13909) has been only mentioned in studies regarding the nomenclature of CES genes or their classification in specific CES families, however no one seems to have studied further about its enzymatic activity and its role in specific tissues or its implication in specific enzymatic pathways.

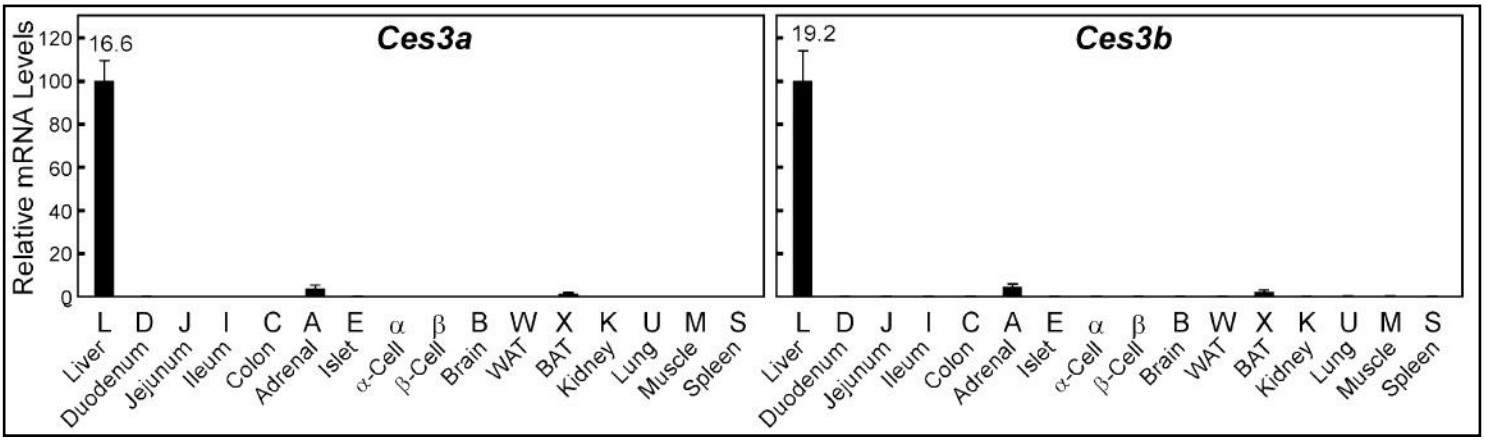

Figure 48. Tissue distribution of the carboxylesterase 3 family. The relative mRNA levels are depicted for mouse liver (L), duodenum (D), jejunum (J), ileum (I), colon ©, adrenal (A), islet (E), a-cell line (αTC1, α), and β-cell line (MIN6, β), brain (B), epididymal white adipose tissue (WAT, W), intrascapular brown adipose tissue (BAT, X), kidney (K), lung (U), muscle (M) and spleen (S). Tissues were obtained from adult (3-5 months of age) C57B1/6 mice that were fed a standard low-fat rodent diet. Values represent the means ±SEM of three independent samples for each tissue or cell line (Jones *et al.*, 2013).

Mouse Ces3a and Ces3b genes share 98% identity for the sequence for exons 2-4, with nearly identical amino acid sequences involving five amino acid substitutions for these genes. On the other hand mouse Ces3a and Ces3b introns 2 and 3 demonstrated distinct nucleotide sequences which were 85 and 94% identical, respectively. Moreover, BLAT analyses of the mouse genome showed distinct wet closely localized predicted positions for these genes, which were separated by ~26 kb of DNA on mouse chromosome 8 (Holmes *et al*., 2010a). Such data support the hypothesis made years ago by different groups that there are two closely localized and related CES3-like genes in the mouse genome, for which the subunits have been previously designated as esterase-31 (for Ces3a) (Aida *et al.*, 1993) and esterase-31L (also called Ces3L for Ces3b) (Holmes *et al*., 2010a).

In 2007 Lein *et al.* examined the distribution of Ces3a and Ces3b within different regions of the brain and demonstrated that the cerebellum fold exhibited the highest staining levels for both transcripts, however staining was detectable in other areas of the brain as well, such as the olfactory bulb, amygdalar nuclei, the hippocampus, the pons, the cortex and the medulla regions (Lein *et al.*, 2007).

Kroetz *et al*. in 1993 reported that the N-glycosylated carbohydrate group is contributing to mammalian CES stability and maintenance of catalytic efficiency (Kroetz *et al.*, 1993). It is proposed that such a property may be shared by the mammalian CES3 subunits, especially by those that contain multiple potential sites for N-glycosylation, like the case of mouse Ces3b (Ces3L) sequence, which contains four such sites (Holmes *et al*., 2010a).

2.1.3. Aim of this study

Preliminary (unpublished) data from our laboratory using microarrays in four 6-weeks old HNF-4α liver knock-out mice (AlbCre;Hnf-4α$^{F/F}$) compared to four 6-weeks old control mice (Hnf-4α$^{F/F}$), revealed a significant number of genes belonging to the family of mouse carboxylesterases, which are up- or down-regulated in the absence of HNF-4α from mouse liver. Among the carboxylesterases down-regulated, we chose mouse Ces3b, since it was the one with the highest fold change between AlbCre;Hnf-4α$^{F/F}$ and Hnf-4α$^{F/F}$. What was observed from microarrays experiments was that mouse Ces3b is significantly down-regulated in the absence of HNF-4α from the liver.

So, the aim of this study is to examine whether mouse Ces3b expression is HNF-4α dependent. We intend to identify binding sites for HNF-4α in Ces3b promoter sequence. This will be succeeded, by transient transfection assays in human embryonic kidney (HEK) cells, through the generation and utilization of luciferase reporter vectors containing consequtive 5’ or/and 3’ deletions of the mouse Ces3b promoter region of interest, along with an expression vector for the cDNA of HNF-4α. HEK cells are known not to express HNF-4α (Tanaka *et al.*, 2006), rendering them suitable for HNF-4α overexpression experiments.

2.2. MATERIALS AND METHODS

Materials

Oligonucleotide primers were either by Eurofins Genomics-VBC Biotech (Vienna, Austria) or by Minotech IMBB-FORTH (Crete, Greece). *PfuUltra* high fidelity DNA polymerase (2.5U/µL) was from Agilent Technologies. *Q5* high fidelity DNA polymerase (2U/µL), restriction enzymes *Sac*I and *Xho*I, NEB4 Buffer, BSA 100X and T4 DNA ligase were from New England Biolabs. Dulbecco's modified Eagle's medium (DMEM), penicillin/streptomycin and trypsin/EDTA for cell culture were purchased from Invitrogen/Life Technologies. Fetal bovine serum (FBS) was purchased from BioChrom Labs. The luciferase assay system and the cell lysis buffer were purchased from Promega Corp. ONPG (*o*nitrophenyl β-D-galactopyranoside) was purchased from Sigma-Aldrich. pGL3 basic reporter luciferase vector was from Promega. CMV-βgal and pMT2-HNF4α were described previously (Prokova et al. J Biol Chem. 2002 Mar 15;277(11):9342-50; (Mosialou et al., 2010; Prokova et al., 2002). Salmon sperm DNA was from Invitrogen.

Methods

2.2.1. Quantification of DNA concentration

The ds DNA concentration was measured with the Infinite M200 PRO reader device. The oligonucleotide primer's concentrations (ssDNA) was measured with the same device and the final working concentration of the oligonucleotides was calculated according to the molecular weight (Mr) provided along with each oligonucleotide.

2.2.2. Polymerase Chain Reaction (PCR)

The sequence upstream mouse carboxylesterase 3 beta (Ces3b) transcription start site of transcript variant 1 (NM_144511), which corresponds to the longer transcript and encodes the longer isoform according to National Center for Biotechnology Information (NCBI) , was obtained from www.ensembl.org using the assembly name: GRCm38.p3 C56BL/6J .

In order to isolate mouse Carboxylesterase 3 beta (Ces3b) promoter region of interest, appropriate primers were designed, shown in Table 25, and total genomic DNA extracted from *Mus musculus* tail was used as DNA template.

The PCR program used was the following:

1) 95 °C for 2 minutes
2) 95 °C for 30 seconds
3) 61 °C for 30 seconds
4) 72 °C for 4 minutes

(2–4) x 30 cycles

5) 72 °C for 10 minutes
6) 4 °C for ever
7) End

The PCR reaction was performed with *PfuUltra* high fidelity DNA polymerase (2.5U/µL) (Agilent Technologies) at a final volume of 50 µL.

Promoter Region	Insert Length (bps)	Primer Name		Restriction Enzyme Site of Insertion on pGL4.10 [luc2] vector
		Forward	Reverse	
-3259/+186	3445	Ces3b-3259F	Ces3b+186R	*Sac*I-*Xho*I
-2892/+186	3078	Ces3b-2892F	Ces3b+186R	*Sac*I-*Xho*I
-1956/+186	2142	Ces3b-1956F	Ces3b+186R	*Sac*I-*Xho*I

Table 25. Oligonucleotide primers pairs used to obtain the desired mouse Ces3b fragments from total mouse genomic DNA template. Restriction enzyme sites (*Sac*I or *Xho*I) were designed in each primer sequence, so as to succeed the desirable site of insertion of the DNA fragment in pGL3 basic plasmid vector.

The sequences of the oligonucleotide primers are shown in Table 26.

Primer Name	Sequence	Start	End
Ces3b-3259F	5'- CCCGAGCTCAACTTAACCAGAAACCAGGCTGC -3'	-3259	-3237
Ces3b-2892F	5'- CCCGAGCTCAAACACATAAACCGGACCGC -3'	-2892	-2873
Ces3b-1956F	5'- CCCGAGCTCCTCTGATTATTTGACCGTGTGCAAG -3'	-1956	-1932
Ces3b+186R	5'- CCCGCTCGAGCACTCTTACTGCCTCACTCCA -3'	+186	+166

Table 26. Oligonucleotide primers sequence. Start position and end position of the primer can be seen, as well as, the restriction site and the extra nucleotides added before, so as for the restriction enzyme to cleave efficiently. The exact DNA sequence is provided in the Appendix.

It must be mentioned at this point that the PCR reaction for the following pair of primers: Ces3b-3259F and Ces3b-186R, was unsuccessful in all the different conditions tried.

Further PCR reactions were performed in order to succeed the fragmentation of Ces3b promoter region of interest to different parts.

The PCR program used in this case was the following:

1) 98 °C for 30 seconds
2) 98 °C for 10 seconds
3) 59 °C for 30 seconds } x 30 cycles
4) 72 °C for 2 minutes
5) 72 °C for 5 minutes
6) 4 °C for ever
7) End

The PCR reaction was performed with *Q5* high fidelity DNA polymerase (2U/µL) (New England Biolabs) at a final volume of 50 µL.

The oligonucleotide primers used for the reactions were the following (Table 27):

Promoter Region	Insert Length (bps)	Primer Name		Restriction Enzyme Site of Insertion on pGL4.10 [luc2] vector
		Forward	Reverse	
-1529/+186	1715	Ces3b-1529F	Ces3b+186R	*Sac*I-*Xho*I
-1052/+186	1238	Ces3b-1052F	Ces3b+186R	*Sac*I-*Xho*I
-532/+186	718	Ces3b-532F	Ces3b+186R	*Sac*I-*Xho*I
-159/+186	345	Ces3b-159F	Ces3b+186R	*Sac*I-*Xho*I
-80/+186	266	Ces3b -80F	Ces3b+186R	*Sac*I-*Xho*I
-80/+4	84	Ces3b -80F	Ces3b +4R	*Sac*I-*Xho*I

Table 27. Oligonucleotide primers pairs used to obtain the desired mouse Ces3b fragments from template: (1956/+186) Ces3b-luc on pGL3 basic plasmid vector. Restriction enzyme sites (SacI or XhoI) were designed in each primer sequence, so as to succeed the desirable site of insertion of the DNA fragment in pGL3 basic plasmid vector.

The oligonucleotide primers sequence is shown in Table 28.

Primer Name	Sequence	Start	End
Ces3b-1529F	5'- CCCGAGCTCATCCACTATCTCCCCTGTCAG -3'	-1529	-1509
Ces3b-1052F	5'- CGCGAGCTCAGGGGTCTTGGAAGCTTGTTAG -3'	-1052	-1030
Ces3b-532F	5'- CCCGAGCTCCAGTCTCCATATCCCCAGGTGC -3'	-532	-511
Ces3b-159F	5'- CCCGAGCTCAGAAGACTCTGCTGGTCCTG -3'	-159	-140
Ces3b -80F	5'- CCCCGAGCTCGCAGACCAAGTTAATTTTCACAAG -3'	-80	-57
Ces3b+186R	5'- CCCGCTCGAGCACTCTTACTGCCTCACTCCA -3'	+186	+166
Ces3b +4R	5'- CCCGCTCGAGACGGACTGTGGACGAAGCTG -3'	+4	-16

Table 28. Oligonucleotide primers sequence. Start position and end position of the primer can be seen, as well as, the restriction site and the extra nucleotides added before, so as for the restriction enzyme to cleave efficiently. The exact

3-5 nucleotides at the 5' end of the oligonucleotides where added on the side of the recognition site to cleave efficiently. The additional bases were chosen so that palindromes and primer dimers are not formed.

After the PCR reaction was complete, the PCR products were run in agarose gel of the appropriate concentration (usually 1 %), and the size of the DNA fragments of interest was verified with the appropriate DNA ladder (λ BstEII ladder).

2.2.3. PCR clean-up Gel extraction

The NucleoSpin Gel and PCR Clean-up Kit, by Macherey-Nagel (Düren, Germany), was used, in order to gel extract the PCR product from agarose gel. Briefly the protocol recommended was the following:

Excise DNA fragment/solubilize gel slice

- Take a clean scalpel to excise the DNA fragment from the agarose gel. Remove all excess agarose.
- Determine the weight of the gel slice and transfer it to a clean tube.
- For each 100mgr of agarose gel < 2% add 200µL buffer NTI. For gels containing > 2% agarose, double the volume of buffer NTI.
- Incubate sample for 5-10 minutes at 50 ^{o} C. Vortex the sample briefly every 2-3 minutes until the gel slice is completely dissolved.

Bind DNA

- Place a NucleoSpin Gel and PCR Clean-up Column into a collection tube and load up to 700 µL sample.
- Centrifuge for 30 seconds at 11000 x g. Discard flow-through and place the column back into the collection tube. Load remaining sample if necessary and repeat the centrifugation step.

Wash silica membrane

- Add 700 µL buffer NT3 to the NucleoSpin Gel and PCR Clean-up Column. Centrifuge for 30 seconds at 11000 x g. Discard flow-through and place the column back into the collection tube.
- Repeat previous washing step.

Dry silica membrane

- Centrifuge for 1 minute at 11000 x g to remove buffer NT3 completely.

Elute DNA

- Place the NucleoSpin Gel and PCR Clean-up Column into a new 1,5 mL microcentrifuge tube. Add 30 µL buffer NE and incubate at room temperature (18-25 ^{o} C) for 1 minute. Centrifuge for 1 minute at 11000 x g.

For DNA fragments larger than 1000 bps, buffer NE should be pre-heated at 70 ^{o} C and incubated for 5 minutes.

2.2.4. Digestion Reaction

One unit is defined as the amount of enzyme required to digest 1 µgr of DNA in 1 hour at 37°C in a total reaction volume of 50 µL.

Briefly, 1 µgr of dsDNA from gel extraction and 5 µgr of pGL3 basic vector (Promega) were restricted with SacI-HF enzyme (20U/µL) and XhoI enzyme (20U/µL) and incubated for 3 hours at 37 ^{o} C.

2.2.5. Heat Inactivation

Incubate for 20 minutes at 65 ^{o} C.

After heat inactivation of the enzymes, both insert and plasmid vector were gel extracted with the PCR clean-up Gel extraction protocol mentioned before, and the DNA concentration was quantified with the method previously described.

2.2.6. Ligation

The ligation reaction was performed with the use of the T4 DNA ligase enzyme, by New England Biolabs. Different vector to insert molar ratios and different vector concentrations were tried, in order to succeed the desired construct at a final reaction volume of 10 µL or 15 µL.

Incubation took place at 16 ^{o} C or 4 ^{o} C overnight (12-16 hours).

2.2.7. Transformation of bacterial strains with the use of competent cells

The bacterial strain that was used in the experiments is the DH10 beta strain, which is an *Escherichia coli* strain.

- Thaw competent cells on ice.
- Add either a part or the whole ligation reaction (total volume 15µL) to 100µL of DH10 beta competent cells.
- Mix gently by pipetting up and down or flicking the tube 4-5 times to mix the cells and DNA.
- Incubate for 30 minutes on ice.
- Heat shock for 45 seconds at 42 ^{o} C.
- Incubate on ice for 1 minute.
- Add 900µL fresh LB medium at room temperature (18-25 ^{o} C) without antibiotic.
- Incubate for 1 hour at 37 ^{o} C.
- In the meantime, prewarm fresh selection LB plates at 37 ^{o} C.
- Centrifuge at 3000 rpm for 5 minutes at room temperature (18-25 ^{o} C).
- Discard ~900µL of the flow through and resuspend the pellet in the volume of flow through remaining.
- With the use of a bacterial glass rod appropriate for plating, spread ~100µL of the cells and the ligation mixture onto the plates bearing the appropriate antibiotic for selection.

Incubate overnight (12-16 hours) at 37 ^{o} C.

2.2.8. LB plates supplemented with Ampicillin

- Weight the following:
  - 5gr NaCl (Merck, Darmstadt, Germany)
  - 2,5gr Yeast Extract (Sigma-aldrich, St. Louis, USA)
  - 5gr Peptone
- Dissolve the above in deionized water at a final volume of 500mL.

- After the solution is fully dissolved add 7,5gr Agar (Sigma-aldrich, St. Louis, USA).
- Sterilize short afterwards in an autoclave machine.
- After the sterilization is complete, stir until a temperature of ~37 ^{o}C is achieved, then add 1mL Ampicillin and stir for 2-3 minutes.
- Apply the medium in plates (Sarstedt, Nümbrecht, Germany) and allow them to chill at room temperature (18-25 ^{o}C) for 1 hour.
  Store at 4 ^{o}C for future use.

2.2.9. MiniPreps

In order to isolate plasmid DNA from bacterial cultures at short scale (MiniPreps), the Plasmid DNA Purification Kit by QIAGEN was used.

Cultivate and harvest bacterial cells

- At first, prepare a 2mL bacterial culture, by inoculating the appropriate volume of LB medium (plus antibiotic) with a single colony picked from a freshly transformed plate. Shake the culture overnight (12-16 hours) at 37 ^{o}C.
- Transfer 1,5mL of the bacterial culture in a 1,5mL eppendorf.
- Centrifuge the culture at 13000 rpm for 1 minute at room temperature (18-25 ^{o}C). Carefully discard supernatant.

Cell lysis

- Carefully resuspend the pellet of bacterial cells in 300μL Buffer P1+RNase A (Resuspension Buffer).
- Add 300μL Buffer P2 (Lysis Buffer) to the suspension. Mix gently by inverting the tube 6-8 times. Incubate the mixture at room temperature (18-25 ^{o}C) for 2-3 minutes (max. 5 minutes). Do not vortex, as this will release contaminating chromosomal DNA from cellular debris into the suspension.
- Add 300μL pre-cooled Buffer P3 (Neutralization Buffer) (4 ^{o}C) to the suspension. Immediately mix the lysate gently by inverting the tube 6-8 times until a homogeneous suspension containing an off-white flocculate is formed. Incubate the suspension on ice for 10 minutes.
- Centrifuge at 13000 rpm for 10 minutes at room temperature (18-25 ^{o}C).

Precipitation

- Transfer the lysate from the previous step into a clean 2mL eppendorf.
- Add 800μL pre-cooled isopropanol to precipitate the plasmid DNA. Mix carefully and incubate on ice for 10 minutes.
- Centrifuge at 13000 rpm for 15 minutes at room temperature (18-25 ^{o}C). Carefully discard supernatant.

Wash and dry DNA pellet

- Add 500μL pre-cooled 70% ethanol to the pellet. Vortex briefly and centrifuge at 13000 rpm for 5 minutes at room temperature (18-25 ^{o}C).

- Carefully remove ethanol from the tube with a pipette tip.
- Add again 500µL pre-cooled 70% ethanol to the pellet. Vortex briefly and centrifuge at 13000 rpm for 5 minutes at room temperature (18-25 ^{o}C).
- Carefully remove ethanol from the tube with a pipette tip.
- Allow the pellet to dry at room temperature (18-25 ^{o}C) for 10 minutes.

Reconstitute DNA

- Dissolve pellet in 35µL of sterile deionized H_2O.
- Allow pellet to dissolve by incubating overnight at 4 ^{o}C.

In order to select the positive for the insert colonies, plasmid DNA from each MiniPrep was run in agarose gel of the appropriate concentration (1 %) and the DNA fragment of interest was verified with the use of the appropriate ladder (λ BstEII ladder), while whether the enzymes have performed the restriction reaction successfully or not, was verified by running unrestricted plasmid vector and comparing it with the MiniPreps.

The positive colony was selected to be grown in MidiPreps.

2.2.10. LB medium

- Weight the following:
  - 10gr NaCl (Merck, Darmstadt, Germany)
  - 5gr Yeast Extract (Sigma-aldrich, St. Louis, USA)
  - 10gr Peptone
- Dissolve the above in deionized water at a final volume of 1000mL.
- Sterilize short afterwards in an autoclave machine.

2.2.11. Glycerol Stock

- Prepare a 2mL bacterial culture, by inoculating the appropriate volume of LB medium (plus antibiotic) with a single colony picked from a freshly streaked plate. Shake the culture for overnight (12-16 hours) at 37 ^{o}C.
- In a 1,5mL eppendorf add 300µL sterile glycerol (AppliChem, Darmstadt, Germany) and 700µL of the bacterial culture.
- Vortex briefly so as a homogenous mix is formed.

Store at -80 ^{o}C for future use.

2.2.12. MidiPreps

In order to isolate plasmid DNA from bacterial cultures at medium scale (MidiPreps), the Plasmid DNA purification NucleoBond PC 100 by Macherey-Nagel (Düren-Germany), was used.

Cultivate and harvest bacterial cells

- At first, prepare a 2mL bacterial culture, by inoculating the appropriate volume of LB medium (plus antibiotic) with a single colony picked from a freshly streaked plate. Shake the culture for 3-4 hours at 37 ^{o}C.
- Then, set up an overnight culture, by transferring the 2mL bacterial culture at a final volume of 100mL LB medium (plus the appropriate antibiotic) and shake this culture overnight (12-16 hours) at 37 ^{o}C.
- Centrifuge the culture at 5000 rpm for 10 minutes at 4 ^{o}C. Carefully discard supernatant. Repeat centrifugation, so as to obtain a bacterial pellet from the total volume of the bacterial culture.

Cell lysis

- Carefully resuspend the pellet of bacterial cells in 8mL Buffer S1+RNase A (Resuspension Buffer).
- Add 8mL Buffer S2 (Lysis Buffer) to the suspension. Mix gently by inverting the tube 6-8 times. Incubate the mixture at room temperature (18-25 ^{o}C) for 2-3 minutes (max. 5 minutes). Do not vortex, as this will release contaminating chromosomal DNA from cellular debris into the suspension.
- Add 8mL pre-cooled Buffer S3 (Neutralization Buffer) (4 ^{o}C) to the suspension. Immediately mix the lysate gently by inverting the tube 6-8 times until a homogeneous suspension containing an off-white flocculate is formed. Incubate the suspension on ice for 5 minutes.

Equilibration of the column

- Equilibrate a NucleoBond AX 100 (Midi) Column with 2,5mL Buffer N2 (Equilibration Buffer). Allow the column to empty by gravity flow. Discard flow-through.

Clarification of the lysate

- Centifuge the lysate at 5000 rpm for 10 minutes at 4 ^{o}C.
- Place a NucleoBond Folded Filter in a funnel of appropriate size. Wet the filter with a few drops of Buffer N2 (Equilibration Buffer) and load the bacterial lysate obtained from the centrifugation onto the wet filter. Collect the flow-through in a separate vessel and proceed to the next step.

Binding

- Load half of the volume of the cleared lysate from the previous step onto the NucleoBond Column. Allow the column to empty by gravity flow.

Washing

- Wash the column with 10mL Buffer N3 (Wash Buffer). Discard flow through.
- Repeat the previous step and discard the flow through.

Elution

- Place the column in a clean tube and elute the plasmid DNA with 5mL Buffer N5 (Elution Buffer).

Precipitation

- Add 3,5mL room-temperature isopropanol to precipitate the eluted plasmid DNA. Mix carefully and centrifuge at 15000 x g for 30 minutes at 4 ^{o}C. Carefully discard supernatant.

At this step, store pellet for a while at 4 ^{o}C and move back to the equilibration of the column step. Equilibrate again the column with 2,5mL Buffer N2 (Equilibration Buffer). Allow the column to empty by gravity flow. Discard flow-through.

Then, proceed to the binding step, and load the second half of the volume of the cleared lysate onto the NucleoBond Column. Allow the column to empty by gravity flow.

Proceed to the Washing, the Elution and the Precipitation (add 3,5mL room-temperature isopropanol to precipitate the eluted plasmid DNA and mix carefully) steps as indicated before.

Place the precipitated plasmid DNA to the tube containing the first pellet (which is stored for a while at 4 ^{o}C and centrifuge at 15000 x g for 30 minutes at 4 ^{o}C. Carefully discard supernatant.

Wash and dry DNA pellet

- Add 2,5mL room-temperature 70% ethanol to the pellet. Vortex briefly and centrifuge at 15000 x g for 10 minutes at room temperature (18-25 ^{o}C).
- Carefully remove ethanol from the tube with a pipette tip. Allow the pellet to dry at room temperature (18-25 ^{o}C) for 10 minutes.

Reconstitute DNA

- Dissolve pellet in 125µL of buffer TE.
- Allow pellet to dissolve by incubating overnight at 4 ^{o}C.

After the desired plasmid construct of interest has been grown in medium scale, the activity of the promoter is tested with the Luciferase Assay method.

TE Buffer

Tris-Cl	10 mM PH 8
EDTA	1 mM

2.2.13. Plasmid Constructions

The plasmid constructs (Figure 49) obtained from molecular cloning, on pGL3 basic vector, were the following:

- ✓ (-1956/+186) Ces3b-luc
- ✓ (-1529/+186) Ces3b-luc
- ✓ (-1052/+186) Ces3b-luc
- ✓ (-532/+186) Ces3b-luc
- ✓ (-159/+186) Ces3b-luc
- ✓ (-80/+186) Ces3b-luc
- ✓ (-80/+4) Ces3b-luc

These constructions are also available in more detail, in the Appendix (Figures 54, 55, 56, 57, 58, 59 and 60).

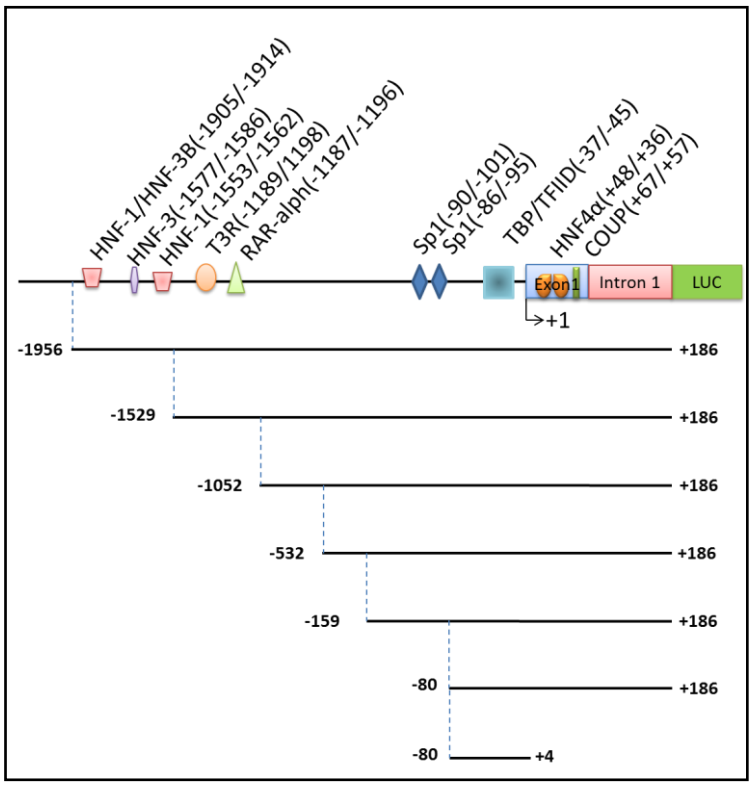

Figure 49. Plasmid constructions on mouse Ces3b promoter fragmentations on pGL3 basic vector. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

2.2.14. Cell lines

The cell line used was the human embryonic kidney cell line HEK293T. HEK233T cells are incubated at 37 ^{o} C, 5% CO_2, 100% humidity and at Dulbecco's Modified Eagle Medium 1X (DMEM) (Life Technologies), which contains 4,5 g/L D-glucose, supplemented with 10% Fetal Bovine Serum (FBS) (Biochrom) and 1% penicillin/streptomycin (P/S) (Life Technologies).

2.2.15. Cell culture-Passages

HEK293T cell line is characterized by the adherence of the cells to the surface of the culture flask. Cell passaging is performed when cell confluence is about 80%-90% in the culture flask and Trypsin-EDTA is used to disrupt cell adherence. Briefly the protocol is the following:

- Remove culture medium from the flask by aspiration.
- Wash the monolayer with 2 mL PBS (phosphate buffer saline) 1X.
- Remove PBS 1X by aspiration and add 1 mL cold Trypsin-EDTA.
- Incubate at 37 ^{o} C, 5% CO_2, 100% humidity for 2-3 minutes.
- Deactivate Trypsin-EDTA by adding 3 volumes of complete medium.

In the case of HepG2 cells, use a pipette tip in order to resuspend cells and to break up the clumps.

- Evaluate cell number in the cell suspension using a Neubauer Hemocytometer.
- Culture cells in complete medium under the conditions mentioned before.

2.2.16. Transient Co-Transfections

Transient co-transfections were performed using the calcium phosphate [$Ca_3(PO_4)_2$] co-precipitation method. The aim of transientco- transfection assay is to evaluate the luciferase activity (Luciferase assay). Such transfections are carried out in 6-well-plates for HEK293T cell line.

Day 0

Appropriate number of cells is placed in 6-well-plates, 250000 cells HEK293T/well, and incubated at 37 ^{o}C, 5 % CO_2 and 100 % humidity overnight (12-16 hours).

Day 1

Prewarm 2X Hepes Buffered Saline (HBS), $CaCl_2$ 2M, water for injection (WFI) and complete medium. Use freshly diluted plasmid constructs, with water for injection, at a final concentration of 0,5 μgr/μL.

- At the beginning, dilute the desired volume of plasmid constructs at a final concentration of 0,5 μgr/μL.
- Vortex and short spin the diluted plasmid constructs.
- Set up the Pre-Mixes in 1,5 mL tube, by adding WFI, $CaCl_2$ 2M, salmon sperm DNA, CMV-β gal and plasmid constructs of interest. Use the aforementioned order when adding each one of them to the Pre-Mix tube.
  It must be noted at this point, that salmon sperm DNA is being used so as to counterbalance the amount of DNA among the samples.
  In addition, in experiments which include the evaluation of luciferase activity, the expression vector of beta galactosidase (1 μgr/well) is being used, so as to counterbalance the effectiveness of the transfection among the samples.
- Vortex and short spin Pre-Mixes.
- In 2 mL tubes (named as Mix) add equal volume (250 μL) of HBS/tube.
- Place the Mix tube on a vortex device and add the equivalent Pre-Mix drop by drop while vortexing the Mix tube. At this point, set up a timer at 15 minutes and wait.

The conditions for plasmid constructs for the promoter of the Ces3b gene are shown in Table 29.

- While waiting, discard old medium from the 6-well-plates and add 1 mL/well of fresh medium.
- After the 15 minute incubation add 250 μL of the Mix to each well, so as to have a duplicate.
- Incubate at 37 ^{o}C, 5 % CO_2 and 100 % humidity overnight (12-16 hours).

Plate	Cell line (number of cells/well)	DNA (µgr)/duplicate				$CaCl_2$ 2M (µL)/ duplicate	H_2O (µL)/ duplicate	2X HBS (µL)/ duplicate
	HEK293T	Construct of interest (µgr)	CMV-beta gal (µgr)	PMT2-HNF4 (µgr)	ssDNA (µgr)			
6-well plate	250000	2	2	-	8	31	195	250
6-well plate	250000	2	2	2	6	31	195	250
6-well plate	250000	2	2	4	4	31	195	250

Table 29. Transient co-transfection conditions using HEK293T cells.

HBS 2X

NaCl 274 mM

KCl 10 mM

$Na_2HPO_4.H_2O$ 1,5 mM

Dextrose 12 mM

Hepes 42 mM

PH 7±0,1

Day 2

- Change the medium by adding fresh medium (1 mL/well).
- Incubate at 37 ^{o}C, 5 % CO_2 and 100 % humidity overnight (12-16 hours).

2.2.17. Reporter Assays

For reporter assays the cells were plated in 6-well-plates and were subsequently transfected by the calcium phosphate method. Luciferase assays were performed using the luciferase assay kit from Promega Corp. according to the manufacturer's instructions. Normalization for transfection efficiency was performed by β-galactosidase assays.

- Discard medium from 6-well-plates.
- Add 1 mL/well 1X PBS (phosphate buffer saline) and shake gently.
- Discard PBS from 6-well-plates.

- Add 150 µL/well (for HepG2) or 200 µL/well (for HEK293T) of 1X Lysis buffer (Promega Lysis Buffer).
- Incubate on a shaker device at 260 rpm, for 10 minutes at room temperature (18-25 ^{o} C).
- Collect cells with the help of a cell scrapper and place the samples at -80 ^{o} C for at least 15 minutes.
- Thaw the samples for 2 minutes at 37 ^{o} C and vortex briefly afterwards.
- Centrifuge for 5 minutes at 13000 rpm at room temperature (18-25 ^{o} C).

After centrifugation, cell extracts are being used to evaluate luciferase and beta galactosidase activity.

Measurement of Luciferase activity

60 µL of cell extract are mixed with 60 µL of substrate (Promega) and the tube is placed in the luminometer device, where released relative light units are being counted.

Measurement of beta galactosidase activity

For each sample a mix is being prepared as following:

- 149 µL Sodium Phosphate Buffer
- 44 µL ONPG
- 2 µL 100X Salt Buffer

So, a final mix is being prepared, according to the number of samples each time.

In a 96-well-plate, add 5 µL of each sample cell extract. 5 µL of 1X Lysis buffer, are being used as blank.

With the use of Infinite M200 PRO reader device, 195 µL of the mix are being injected in each well at 37 ^{o} C. A measurement of the absorbance at 410 nm is being taken every 2 minutes for 25 cycles.

Sodium Phosphate buffer (P Buffer) 0,1M, pH 7.3:

Adjust the pH of Na_2HPO_4 0,1M solution to 7.3 using NaH_2PO_4 o,1M buffer.

O-nitrophenyl-galactopuranoside (ONPG) buffer:

ONPG is being diluted in P buffer at a final concentration of 8 mgr/mL.

100X Salt buffer:

KCl 3M	333 µL
$MgCl_2$	100 µL
β-mercaptoethanol	352 µL
Sterile H_2O	215 µL
Total Volume	1000 µL

2.2.18. Bioinformatics analysis

A bioinformatics analysis was performed using the online program AliBaba2.1 from www.gene-regulation.com, a program which gives a prediction of the transcription factor binding sites of a desired DNA sequence. We screened -3500/+186 Ces3b gene sequence for potential transcription factor binding sites. Moreover, another online tool was used, which is an HNF-4 Binding Site Scanner (http://nrmotif.ucr.edu/fuzzhtmlform.html) provided by Transcription factor encyclopedia (http://www.cisreg.ca/cgi-bin/tfe/home.pl), was used to search for potential HNF-4α binding sites in region spanning from 532 to 186 of Ces3b gene.

2.2.19. Statistical analysis

Data are presented as the average of normalized luciferase values %. Normalization was performed according to beta-galactosidase measurements. Microsoft Excel 2010 program was used for the statistical analysis of the results using t-test analysis, which returns the probability associated with a Student's two-sample equal variance (homoscedastic) t-test with a two-tailed distribution. For p values<0.053 the difference observed was statistically significant.

2.3. RESULTS

2.3.1. Overexpression of HNF-4 transcription factor up-regulates mouse Ces3b promoter activity

As mentioned above, microarrays in HNF-4 liver knock-out mice (H4LivKO) revealed a large number of genes whose expression was reduced in the absence of HNF-4 from the liver. Among these genes, mouse Ces3b was chosen for further investigation, since it was the carboxylesterase with the most significant fold change reduction.

In order to examine the effect of HNF-4 in Ces3b expression and the possible regulation of the latter by HNF-4, we constructed different plasmids bearing fragmentations of the mouse Ces3b promoter region (the mouse Ces3b promoter, first exon and part of first intron sequence is shown in the Appendix) in a luciferase reporter vector (pGL3 basic). The fragments of Ces3b promoter cloned in pGL3 basic vector were the following: (-1956/+186), (-1529/+186), (-1052/+186), (-532/+186), (-159/+186), (-80/+186) and (-80/+4) (Figure 49). The purpose of these plasmid constructions was to use them in transient transfection assays, so as to examine whether the presence of HNF-4 transcription factor regulates mouse Ces3b promoter activity. In order to succeed that, we transiently co-transfected HEK293T cells with the different plasmid constructs for Ces3b fragmentations, along with the absence or the presence of different concentrations of HNF-4 expression vector and a beta-galactosidase expression vector and we evaluated the relative luciferase activity (%) in each case.

The results from the transient co-transfection experiments are shown in Figure 50. HNF-4 was able to up-regulate the activity of the mouse Ces3b promoter fragments -1956/+186, -1529/+186, -1052/+186, -532/+186 and -159/+186 by ~5-fold in a dose-dependent manner. In the case of -80/+186 fragment, a stronger 16-fold increase in promoter activity was observed which was compromised when the region +5 to +186 of the promoter was deleted. These data suggest the potential existence of an HNF-4 binding site between +4 to +186 region, a region which spans the first exon of mouse Ces3b up to nucleotide at position +96, and part of the first intron (+97/+186).

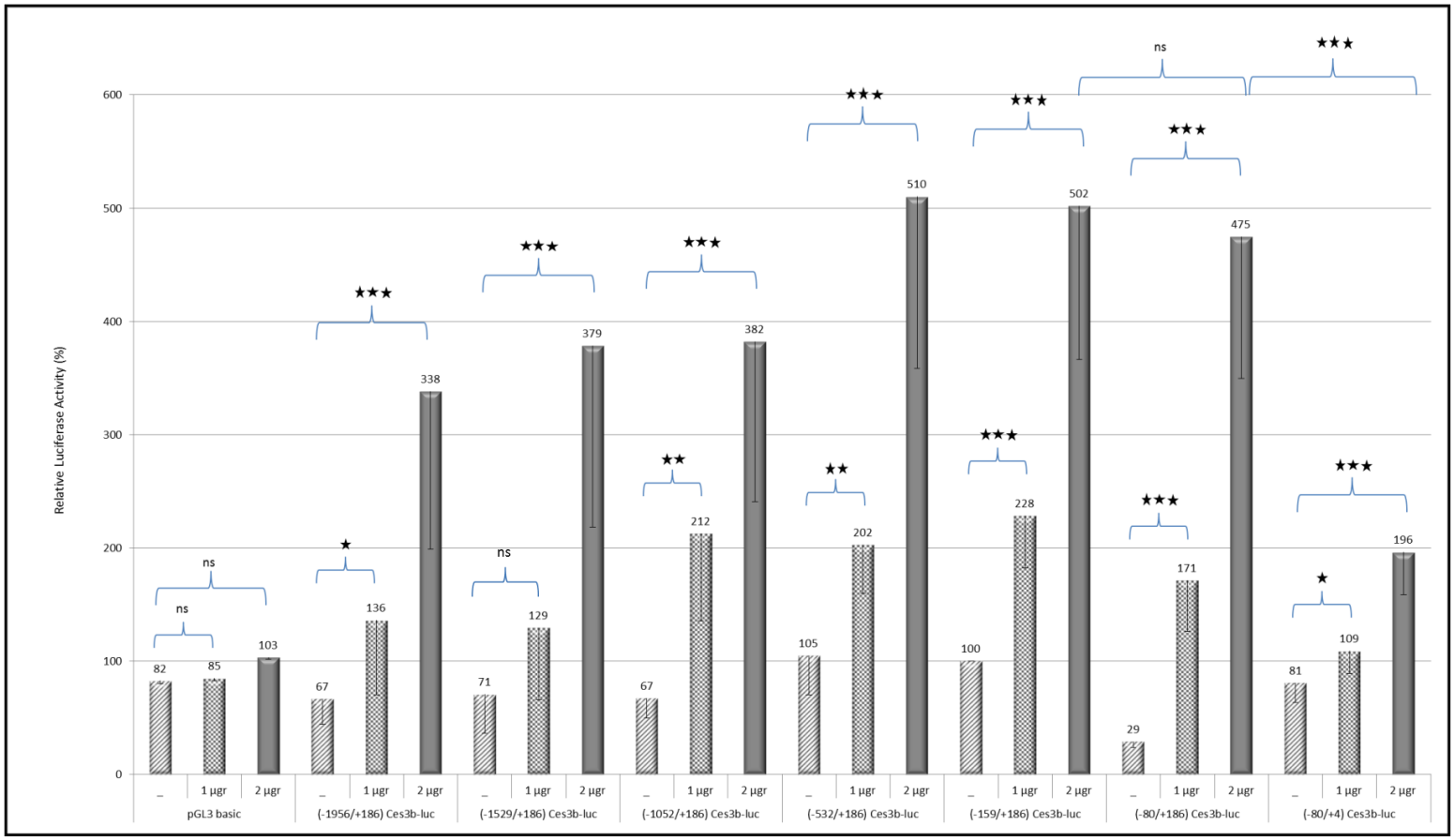

Figure 50. HNF-4 overexpression up-regulates mouse Ces3b promoter activity. HEK293T cells were transiently co-transfected with the luciferase reported plasmids expressing different fragments of the mouse Ces3b promoter (1µgr) along with an HNF-4 expression vector (1µgr or 2µgr) and a beta-galactosidase expression vector (1µgr). From left to right, the mouse Ces3b promoter fragments are the following: (-1956/+186) Ces3b-luc, (-1529/+186) Ces3b-luc, (-1052/+186) Ces3b-luc, (-532/+186) Ces3b-luc, (-159/+186) Ces3b-luc, (-80/+186) Ces3b-luc and (-80/+4) Ces3b-luc, whereas empty pGL3 basic reporter vector was used as a control. A statistical significance was observed for *p* value<0.053. Asterisks represent the statistical significance, * for *p* value<0.053, ** for *p* value<0.01, * for *p* value<0.001 and ns for non-significant.**

We performed a bioinformatics analysis using the online program AliBaba2.1 from www.gene-regulation.com, a program which gives a prediction of the transcription factor binding sites of a desired DNA sequence. Indeed, analysis of Ces3b promoter sequence (-3500/+186) with this tool revealed a wide variety of potential transcription factor binding sites in our promoter sequence. More specifically, an HNF-4α binding motif was recognized spanning from -2664 to -2673 upstream of +1 of Ces3b promoter (Figure 51). Moreover, the exact same region (-2664 to -2673) was identified for another transcription factor as well, COUP (Figure 51).

```
==========================================================================
seq(  780..  839)-2720actgcagataaatatacacacccttgctcaaattctcttcaccccaaccaaaggtcattc-2661
Class      lbp  rbp
2.1.2.11   827  836                                                  nyAArGkTCA
2.1.2.11.1 +R03881                                                   ccaaagttca
2.1.2.11.1 -R01183                                                   ttaaagttca
2.1.2.11.1 +R03035                                                   gcaagggtca
2.1.2.11.1 -R02657                                                   gcaaaggtca
2.1.2.11.1 +R03034                                                   gcaaggttca
                                                                     ||||||||||
                                                                     ||||||||||
                                                                      ||||||||
                                                                      ||| | |||
                                                                       || | |||

Segments:
3.3.2.0    782  791     ===HNF-3==
2.2.1.1    783  795      ====GATA-1===
4.5.1.0    788  797          ====TBP===
2.3.1.0    800  809                      ====Sp1===
2.3.2.2    802  811                        ===LyF-1==
1.1.3.0    803  812                         =C/EBPalp=
3.1.2.2    805  814                           ===Oct-1==
2.3.1.0    816  825                                     ====Sp1===
2.1.2.1    827  836                                                  =RAR-alph=
2.1.2.3    827  836                                                  =T3R-alpha
2.1.2.7    827  836                                                  ===FKBP59=
2.1.2.10   827  836                                                  ====COUP==
2.1.2.11   827  836                                                  =HNF-4alp=
==========================================================================
seq(  840..  899)-2660catttcttcatctacattctaactgttcacaccctaggctttcagcgtctgcttccacct-2661
Class      lbp  rbp
Segments:
3.1.2.1    841  850    ===Pit-1a=
2.1.1.1    858  867                  =====GR===
==========================================================================
seq(  900..  959)-2600tcttgtctggttcctccagtcacagcttcctgctccctgatcttcatgcccaacacctct-2541
Class      lbp  rbp
Segments:
2.1.1.4    915  924                =====ER===
2.3.1.0    925  934                         ====Sp1===
2.2.1.1    930  939                              ===GATA-1=
2.1.2.2    933  942                                 ==RXR-beta
3.1.2.2    940  949                                        ===Oct-1==
```

```
==========================================================================
seq(  780..  839)-2720actgcagataaatatacacacccttgctcaaattctcttcaccccaaccaaaggtcattc-2661
Class      lbp  rbp
2.1.2.10   827  836                                                  nmwmArGTCA
2.1.2.10.5 +R03931                                                   tcacaggtca
2.1.2.10.5 +R03932                                                   caaaaagtca
2.1.2.10.5 +R01193                                                   tcaaaggtca
2.1.2.10.5 -R02657                                                   gcaaaggtca
2.1.2.10.5 -R03405                                                   ccaaaggtca
2.1.2.10.5 -R01465                                                   cctcaggtca
                                                                     ||||||||||
                                                                     ||||||||||
                                                                      ||||||||
                                                                      || ||||||
                                                                         | ||||

Segments:
3.3.2.0    782  791     ===HNF-3==
2.2.1.1    783  795      ====GATA-1===
4.5.1.0    788  797          ====TBP===
2.3.1.0    800  809                      ====Sp1===
2.3.2.2    802  811                        ===LyF-1==
1.1.3.0    803  812                         =C/EBPalp=
3.1.2.2    805  814                           ===Oct-1==
2.3.1.0    816  825                                     ====Sp1===
2.1.2.1    827  836                                                  =RAR-alph=
2.1.2.3    827  836                                                  =T3R-alpha
2.1.2.7    827  836                                                  ===FKBP59=
2.1.2.10   827  836                                                  ====COUP==
2.1.2.11   827  836                                                  =HNF-4alp=
==========================================================================
seq(  840..  899)-2660catttcttcatctacattctaactgttcacaccctaggctttcagcgtctgcttccacct-2661
Class      lbp  rbp
Segments:
3.1.2.1    841  850    ===Pit-1a=
2.1.1.1    858  867                  =====GR===
==========================================================================
seq(  900..  959)-2600tcttgtctggttcctccagtcacagcttcctgctccctgatcttcatgcccaacacctct-2541
Segments:
2.1.1.4    915  924                =====ER===
2.3.1.0    925  934                         ====Sp1===
2.2.1.1    930  939                              ===GATA-1=
2.1.2.2    933  942                                 ==RXR-beta
3.1.2.2    940  949                                        ===Oct-1==
```

Figure 51. Bioinformatics analysis of -3500/+186 sequence of mouse Ces3b gene using AliBaba2.1 online tool from www.gene-regulation.com. Part of this sequence (-2720/-2541) is being demonstrated in this figure. An HNF-4α (underlined with orange) as well as a COUP (underlined with yellow) potential binding site is being recognized, from -2664 to -2673.

Even though the analysis show an HNF-4α binding motif upstream the position -1956, another motif should exist downstream from position -1956 since we observed up-regulation in the activity of all our promoter fragments that lack this distal region. We focused our interest, in the region between +4 to +186, since as shown in Figure 50, luciferase levels between constructs (-80/+186) Ces3b-luc and (-80/+4) Ces3b-luc differed significantly, with a great reduction in the relative luciferase activity (%) observed in (-80/+4) Ces3b-luc, indicating that there might was a site for HNF-4 which was lost from fragmentation -80/+186 to fragmentation -80/+4.

So, using again AliBaba2.1 online tool, as well as another online tool, which is an HNF-4 Binding Site Scanner (http://nrmotif.ucr.edu/fuzzhtmlform.html) provided by Transcription factor encyclopedia (http://www.cisreg.ca/cgi-bin/tfe/home.pl), we examined the region upstream and downstream to +1 for potential transcription factor binding sites. Analysis with AliBaba2.1 online tool, revealed a binding motif for COUP transcription factor at position +57 to +67 (Figure 52), whereas analysis with the online HNF-4 Binding Site Scanner revealed an HNF-4α binding motif at position +36 to +48 (Figure 53).

```
================================================================================
seq( 3420.. 3479) -80 gcagaccaagttaattttcacaagataagaggtactataaaaggcaagtcctcccagctt  -21
Class      lbp  rbp
Segments:
1.1.3.0     3435 3444                 =C/EBPalp=
2.2.1.1     3439 3450                     ====GATA-1==
4.5.1.0     3455 3464                                 ====TBP===
9.9.820     3455 3464                                 ===TFIID==
1.1.3.0     3478 3487                                                        =C
================================================================================
seq( 3480.. 3539) -20 tccacagcttcgtccacagtccgttatgacaaatatgaggacaatgataccagctgggtc +40
Class      lbp  rbp
Segments:
1.1.3.0     3478 3487  /EBPalp=
4.3.2.0     3507 3516                            ====SRF===
4.3.2.0     3513 3522                                 ====SRF===
2.2.1.1     3519 3528                                      ===GATA-1=
3.1.2.2     3520 3529                                       ===Oct-1==
1.1.3.0     3522 3531                                         =C/EBPalp=
1.2.1.0     3527 3536                                              ====HEB===
================================================================================
seq( 3540.. 3599) +41 cagtgtcctagtctgggtgacctgtctgctcctggcatttgttaccacagtcactggtaa +100
Class      lbp  rbp
2.1.2.10    3556 3565                 nTGACCTnys
2.1.2.10.5 -R01812                    gtgaccttcc
2.1.2.10.5 -R03931                    gtgacctgtg
2.1.2.10.5 -R01457                    ttgacctccc
2.1.2.10.5 +R02657                    gtgacctttg
2.1.2.10.5 +R03405                    ctgacctttg
2.1.2.10.5 -R01193                    ttgacctttg
                                      ||||||||||
                                      ||||||||||
                                       |||||| ||
                                       ||||||
                                       ||||||

Segments:
1.1.1.1     3551 3560            ===c-Jun==
2.1.2.10    3556 3565                 ====COUP==
2.3.3.0     3556 3565                 =CPE bind=
2.3.1.0     3570 3579                                ====YY1===
1.1.3.0     3571 3580                                 =C/EBPbeta
1.1.3.0     3580 3589                                          =C/EBPalp=
```

Figure 52. Bioinformatics analysis of -159/+186 sequence of mouse Ces3b gene using AliBaba2.1 online tool from www.gene-regulation.com. Part of this sequence (-80/+100) is being demonstrated in this figure. A COUP potential binding site is being recognized, from +57 to +67, underlined with yellow. +1 position is being indicated in the blue box. The translation initiation codon (ATG) is underlined with green. The potential HNF-4α binding site shown in Figure 53 is underlined with orange.

nrmotif.ucr.edu/cgi-bin/htmlfuzz2.cgi

HNF4 Motif Finder

back

```
Fri Oct 30 08:56:37 PDT 2015
-----------------------Coordinates for Exact Match--------------------

        H4seq    H4num    SeqName start    end      length

GGACACTGGACCC PBM.1028 CTCTGATTATTTGACC 1988    1976    13      -       pattern1



-----------------------Sequence for Exact Match--------------------
```

CTCTGATTATTTGACC

```
             1690      1700      1710      1720      1730      1740
     ----:----|----:----|----:----|----:----|----:----|----:----|
-260 accatgccaggcaagagtagaatcatccacagaggaggggtggtggtggtgggagtgggg -201

             1750      1760      1770      1780      1790      1800
     ----:----|----:----|----:----|----:----|----:----|----:----|
-200 gtagtgggtggggtgagggtgttaaatagtcagacctctccagaagactctgctggtcct -141

             1810      1820      1830      1840      1850      1860
     ----:----|----:----|----:----|----:----|----:----|----:----|
-140 gctctgagaacaggcagagcgtagaatacacaattccagagctggcctccaaccaatcgg -81

             1870      1880      1890      1900      1910      1920
     ----:----|----:----|----:----|----:----|----:----|----:----|
-80  gcagaccaagttaattttcacaagataagaggtactataaaaggcaagtcctcccagctt -21

             1930      1940      1950      1960      1970      1980
     ----:----|----:----|----:----|----:----|----:----|----:----|
-20  tccacagcttcgtccacagtccgttatgacaaatatgaggacaatgataccagctGGGTC +40

             1990      2000      2010      2020      2030      2040
     ----:----|----:----|----:----|----:----|----:----|----:----|
+41  CAGTGTCCtagtctgggtgacctgtctgctcctggcatttgttaccacagtcactggtaa +100

             2050      2060      2070      2080      2090      2100
     ----:----|----:----|----:----|----:----|----:----|----:----|
+101 gacatactccctgaaccagtgaagtggggcttccagctccgtgggaaacctgaggaaggg +160

             2110      2120
     ----:----|----:----|----:-
+161 gggactggagtgaggcagtaagagtg                                   +186
```

Figure 53. Analysis of -532/+186 sequence of mouse Ces3b gene using HNF-4 Binding Site Scanner (http://nrmotif.ucr.edu/fuzzhtmlform.html). Part of this sequence (-260/+186) is being demonstrated in this figure. An HNF-4α binding site is being demonstrated in red letters, from +36 to +67. +1 position is being indicated in the blue box. The translation initiation codon (ATG) is underlined with green.

2.4. DISCUSSION

Carboxylesterases are key participants in a variety of processes, such as drug metabolism through hydrolysis of a wide range of amide- and ester-containing compounds, biotransformation of numerous drugs and prodrugs such as irinotecan, detoxification of environmental toxicants and conversion of pro-carcinogens into carcinogens, a process taking place in the liver. Numerous endogenous compounds are substrates for carboxylesterases such as short- and long-chain acyl-glycerols, palmitoyl-coenzyme A, as well as long- and medium-chain acylcarnitines (Staudinger *et al.*, 2010). Carboxylesterases bear several potential binding sites for transcription factors such as Sp1, Sp3, USF1, NFkB, C/EBP, NF-1, PPARa, SREBP, GR, HNF-1, HNF-3 and HNF-4 (Satoh *et al.*, 2006). In 2006, a study from Furihata *et al.* demonstrated that HNF-4α plays a pivotal role in the transcriptional regulation of mouse Ces2 gene in the liver, by enhancing its transcription. It was shown that involvement of HNF-4α is among the reasons for high liver expression levels of mouse Ces2, and that bile acid can repress its transcription through repression of HNF-4α mediated transactivation (Furihata *et al.*, 2006).

Recently, studies in our laboratory demonstrated through microarrays in HNF-4α liver knock-out mice, that a significant number of genes are down-regulated in the absence of HNF-4 from mouse liver. Among these genes, several genes from the family of carboxylesterases were identified, with Ces3b being among those with the highest fold change decrease in expression (unpublished data). So it was reasonable to hypothesize, that perhaps the expression levels of mouse Ces3b in the liver are HNF-4α-dependent.

HNF-4α is mainly expressed in liver, pancreas, intestine and kidney and appears to be critical for transcriptional regulation of many hepatic genes such as CAR, Cyp7a1 and genes involved in a variety of hepatic functions, such as cholesterol and glucose metabolism, lipid homeostasis, as well as drug metabolism (Hosokawa *et al.*, 2007; Staudinger *et al.*, 2010). It is well known that HNF-4α binds predominantly direct repeats with a spacing of one nucleotide (DR1), whereas in most cases the spacer nucleotide tends to be an adenine (A). Over the years, a great variety of motifs for HNF-4α have been identified in various gene targets and HNF-4α is predicted to bind thousands of different variations of the DR1 (Badis *et al.*, 2009). So it is reasonable that some bioinformatics tools may recognize a potential HNF-4α binding site and others may not be able to recognize the same site, and this is due to the specificity and sensitivity of each tool. For instance, an HNF-4α binding motif is recognized by the online tool HNF-4 Binding Site Scanner (http://nrmotif.ucr.edu/fuzzhtmlform.html) at position +36 to +48 (Figure 53), however the online tool AliBaba2.1 (from www.gene-regulation.com) does not identify this motif (Figure 52). Taking into account the motif shown in Figure 53, 5'-GGGTCCAGTGTCC-3', it is obvious that we have to deal with a DR1 binding motif. As mentioned earlier, this specific motif is located between +36 and +48 downstream

from +1, an area that belongs to the first exon of Ces3b gene (Figure 49). Studies over the years have quite well demonstrated the regulation of promoter regions by transcription factors binding to exonic regions. For example, Fernandez-Rachubinski *et al.* in 1996 demonstrated that regions flanking the first exon of human antithrombin gene can regulate its expression. This was demonstrated through evaluation of the serial deletions performed, in regions downstream and upstream of +1, by a luciferase-based reporter assay in transiently transfected HepG2, COS1, Hela and BSC40 cells (Fernandez-Rachubinski *et al.*, 1996). So it is reasonable to hypothesize that the +36/+48 HNF-4α motif observed in the first exon of Ces3b gene, may be the one responsible for the increased Ces3b promoter activity in the presence of HNF-4.

Another important transcription factor motif has been identified in the region downstream to +1 in Ces3b gene, a motif for COUP (5'-GTGACCTGTCT-3'), at position +57 to +67inside the first exon (Figure 52). Chicken ovalbumin upstream promoter (COUP) transcription factors (COUP-TFs) belong to the family of orphan receptors of steroid/thyroid hormone receptor superfamily. They exhibit important roles in the regulation of neurogenesis, organogenesis and differentiation during embryonic development. Two genes have been identified in mammals, COUP-TFI (also designated as EAR3) and COUP-TFII (also known as ARP-1). COUP-TFs homodimerize or heterodimerize with retinoid X receptor (RXR) and a few other nuclear receptors, binding to a variety of response elements, which contain imperfect 5'-AGGTCA-3' inverted or direct repeats with various spacings and palindromes (Park *et al*., 2003). It has been reported in several studies, that HNF-4α and COUPTF2 share common recognition sequence, which barely resemble the classical DR1s, in the promoters of certain genes, resulting in a competition for transcriptional regulation (Mietus-Snyder *et al.*, 1992; Nakshartri *et al.*, 1998; McNair *et al*., 2000). In these promoters COUP-TF acts to repress HNF-4-mediated activation, through competition for the same binding site (Ladias *et al.*, 1992; Mietus-Snyder *et al.*, 1992). So if we consider, that HNF-4α binds to COUP-TF sites, it is reasonable to hypothesize that perhaps the COUP site downstream from position +4 (+57 to +67) in Ces3b gene, is the motif where HNF-4α binds in our experiments. Therefore, when this site is lost in plasmid construction (-80/+4) Ces3b-luc the expression levels of luciferase are significantly reduced (Figure 50).

Future experiments however need to be done, in order to be sure that indeed HNF-4α binds to Ces3b gene in a region spaning the first exon, and through this binding up-regulates the transcription of the gene. First of all, plasmids (-532/+4) Ces3b-luc and (-159/+4) Ces3b-luc have to be constructed, in order to compare different upstream ends of the promoter with the same downstream end along with the construct (-80/+4) Ces3b-luc that we already have. Moreover, in order to examine, whether HNF-4α binds to HNF-4α motif recognized at position

+36/+48 or to COUP motif recognized at position +57/+67, another plasmid construction has to be made, where the upstream end will be at position -80, whereas the downstream end will include the HNF-4α site and exclude the COUP site, so it must be somewhere between +49 to +56. Furthermore, Chromatin Immunoprecipitation (ChIP) assays have to be performed, using an antibody against HNF-4α and analyze thereafter the immunoprecipitated chromatin by PCR using specific primers, in order to determine exactly the DNA binding site of HNF-4α between the +4/+186 region. What is more, DNA Affinity Precipitation (DNAP) would be another valuable assay to demonstrate the binding of HNF-4α to a specific DNA sequence. An oligonucleotide primer coupled with beads could be designed, so as to detect nuclear receptor binding by SDS-PAGE and immunoblotting using specific antibody for HNF-4α. Finally, Site-Directed Mutagenesis could be performed to the target site, so as to prove that when altering one or two nucleotides in the motif sequence, HNF-4α is not able to bind and regulate transcription of Ces3b gene.

Since Ces3b belongs to the family of carboxylesterases and owning to the fact that it is mainly expressed in the liver, this enzyme might be implicated in processes such as hydrolysis of endogenous compounds, metabolic activation or detoxification of various drugs and environmental toxicants, the assembly of low-density lipoprotein particles in the liver and several other metabolic reactions. So, unraveling possible ways of Ces3b transcriptional regulation might be critical to elucidate the mechanisms by which this gene is implicated in processes taking place in the mouse liver, a fact that will give us valuable information for possible role, implication in metabolism and regulation of its homolog in human.

2.5. REFERENCES

- Aida Kaoru, Moore Rickie, Negishi Masahiko (1993) Cloning and nucleotide sequence of a novel, male-predominant carboxylesterase in mouse liver. *Biochemica et Biophysica Acta* **1174**: 72-74.
- Alam Mustafa, Vance Dennis E., Lehner Richard (2002a) Structure-function analysis of human triacylglycerol hydrolase by site-directed mutagenesis: identification of the catalytic triad and a glycosylation site. *Biochemistry* **41**: 6679-6687.
- Alam Mustafa, Ho Samuel, Vance Dennis E., Lehner Richard (2002b) Heterologous expression, purification and characterization of human triacylglycerol hydrolase. *Protein Expression and Purification* **24**: 33-42.
- Aldridge WN. (1953) Serum esterases. I. Two types of esterase (A and B) hydrolysing p-nitrophenyl acetate, propionate and butyrate, and a method for their determination. *Biochem J* **53**: 110–117.
- Babeu Jean-Philippe, Boudreau François (2014) Hepatocyte nuclear factor 4-alpha involvement in liver and intestinal inflammatory networks. *World J Gastroenterol* **20(1)**: 22-30.
- Badis G., Berger M.F., Philippakis A.A., Talukder S., Gehrke A.R., Jaeger S.A., Chan E.T., Metzler G., Vedenko A., Chen X., Kuznetsov H., Wang C.F., Coburn D., Newburger D.E., Morris Q., Hughes T.R., Bulyk M.L. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* **324(5935)**: 1720-1723.
- Bahar Goksin Fatma, Ohura Kayoko, Ogihara Takuo, Imai Teruko (2012) Species difference of esterase expression and hydrolase activity in plasma. *Journal of Pharmaceutical Sciences* **101**(10): 3979-3988.
- Barrero M.J., Malik S. (2006) Two functional modes of a nuclear receptor-recruited arginine methyltransferase in transcriptional activation. *Mol Cell* **24(2)**: 233-243.
- Becker A, Bottcher A, Lackner KJ, Fehringer P, Notka F, et al. (1994) Purification, cloning and expression of a human enzyme with acyl coenzyme A: cholesterol acyltransferase activity, which is identical to liver carboxylesterase. *Arterioscler Thromb.* **14**: 1346–1355.
- Bencharit S, Morton CL, Xue Y, Potter PM, Redinbo MR. (2003) Structural basis of heroin and cocaine metabolism by a promiscuous human drug-processing enzyme. *Nat Struct Biol*. **10**: 349–356.
- Bencharit S, Edwards CC, Morton CL, Howard-Williams EL, Kuhn P, et al. (2006) Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1. *J Mol Biol.* **363**: 201–214.
- Bergmann F, Segal R, Rimon S. (1957) A new type of esterase in hog-kidney extract. *Biochem J* **67**: 481–486.
- Bogan A.A., Dallas-Yang Q., Ruse M.D. Jr., Maeda Y., Jiang G., Nepomuceno L., Scanlan T.S., Cohen F.E., Sladek F.M. (2000) Analysis of protein dimerization

and ligand binding of orphan receptor HNF4alpha. *J Mol Biol.* **302(4)**: 831-851.

- Bolotin E, Schnabl J, Sladek FM. HNF4A (Homo sapiens). In Transcription Factor Encyclopedia (http://www.cisreg.ca/tfe) (Updated August 26, 2010) This continually updated, on-line review provides an overview of the most important aspects of HNF4α as well as comprehensive lists of target genes and interacting proteins with direct links to the literature.
- Bolotin E., Liao H., Ta T.C., Yang C., Hwang-Verslues W., Evans J.R., Jiang T., Sladek F.M. (2010) Integrated approach for the identification of human hepatocyte nuclear factor 4α target genes using protein binding microarrays. *Hepatology* **51(2)**: 642-653.
- Chen W.S., Manova K., Weinstein D.C., Duncan S.A., Plump A.S., Prezioso V.R., Bachvarova R.F., Darnell J.E. Jr. (1994) Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. *Genes Dev.* **8(20)**: 2466-2477.
- Chen T. (2008) Nuclear receptor drug discovery. *Curr. Opin. Chem. Biol.* **12**: 418–426.
- Cotnoir-White D., Laperriere D., Mader S. (2011) Evolution of the repertoire of nuclear receptor binding sites in genomes. *Mol. Cell Endocrinol.* **334**: 76–82.
- Cygler M, Schrag JD, Sussman JL, Harel M, Silman I, et al. (1993) Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases and related proteins. *Protein Sci.* **2**: 366–382.
- Diczfalusy MA, Bjorkkem I, Einarsson C, Hillebrant CG, Alexson SE. (2001) Characterization of enzymes involved in formation of ethyl esters of long-chain fatty acids. *J Lipid Res.* **42**: 1025–1032.
- Dolinsky VW, Sipione S, Lehner R, Vance DE. (2001) The cloning and expression of murine triacylglycerol hydrolase cDNA and the structure of the corresponding gene. *Biochim Biophys Acta* **1532**: 162–172.
- Drewes T., Senkel S., Holewa B., Ryffel G.U. (1996) Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. *Mol Cell Biol* **16**: 925-931.
- Evans R.M. (1988) The steroid and thyroid hormone receptor superfamily. *Science* **240**: 889–895.
- Fang Bin, Mane-Padros Daniel, Bolotin Eugene, Jiang Tao, Sladek Frances M. (2012) Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. *Nucleic Acids Research* **40(12)**: 5343-5356.
- Fernandez-Rachubinski Françoise A., Weineri Joel H., Blajchman Morris A. (1996) Regions flanking exon 1 regulation constitutive expression of the human antithrombin gene. *The Journal of Biological Chemistry* **271(46)**: 29502-29512.

- Fleming CD, Bencharit S, Edwards CC, Hyatt JL, Tsurkan L, et al. (2005) Structural insights into drug processing by human carboxylesterase 1: tamoxifen, mevastatin, and inhibition by Benzil. *J Mol Biol.* **352**: 165–177.
- Furihata T, Hosokawa M, Nakata F, Satoh T, Chiba K. (2003) Purification, molecular cloning, and functional expression of inducible liver acylcarnitine hydrolase in C57BL/6 mouse, belonging to the carboxylesterase multigene family. *Arch Biochem Biophys.* **416**: 101–109.
- Furihata T., Hosokawa M., Masuda M., Satoh T., Chiba K. (2006) Hepatocyte nuclear factor-4alpha plays pivotal roles in the regulation of mouse carboxylesterase 2 gene transcription in mouse liver. *Arch. Biochem. Biophys.* **447**: 107–117.
- Furuta H., Iwasaki N., Oda N., Hinokio Y., Horikawa Y., Yamagata K., Yano N., Sugahiro J., Ogata M., Ohgawara H., Omori Y., Iwamoto Y., Bell G.I. Organization and partial sequence of the hepatocyte nuclear factor-4 alpha/MODY1 gene and identification of a missense mutation, R127W, in a Japanese family with MODY. *Diabetes* **46**: 1652-1657.
- Genetta TL, D'Eustachio P, Kadner SS, Finlay TH. (1988) cDNA cloning of esterase 1, the major esterase activity in mouse plasma. *Biochem Biophys Res Commun*. **151**: 1364–1370.
- Ghosh S. (2000) Cholesteryl ester hydrolase in human monocyte/macrophage: cloning, sequencing and expression of full-length cDNA. *Physiol Genomics* **2**: 1–8.
- Ghosh S, Mallonee DH, Grogan WM. (1995) Molecular cloning and expression of rat hepatic neutral cholesteryl ester hydrolase. *Biochim Biophys Acta* **1259**: 305–312.
- Gonzalez FJ. (1998) The study of xenobiotic-metabolizing enzymes and their role in toxicity in vivo using targeted gene disruption. *Toxicol Lett* **102–103**: 161–166.
- Green V.J., Kokkotou E., Ladias J.A. (1998) Critical structural elements and multitarget protein interactions of the transcriptional activator AF-1 of hepatocyte nuclear factor 4. *J Biol Chem.* **273(45)**: 29950-29957.
- Guo H., Gao C., Mi Z., Wai P.Y., Kuo P.C. (2006) Phosphorylation of Ser158 regulates inflammatory redox-dependent hepatocyte nuclear factor-4alpha transcriptional activity. *Biochem J.* **394**: 379-387.
- Hadzopoulou-Cladaras M., Kistanova E., Evagelopoulou C., Zeng S., Cladaras C., Ladias J.A. (1997) Functional domains of the nuclear receptor hepatocyte nuclear factor 4. *J Biol Chem.* **272(1)**: 539-550.
- Hansen S.K., Párrizas M., Jensen M.L., Pruhova S., Ek J., Boj S.F., Johansen A., Maestro M.A., Rivera F., Eiberg H., Andel M., Lebl J., Pedersen O., Ferrer J., Hansen T. (2002) Genetic evidence that HNF-1alpha-dependent

transcriptional control of HNF-4alpha is essential for human pancreatic beta cell function. *J Clin Invest* **110**: 827-833.

- Harries L.W., Locke J.M., Shields B., Hanley N.A., Hanley K.P., Steele A., Njølstad P.R., Ellard S., Hattersley A.T. (2008) The diabetic phenotype in HNF4A mutation carriers is moderated by the expression of HNF4A isoforms from the P1 promoter during fetal development. *Diabetes* **57(6)**: 1745-1752.
- Hayhurst G.P., Lee Y.H., Lambert G., Ward J.M., Gonzalez F.J. (2001) Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. *Moll Cell Biol.* **21(4)**: 1393-1403.
- Holmes RS, Cox LA, VandeBerg JL. (2008a) Mammalian carboxylesterase 5: comparative biochemistry and genomics. *Comp Biochem Physiol D Genomics Proteomics* **3**: 195–204.
- Holmes RS, Chan J, Cox LA, Murphy WJ, VandeBerg JL. (2008b) Opossum carboxylesterases: sequences, phylogeny and evidence for CES duplication events predating the marsupial-eutherian common ancestor. *BMC Evol Biol.* **8**: 54.
- Holmes RS, VandeBerg JL, Cox LA. (2009a) A new class of mammalian carboxylesterase *CES6*. *Comp Biochem Physiol Part D Genomics Proteomics* **4**: 209–217.
- Holmes RS, Glenn JP, VandeBerg JL, Cox LA. (2009b) Baboon carboxylesterases 1 and 2: sequences, structures and phylogenetic relationships with human and other primate carboxylesterases. *J Med Primatol*. **38**: 27–38.
- Holmes RS, Cox LA, VandeBerg JL. (2010a) Mammalian carboxylesterase 3: comparative genomics and proteomics. *Genetica* **138(7)**: 695–708.
- Holmes RS, Wright MW, Laulederkind SJF, Cox LA, Hosokawa M, Imai T, Ishibashi S, Lehner R, Miyazaki M, Perkins EJ, Potter PM, Redinbo MR, Robert J, Satoh T, Yamashita T, Yan B, Yokoi T, Zehner R, Maltais LJ (2010b) Recommended nomenclature for five mammalian carboxylesterase gene families: human, mouse, and rat genes and proteins. *Mamm Genome* **21(9-10)**: 427-441.
- Hong Y.H., Varanasi U.S., Yang W., Leff T. (2003) AMP-activated protein kinase regulates HNF4alpha transcriptional activity by inhibiting dimer formation and decreasing protein stability. *J Biol Chem.* **278(30)**: 27495-27501.
- Hosokawa Masakiyo, Furihata Tomomi, Yaginuma Yumiko, Yamamoto Naoko, Koyano Nao, Fujii Ayako, Nagahara Yuko, Satoh Testuo, Chiba Kan (2007) Genomic structure and transcriptional regulation of the rat, mouse and human carboxylesterase genes. *Drug Metabolism Reviews* **39**: 1-15.
- Huang J., Karakucuk V., Levitsky L.L., Rhoads D.B. (2008) Expression of HNF4alpha variants in pancreatic islets and Ins-1 beta cells. *Diabetes Metab Res Rev.* **24(7)**: 533-543.
- Hwang-Verslues Wendy W., Sladek Frances M. (2010) HNF4α-role in drug metabolism and potential drug target? *Current Opinion in Pharmacology* **10**: 698-705.

- Ishikawa F., Nose K., Shibanuma M. (2008) Downregulation of hepatocyte nuclear factor-4alpha and its role in regulation of gene expression by TGF-beta in mammary epithelial cells. *Exp Cell Res* **314(10)**: 2131-2140,
- Jiang G., Nepomuceno L., Hopkins K., Sladek F.M. (1995) Exclusive homodimerization of the orphan receptor hepatocyte nuclear factor 4 defines a new subclass of nuclear receptors. *Mol Cell Biol.* **15(9)**: 5131-5143.
- Jiang G., Lee U., Sladek F.M. (1997a) Proposed mechanism for the stabilization of nuclear receptor DNA binding via protein dimerization. *Mol Cell Biol.* **17(11)**: 6546-6554.
- Jiang Guoqiang, Sladek Frances M. (1997b) The DNA binding domain of hepatocyte nuclear factor 4 mediates cooperative, specific binding to DNA and heterodimerization with the Retinoid X Receptor α. *The Journal of Biological Chemistry* **272(2)**: 1218-1225.
- Jiang G., Nepomuceno L., Yang Q., Sladek F.M. (1997c) Serine/threonine phosphorylation of orphan receptor hepatocyte nuclear factor 4. *Arch Biochem Biophys.* **340(1)**: 1-9.
- Jones Ryan D., Taylor Anna M., Tong Ernest Y., Repa Joyce J. (2013) Carboxylesterases are uniquely expressed among tissues and regulated by nuclear hormone receptors in the mouse. *Drug Metab Dispos* **41**: 40-49.
- Krishnasamy R, Teng AL, Dhand R, Schultz RM, Gross NJ. (1998) Molecular cloning, characterization and differential expression pattern of mouse lung surfactant convertase. *Am J Physiol Lung Mol Cell Biol*. **275**: L969–L975.
- Kritis A.A., Argyrokastritis A., Moschonas N.K., Power S., Katrakili N., Zannis V.I., Cereghini S., Talianidis I. (1996) Isolation and characterization of a third isoform of human hepatocyte nuclear factor 4. *Gene* **173**: 275-280.
- Kroetz DL, McBride OW, Gonzalez FJ. (1993) Glycosylation-dependent activity of Baculovirus-expressed human liver carboxylesterases: cDNA cloning and characterization of two highly similar enzyme forms. *Biochemistry* **32**: 11606–11617.
- Ktistaki E., Ktistakis N.T., Papadogeorgaki E., Talianidis I. (1995) Recruitment of hepatocyte nuclear factor 4 into specific intranuclear compartments depends on tyrosine phosphorylation that affects its DNA-binding and transactivation potential. *Proc Natl Acad Sci U S A* **92(21)**: 9876-9880.
- Kurokawa R., Yu V., Naar A., Kyakumoto S., Han Z., Silverman S., Rosenfeld M., Glass C. (1993) Differential orientations of the DNA-binding domain and carboxyterminal dimerization interace regulate binding site selection by nuclear receptor heterodimers. *Genes Dev.* **7:** 1423–1435.
- Ladias J. A. A., Cladaras-Hadzopoulou M., Kardassis D., Cardot P., Cheng J., Zannis V. I., Cladaras C. (1992) Transcriptional regulation of human apolipoprotein genes ApoB, ApoCIII, and ApoAII by members of the steroid hormone

receptor superfamily HNF-4, ARP-1, EAR-2, and EAR-3. *J. Biol. Chem.* **267**: 15849-15860.

- Langmann T, Becker A, Aslanidis C, Notka F, Ulrich H, et al. (1997) Structural organization and characterization of the promoter region of a human carboxylesterase gene. *Biochim Biophys Acta.* **1350**: 65–74.
- Laudet V., Thompson B., Pratt W., Claessens F., Kraus L., Baniahmad A., Edwards D., Trifiro M., Chrousos G., Gustafsson J. et al. (2004) The nuclear receptor superfamily. In: McEwan,I. (ed.). *Essays in Biochemistry* **40** Portland Press Ltd, London.
- Lehner Richard, Cui Zheng, Vance Dennis E. (1999) Subcellular localization, developmental expression and characterization of a liver triacylglycerol hydrolase. *Biochem. J.* **338**: 761-768.
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen T-M, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong H-W, Dougherty JG, Duncan BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR, Fliss TP, Frensley C, Gates SN, Glattfelder KG, Halverson KR, Hart MR, Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawal R, Kidney JM, Knapik RH, Kuan CL, Lake JH, Laramee AR, Larsen KD, Lau C, Lemon TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE, Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA, Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova NV, Sivisay T, Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN, Stumpf K-R, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY, Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B, Zwingman TA, Jones AR. (2007) Genome wide atlas of gene expression in the mouse brain. *Nature* **145**: 168–176.
- Li J., Ning G., Duncan S.A. (2000) Mammalian hepatocyte differentiation requires the transcription factor HNF-4α. *Genes Dev* **14**: 464-474.
- Lockridge O, Adkins S, La Due BN. (1987) Location of disulfide bonds within the sequence of human serum cholinesterase. *J Biol Chem.* **262**: 12945–12952.
- Marsh S, Xiao M, Yu J, Ahluwalia R, Minton M, et al. (2004) Pharmacogenomic assessment of carboxylesterases 1 and 2. *Genomics* **84**: 661–668.
- McNair Alan, Cereghini Silvia, Brand Heike, Smith Terry, Breillat Christelle, Gannon Frank (2000) Synergistic activation of the Atlantic salmon hepatocyte nuclear factor (HNF) 1 promoter by the orphan nuclear receptors HNF4 and chicken ovalbumin upstream promoter transcription factor I (COUP-TFI). *Biochem. J.* **352**: 557-564.

- Mietus-Snyder M., Sladek F.M., Ginsburg G.S., Kuo C.F., Ladias J.A., Darnell J.E. Jr, Karathanasis S.K. (1992) Antagonism between apolipoprotein AI regulatory protein 1, Ear3/COUP-TF, and hepatocyte nuclear factor 4 modulates apolipoprotein CIII gene expression in liver and intestinal cells. *Mol. Cell. Biol.* **12**: 1708–1718.
- Miyazaki M, Kamiie K, Soeta S, Taira H, Yamashita T.(2003) Molecular cloning and characterization of a novel carboxylest-erase-like protein that is physiologically present at high concentrations in the urine of domestic cats (Felis catus). *Biochem J.* **370**: 101–110.
- Miyazaki M, Yamashita T, Suzuki Y, Saito Y, Soeta S, et al. (2006) A major urinary protein of the domestic cat regulates the production of felinine, a putative pheromone precursor. *Chem Biol.* **13**: 1070–1079.
- Munger JS, Shi GP, Mark EA, Chin DT, Gerard C, et al. (1991) A serine esterase released by human alveolar macrophages is closely related to liver microsomal carboxylesterases. *J Biol Chem*. **266**: 18832–18838.
- Nakhei H., Lingott A., Lemm I., Ryffel G.U. (1998) An alternative splice variant of the tissue specific transcription factor HNF4alpha predominates in undifferentiated murine cell types. *Nucleic Acids Res* **26**: 497-504.
- Nakshatri H., Bhat-Nakshatri P. (1998) Multiple parameters determine the specificity of transcriptional response by nuclear receptors HNF-4, ARP-1, PPAR, RAR and RXR through common response elements. *Nucleic Acids Res.* **26**: 2491–2499.
- Nammo T., Yamagata K., Tanaka T., Kodama T., Sladek F.M., Fukui K., Katsube F., Sato Y., Miyagawa J., Shimomura I. (2008) Expression of HNF-4alpha (MODY1), HNF-1beta (MODY5), and HNF-1alpha (MODY3) proteins in the developing mouse pancreas. *Gene Expr Patterns* **8(2)**: 96-106.
- Odom D.T., Zizlsperger N., Gordon D.B., Bell G.W., Rinaldi N.J., Murray H.L., Volkert T.L., Schreiber J., Rolfe P.A., Gifford D.K., Fraenkel E., Bell G.I., Young R.A. (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303(5662)**: 1378-1381.
- Ohtsuka H, Inoue S, Kameyama M. (2003) Intracellular conversion of irinotecan to its active form, SN-38, by native carboxylesterase in human non-small cell lung cancer. *Lung Cancer* **41**: 87–198.
- Overington J.P., Al-Lazikani B. and Hopkins A.L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* **5**: 993–996.
- Ovnic M, Swank RT, Fletcher C, Zhen L, Novak EK, et al. (1991) Characterization and functional expression of a cDNA encoding egasyn (esterase-22): the endoplasmic reticulum-targeting protein of betaglucuronidase. *Genomics* **11**: 956–967.

- Park Joo-In, Tsai Sophia Y., Tsai Ming-Jer. (2003) Molecular mechanism of chicken ovalbumin upstream promoter-transcription factor (COUP-TF) actions. *Keio J Med* **52(3)**: 174-181.
- Potter PM, Wolverton JS, Morton CL, Wierdl M, Danks MK. (1998) Cellular localization domains of a rabbit and human carboxylesterase: influence on irinotecan (CPT-11) metabolism by the rabbit enzyme. *Cancer Research* **58**: 3627–3632.
- Redinbo MR, Potter PM. (2005) Mammalian carboxylesterases: from drug targets to protein therapeutics. *Drug Discov Today* **10**: 313–320.
- Robbi M, Beaufay H. (1983) Purification and characterization of various esterases from rat liver. *Eur J Biochem*. **137**: 293–301.
- Ruppert C, Bagheri A, Markart P, Schmidt R, Seegar W, et al. (2006) Liver carboxylesterase cleaves surfactant protein (SP-B) and promotes surfactant subtype conversion. *Biochem Biophys Res Commun.* **348**: 1449–1454.
- Sanghani SP, Quinney SK, Fredenberg TB, Davis WI, Murray DJ, et al. (2004) Hydrolysis of irinotecan and its oxidative metabolites, 7-ethyl-10-[4-N(5-aminopentanoic acid)-1-piperidino] carbonyloxycampothecin and 7-ethyl-10-[4-(1-piperidino)-1 amino]-carbonyloxycamptothecin, by human carboxylesterases CES1A1, CES2, and a newly expressed carboxylesterase isoenzyme, CES3. *Drug Metab Dispos.* **32**: 505–511.
- Sanghani SP, Sanghani PC, Schiel MA, Bosron WF. (2009) Human carboxylesterases: an update on CES1, CES2 and CES3. *Protein Pept Lett* **16**: 1207–1214.
- Satoh Tetsuo, Hosokawa Masakiyo (1998) Carboxylesterases: From molecules to functions. *Annu. Rev. Pharmacol. Toxicol.* **38**: 257-288.
- Satoh Tetsuo, Taylor Palmer, Bosron William F., Sanghani Sonal P., Hosokawa Masakiyo, La Du Bert n. (2002) Current progress on esterases: From molecular structure to function. *Drug Metabolism and Disposition* **30**: 488-493
- Satoh Tetsuo, Hosokawa Masakiyo (2006) Structure, function and regulation of carboxylesterases. *Chemico-Biological Interactions* **162**: 195-211.
- Schewer H, Langmann T, Daig R, Becker A, Aslandis C, et al. (1997) Molecular cloning and characterization of a novel putative carboxylesterase, present in human intestine and liver. *Biochem Biophys Res Commun*. **233**: 117–120.
- Shibita F, Takagi Y, Kitajima M, Kuroda T, Omura T. (1993) Molecular cloning and characterization of a human carboxylesterase gene. *Genomics* **17**: 76–82.
- Sladek F.M., Zhong W.M, Lai E., Darnell J.E. Jr. (1990) Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev.* **4(12B)**: 2353-2365.
- Sladek F.M., Ruse M.D. Jr, Nepomuceno L., Huang S.M., Stallcup M.R. (1999) Modulation of transcriptional activation and coactivator interaction by a

splicing variation in the F domain of nuclear receptor hepatocyte nuclear factor 4alpha1. *Mol Cell Biol.* **19(10)**: 6509-6522.

- Soutoglou E., Katrakili N., Talianidis I. (2000) Acetylation regulates transcription factor activity at multiple levels. *Mol Cell* **5(4)**: 745-751.
- Staudinger Jeff L., Xu Chenshu, Cui Yue J., Klaassen Curtis D. (2010) Nuclear receptor-mediated regulation of carboxylesterase expression and activity. *Expert Opin Drug Metab Toxicol.* **6**(3): 261-271.
- Sugai M., Umezu H., Yamamoto T., Jiang S., Iwanari H., Tanaka T., Hamakubo T., Kodama T., Naito M. (2008) Expression of hepatocyte nuclear factor 4 alpha in primary ovarian mucinous tumors. *Pathol Int.* **58(11)**: 681-686.
- Sun K., Montana V., Chllappa K., Brelivet Y., Moras D., Maeda Y., Parpura V., Paschal B.M., Sladek F.M. (2007) Phosphorylation of a conserved serine in the deoxyribonucleic acid binding domain of nuclear receptors alters intracellular localization. *Mol Endocrinol.* **21(6)**: 1297-1311.
- Tanaka T, Jiang S, Hotta H, Takano K, Iwanari H, Sumi K, Daigo K, Ohashi R, Sugai M, Ikegame C, Umezu H, Hirayama Y, Midorikawa Y, Hippo Y, Watanabe A, Uchiyama Y, Hasegawa G, Reid PC, Aburatani H, Hamakubo T, Sakai J, Naito M, Kodama T. (2006) Dysregulated expression of P1 and P2 promoter-driven hepatocyte nuclear factor-4α in the pathogenesis of human cancer. *J Pathol* **208**: 662-672.
- Tanimoto K, Kaneyasu M, Shimokuni T, Hiyama K, Nishiyama M. (2007) Human carboxylesterase 1A2 expressed from carboxylesterase 1A1 and 1A2 genes is a potent predictor of CPT-11 cytotoxicity in vitro. *Pharm Genomics* **17**: 1–10.
- Thierry-Mieg D, Thierry-Mieg J. (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol*. **7**(Suppl 1): S12–S14.
- Thomas H., Jaschkowitz K., Bulman M., Frayling T.M., Mitchell S.M., Roosen S., Lingott-Frieg .A, Tack C.J., Ellard S., Ryffel G.U., Hattersley A.T. (2001) A distant upstream promoter of the HNF-4alpha gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum Mol Genet* **10**: 2089-2097.
- Torres-Padilla M.E, Fougère-Deschatrette C., Weiss M.C. (2001) Expression of HNF4alpha isoforms in mouse liver development is regulated by sequential promoter usage and constitutive 3' end splicing. *Mech Dev* **109**: 183-193.
- Torres-Padilla M.E, Sladek F.M., Weiss M.C. (2002) Developmentally regulated N-terminal variants of the nuclear receptor hepatocyte nuclear factor 4alpha mediate multiple interactions through coactivator and corepressor-histone deacetylase complexes. *J Biol Chem.* **277(47)**: 44677-44687.
- Tsujita T, Okuda H. (1993) Palmitoyl-coenzyme A hydrolyzing activity in rat kidney and its relationship with carboxylesterase. *J Lipid Res.* **34**: 1773–1781.

- Umesono K. and Evans R.M. (1989) Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* **57**: 1139–1146.
- Umesono K., Murakami K.K., Thompson C.C., Evans R.M. (1991) Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell* **65**: 1255–1266.
- Von Heijne G. (1983) Patterns of amino acids near signal-sequence cleavage sites. *Eur J Biochem.* **133**: 17–21.
- Wang Huajin, Gilham Dean, Lehner Richard (2007) Proteomic and lipid characterization of apolipoprotein B-free luminal lipid droplets from mouse liver microsomes, implications for very low density lipoprotein assembly. *The Journal of Biological Chemistry* **282**: 33218-33226.
- Watt Alistair J., Garrison Wendy D., Duncan Stephen A. (2003) HNF4: A central regulator of hepatocyte differentiation and function. *Hepatology* **37(6)**: 1249-1253.
- Waxman D.J., Holloway M.G. (2009) Sex differences in the expression of hepatic drug metabolizing enzymes. *Mol Pharmacol.* **76(2)**: 215-228.
- Wiggins David, Gibbons Geoffrey F. (1992) The lipolysis/esterification cycle of hepatic triacylglycerol. Its role in the secretion of very-low-density l ipoprotein and its response to hormones and sulphonylureas. *Biochem. J.* **284**: 557-462.
- Williams ET, Wang H, Wrighton SA, Qian YW, Perkins EJ. (2010) Genomic analysis of the carboxylesterases: identification and classification of novel forms. *Mol Phylogenet Evol.* **57(1)**: 23–34.
- Xie W. (2008) Nuclear Receptors in Drug Metabolism. *Wiley, Hoboken*.
- Xie X., Liao H., Dang H., Pang W., Guan Y., Wang X., Shyy J.Y., Zhu Y., Sladek F.M. (2009) Down-regulation of hepatic HNF4alpha gene expression during hyperinsulinemia via SREBPs. *Mol Endocrinol.* **23(4)**: 434-443.
- Yan B, Matoney L, Yang D. (1999) Human carboxylesterases in term placenta: enzymatic characterization, molecular cloning and evidence for the existence of multiple forms. *Placenta* **20**: 517–525.
- Yuan X., Ta T.C., Lin M., Evans J.R., Dong Y., Bolotin E., Sherman M.A., Forman B.M., Sladek F.M. (2009) Identification of an endogenous ligand bound to a native orphan nuclear receptor. *PLoS One* **4(5)**: e5609.
- Zechel C., Shen X., Chen J., Chen Z., Chambon P., and Gronemeyer, H. (1994a) The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats. *EMBO J.* **13(6)**: 1425–1433.
- Zechel C., Shen X., Chambon P., and Gronemeyer H. (1994b) Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements. *EMBO J.* **13(6)**: 1414–1424.

- Zhang L, Hu Z, Zhu C, Liu Q, Zhou Y, et al. (2009) Identification and characterization of an epididymisspecific gene, Ces7. *Acta Biochim Biophys Sin.* **41**: 809–815.
- Zhang Youcai, Cheng Xingguo, Aleksunes Lauren, Klaasen Curtis D. (2012) Transcription factor-mediated regulation of carboxylesterase enzymes in livers of mice. *Drug Metabolism and Disposition* **40**: 1191-1197.
- Zhao Bin, Bie Jinghua, Wang Jing, Marqueen Stephanie A., Ghosh Shobha (2012) Identification of a novel intracellular cholesteryl ester hydrolase (carboxylesterase 3) in human macrophages: compensatory increase in its expression after carboxylesterase 1 silencing. *Am J Physiol Cell Physiol* **303**: C427-C435.
- Zhen L, Rusiniak ME, Swank RT. (1995) The beta-glucuronidase propeptide contains a serpin-related octamer necessary for complex formation with egasyn esterase and for retention within the endoplasmic reticulum. *J Biol Chem.* **270**: 11912–11920.

2.6. APPENDIX

2.6.1. *Mus musculus* Carboxylesterase 3 beta promoter sequence according to NCBI

```
-3500 TTATTGGCTAAACCTCCAGGCCTTTAAGTACCTCGTTAAAATGGAGATCTGTCTTGAGGC

-3440 TACATGACCTGTGGTCACATTTTCTACCTGTGGAGAGGATTGGGGCATTGTCCTTACATG

-3380 ACTGATGGCCACAAACCTATGGAGAGCTGTGTCCTCGTGTCAGGAGCCTGGTGTCCAGGA

-3320 GGGTGGCAACATGCCTACAATCCCTTGCAGGGTTATCCCCGGCAGGGTCTCCAGAGTTTC
        Ces3b-3259F
-3260 TAACTTAACCAGAAACCAGGCTGCCTTTCACAGTTCCAAATTGGGAGGTGGGGTTTGGGA

-3200 TGGGAGTTGGTAAGAGGGGGTGAACAGGAAATGGGAAATCATTTCAGATGTAAACAAATG

-3140 AATTGATAGAAATAAACAAACAAACAAACAAACAGTTCAAGTGTTTTGTTTGCCTGGAGC

-3080 GCTTTCAAAAAGCAAACAGTGTTAGCCTAAAAGTAGGCTATCATATATCAAAAATGTTAG

-3020 CTTAGGACTAGACATTCTTTATACTCTTTAAATTAGAATTACAGGGTTGATAGTAGGAAT

-2960 AAATACTGTTCTGTTTGCATTTTCTAATTGGGACTGAACTTGTAACTCTGGGCATATAAT
              Ces3b-2892F
-2900 AAAACGTGAAACACATAAACCGGACCGCTTGGAAAATCATTAATCTGTTCTGCAGAACTT

-2840 GTATACAGCAGTGCCTGGCTGCTAGACATTCAGGACCACAGAGTTCCTAGTAGCTGGGAA

-2780 ATCAGAGCTGTGAGATTACCCTGACCATTGGTCACTGACAATTGCCAAATGAAAGAGCAC

-2720 ACTGCAGATAAATATACACACCCTTGCTCAAATTCTCTTCACCCCAACCAAAGGTCATTC

-2660 CATTTCTTCATCTACATTCTAACTGTTCACACCCTAGGCTTTCAGCGTCTGCTTCCACCT

-2600 TCTTGTCTGGTTCCTCCAGTCACAGCTTCCTGCTCCCTGATCTTCATGCCCAACACCTCT

-2540 ATTGAACTGTGACTGCAAGCTCCATGATTGTAGTGACTCCTGTGGGAGTGTACAGGGCTC

-2480 CTGGCATCGTGAAGGAAGCTTCAGAGGCAGGTTTGCTCTGCTCCAAACTCCCACAGGAGC

-2420 TGCCTTTCTTGATCTCATCTAATCCTGCTGATTAGGTGGAAGGAAACTTCTAGAAGAAGG

-2360 TATAAGAAGAGGAATGAGACATGGTAACAGATGATCTTGGCTTTGCGGAAGAGAAGAGAG

-2300 AAAATGGAGCAAAGAACACCTTGGGGAGTGTGGTATGAGTGTGTTTGTGGTGCTAGGGAT

-2240 TGACCCTGAGGCCCCACCCACAAGAGACAAGCATATACCACTGAGCTACATCCCCAGATC

-2180 TCAACAGGATGCTTCTTAGATGAGGGTGGGTTTGCAGCAGGTCGGAAAGCACAGTAGAAG

-2120 TTGGAGGTGCAGAACAGATAGTCCAGGGAAGATAGGAAGGGACTTAGGGCTAGTATCCAT

-2060 GAATGGATTGCAGATGTAGGCTTCGATGAGGCAAACAAAAATGATAGTGTAAGAACTTCT
```

Ces3b-1956F

-2000 GTAAGAATTCAGTGAGAAACAAGACACTTGTTTACTGTCACTAACTCTGATTATTTGACC

-1940 GTGTGCAAGGGTGCATCTAAATTTTCTGTTTATACATGGAGGAGCCTGCCTGTAGAGCAT

-1880 TCAAGATGTGAAGGAAATACGAACATGGGTTTGAAGCAAGCCTGGGCCAATTGTGAGACC

-1820 TTCTATCGAAAAATATATGTGGGCACTGGAGAGAAATCATGGTGGTTAAGAGCACTGGCT

-1760 ACTCTTCCAAAGGACCTATACTTGATCGTCAGTACCTATATGGTAACTCACAACCCTCTT

-1700 TTACACCAGTGGCAAGGGATCCAATATCTTTTTCTGGACACTCAAGACACTGTACACACA

-1640 TGATACACAAACATACATGCAGCAAAACACCCTGAAAAGCATATACACAAAATAAAAACA

Ces3b-1956F

-1580 AATATTTATTTTATTTTATTGGGTAATTTCTTTACTTACATTTCAAATATTATCCACTAT

-1520 CTCCCCTGTCAGAAACCCCCTGTCCCATCCTCCCTCCCCCTGCACTTATGAGGGGTTTCC

-1460 TCCATCCACTTCCCCACAACTACCTCCCCGCCCTCGATTCCCCTAGACTGGGACATCTAT

-1400 TGAGCCTTCCTAGGACCAAGGACTTCTCCCATTTATGTCTATTTTTTTTAAAAAAGTGAA

-1340 AACCCATGCATACATATTCTGTTTAGGACAATTGGAAGGATGTTTTGTAGACATCTCAGA

-1280 ACACCTGAAGGATGCCCCTCCTATAATAATGGTCAAAAGACTATTCTCTGTTTTAAAGAC

-1220 CAGGCTGGGATTCCTTCTACAAGATCACCTCAAGATGCCAACATCCTAGTGACACACTGA

-1160 GCAGCTTCTGTGGTTGTGATCCAGCAGGCTTAGGCCACACTTCCACCCTCAGCATACCTT

Ces3b-1052F

-1100 CATTGTGGCATAAGGTACCAGGACTATGTGCATTCAGAAGGCAAATAAGGGGTCTTGGA

-1040 AGCTTGTTAGATTTTGACAGAAGTTTGCTGCAGGTTCAGAGTTGCTTCCGGGGGTCTCCC

-980 AATGGGATAATGCCTAAATGGAACCAGAATGGTTACTTGAGGACCTTATAAAGGTCGGGG

-920 AAACATCAAGATGCTTCCTAGGCTTGTAAAGTTGCAGACCTCTATATATAAGTCTATGAA

-860 AAATGTTAACATGTAGCCGCAATGCAGTAGGTCTTGGGACATCCCTTATTACAGAGTATG

-800 TGGATGCTGGGCTTAGAGACTTAGAGATGTCTCCTTGCCTCTGGGTTTCAGTCTTTCTTG

-740 GAATGATTGCTCTTTTCTGCTTGCCCAGTCCTCCTTCATAGAATAGGACCTTATCTTGAC

-680 TTGGGCCAGTCATTGGGCCATTTCTGCTCCATTTTTATCCCAGCACATCTCATAGCCAAG

-620 ACAAATTGTGGCTCAAAGATTTCATGGGTGAGTTGATATCCCCTTCCCTTCACTGGAAGT

Ces3b-532F

-560 CCTGCCTGGCTACAGGAGGTGGCCACTTCAGTCTCCATATCCCCAGGTGCTCAGAGTCTC

-500 AGCTAGGGTCATCTCCATAGACTCCCAGGAGCCTCCTCACCCATTCCCAGTTCTCCAGCA

-440 AGTCCCAGATTTGCCCGCCCCCATTTCCATTCTCACTCCCAGCCCTTTCCCAATTCCCAA

```
-380 ACCTGATCCCTATCCCTATTCCCCTCCTCAACTCTGTCCAACCCAGTTCCCTCGCTCCAT

-320 TCACTAAGCTCAGAGTCTGACTGACATGCAATTCAGCCACTGCAGTTGTAAGAGTGACAC

-260 ACCATGCCAGGCAAGAGTAGAATCATCCACAGAGGAGGGGTGGTGGTGGTGGGAGTGGGG
                                                    Ces3b-159F

-200 GTAGTGGGTGGGGTGAGGGTGTTAAATAGTCAGACCTCTCCAGAAGACTCTGCTGGTCCT

-140 GCTCTGAGAACAGGCAGAGCGTAGAATACACAATTCCAGAGCTGGCCTCCAACCAATCGG

     Ces3b-80F

 -80 GCAGACCAAGTTAATTTTCACAAGATAAGAGGTACTATAAAAGGCAAGTCCTCCCAGCTT
                Ces3b +4R

 -20 TCCACAGCTTCGTCCACAGTCCGTTATGACAAATATGAGGACAATGATACCAGCTGGGTC

 +41 CAGTGTCCTAGTCTGGGTGACCTGTCTGCTCCTGGCATTTGTTACCACAGTCACTGGTAA

+101 GACATACTCCCTGAACCAGTGAAGTGGGGCTTCCAGCTCCGTGGGAAACCTGAGGAAGGG
                Ces3b +186R

+161 GGGACTGGAGTGAGGCAGTAAGAGTGATCAGCCTTCTCTGA
```

2.6.2. Ces3b fragmentations plasmid maps

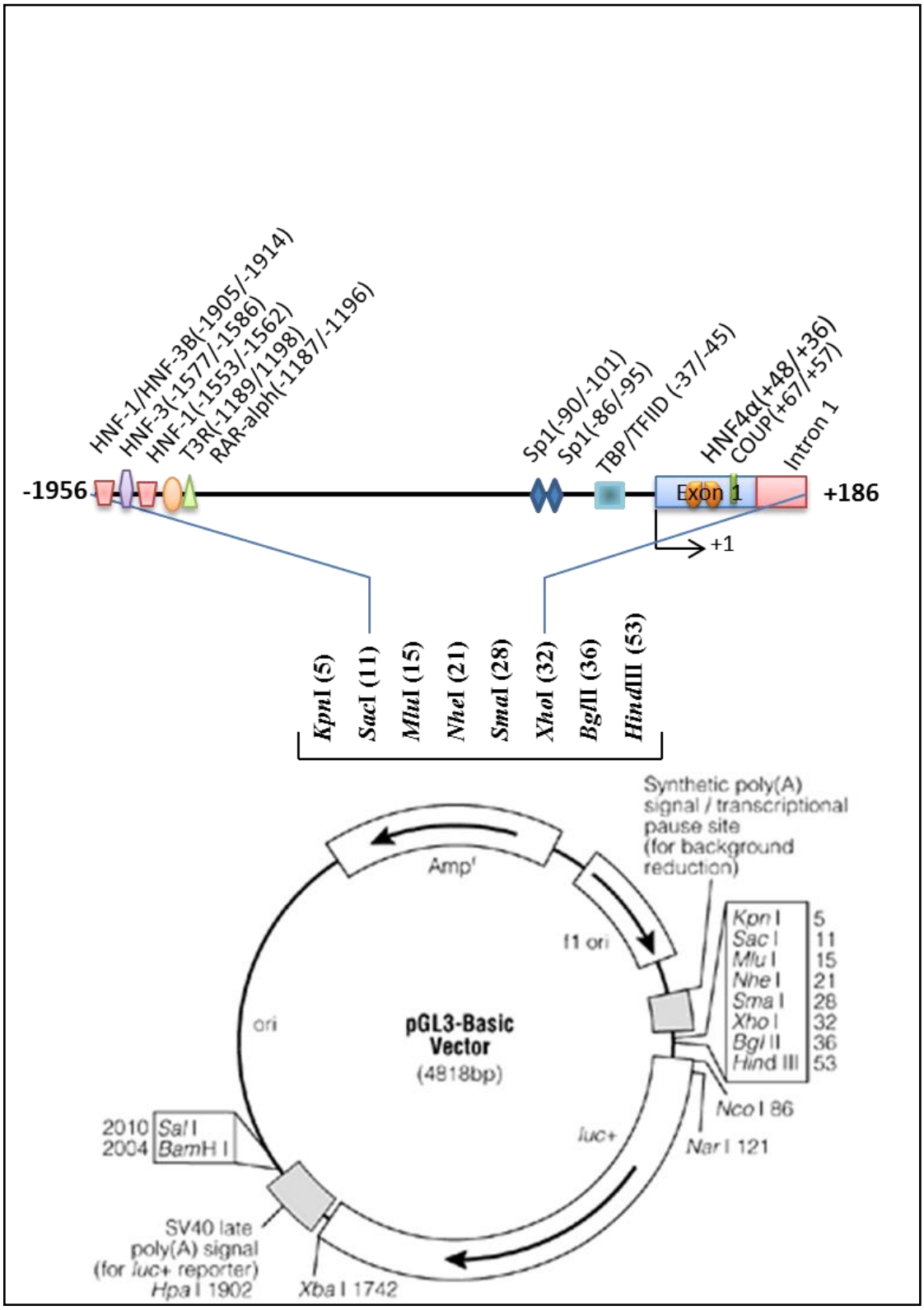

Figure 54. (-1956/+186) Ces3b-luc. Mouse Ces3b gene fragment -1956/+186 was inserted in pGL3 basic vector using *Sac*I and *Xho*I restriction enzymes. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

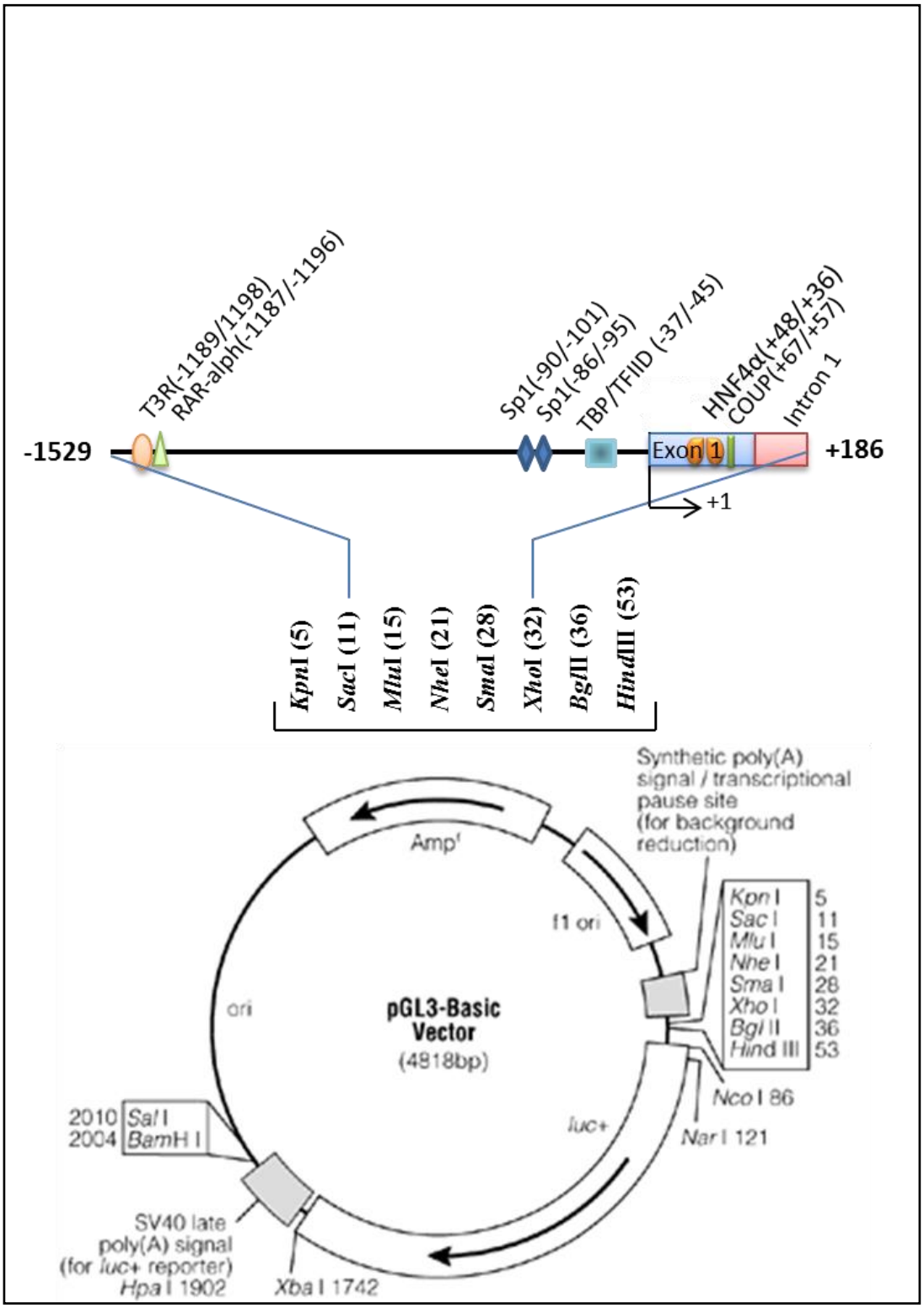

Figure 55. (-1529/+186) Ces3b-luc. Mouse Ces3b gene fragment -1529/+186 was inserted in pGL3 basic vector using *Sac*I and *Xho*I restriction enzymes. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

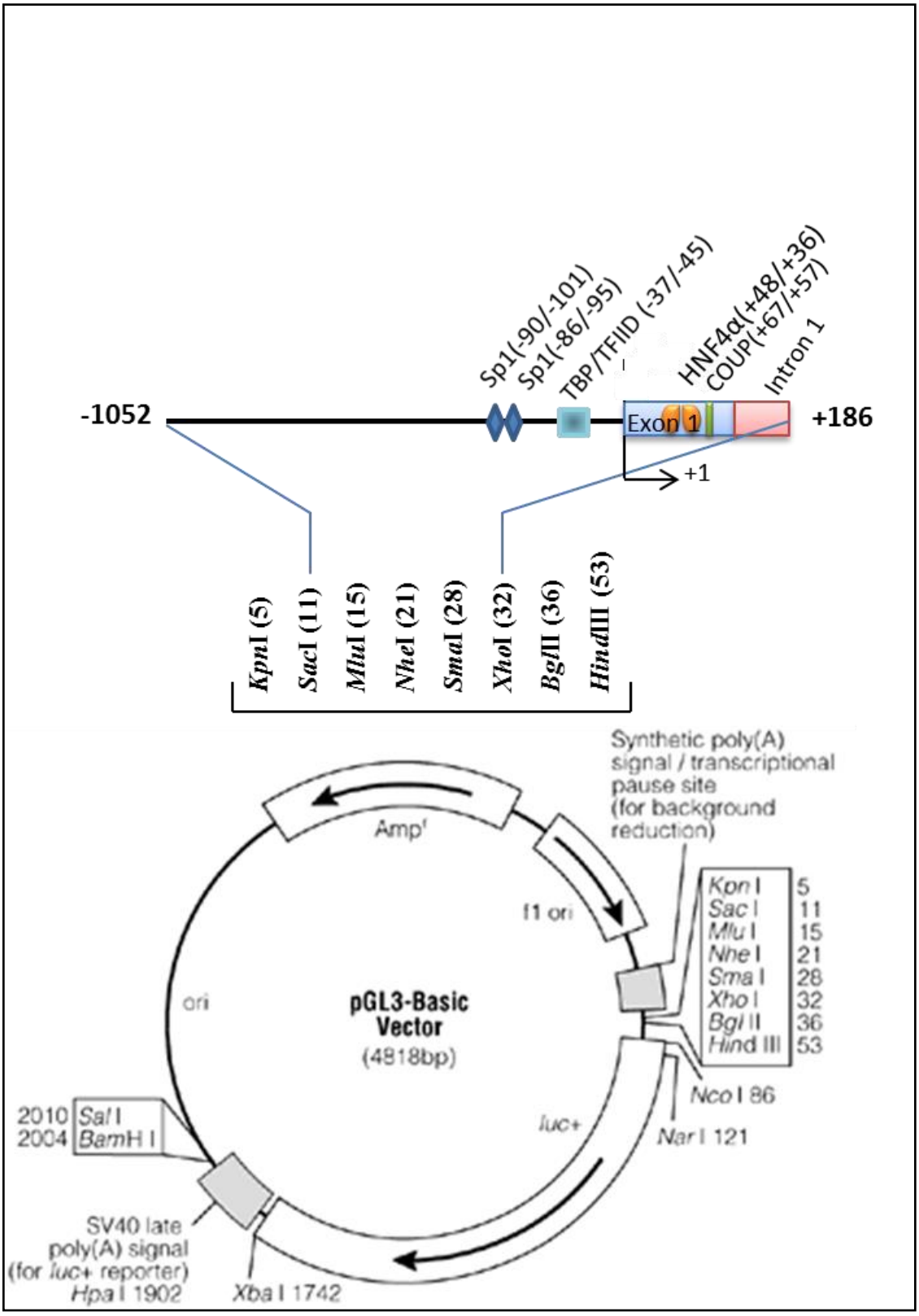

Figure 56. (-1052/+186) Ces3b-luc. Mouse Ces3b gene fragment -1052/+186 was inserted in pGL3 basic vector using *Sac*I and *Xho*I restriction enzymes. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

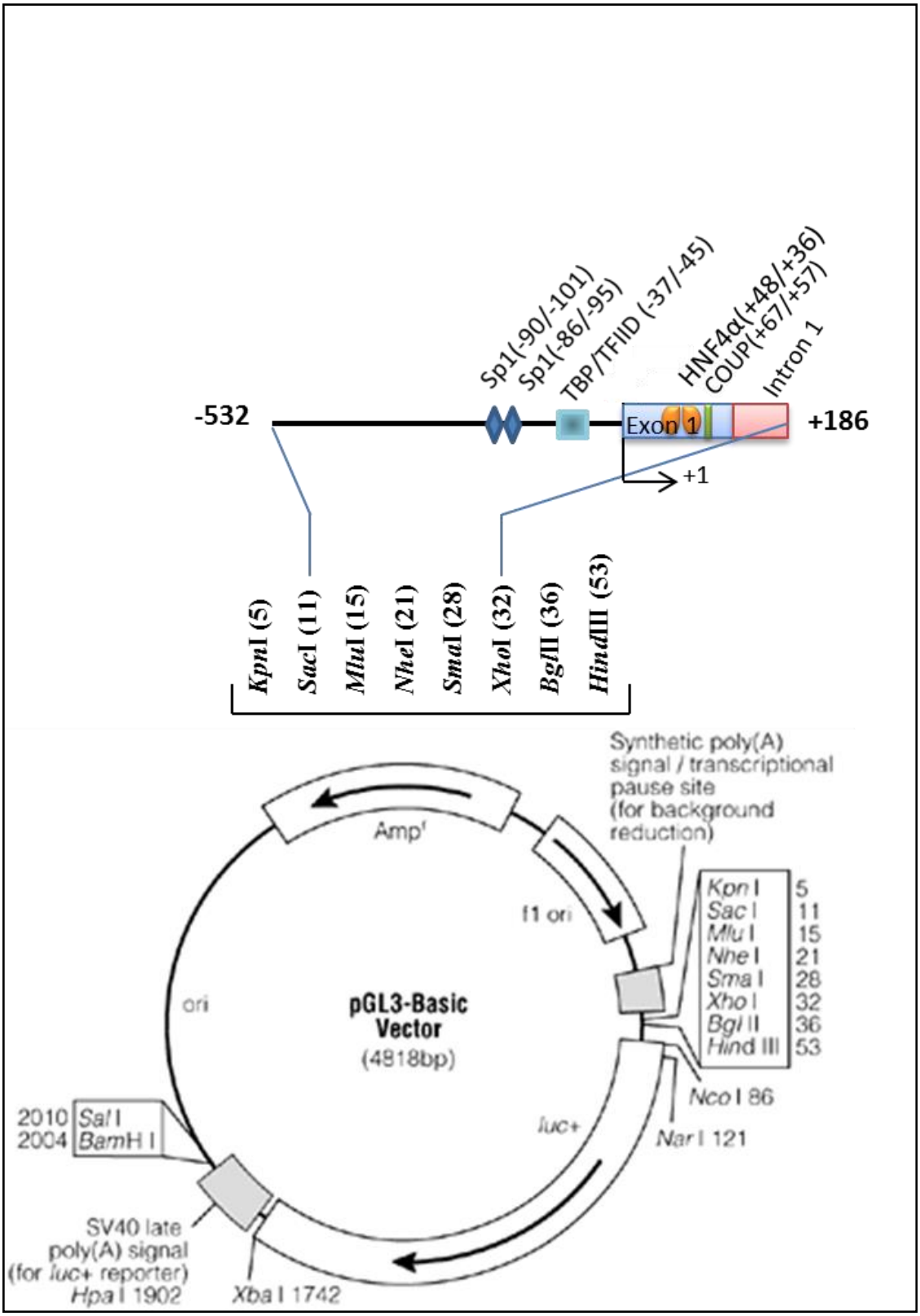

Figure 57. (-532/+186) Ces3b-luc. Mouse Ces3b gene fragment -532/+186 was inserted in pGL3 basic vector using *Sac*I and *Xho*I restriction enzymes. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

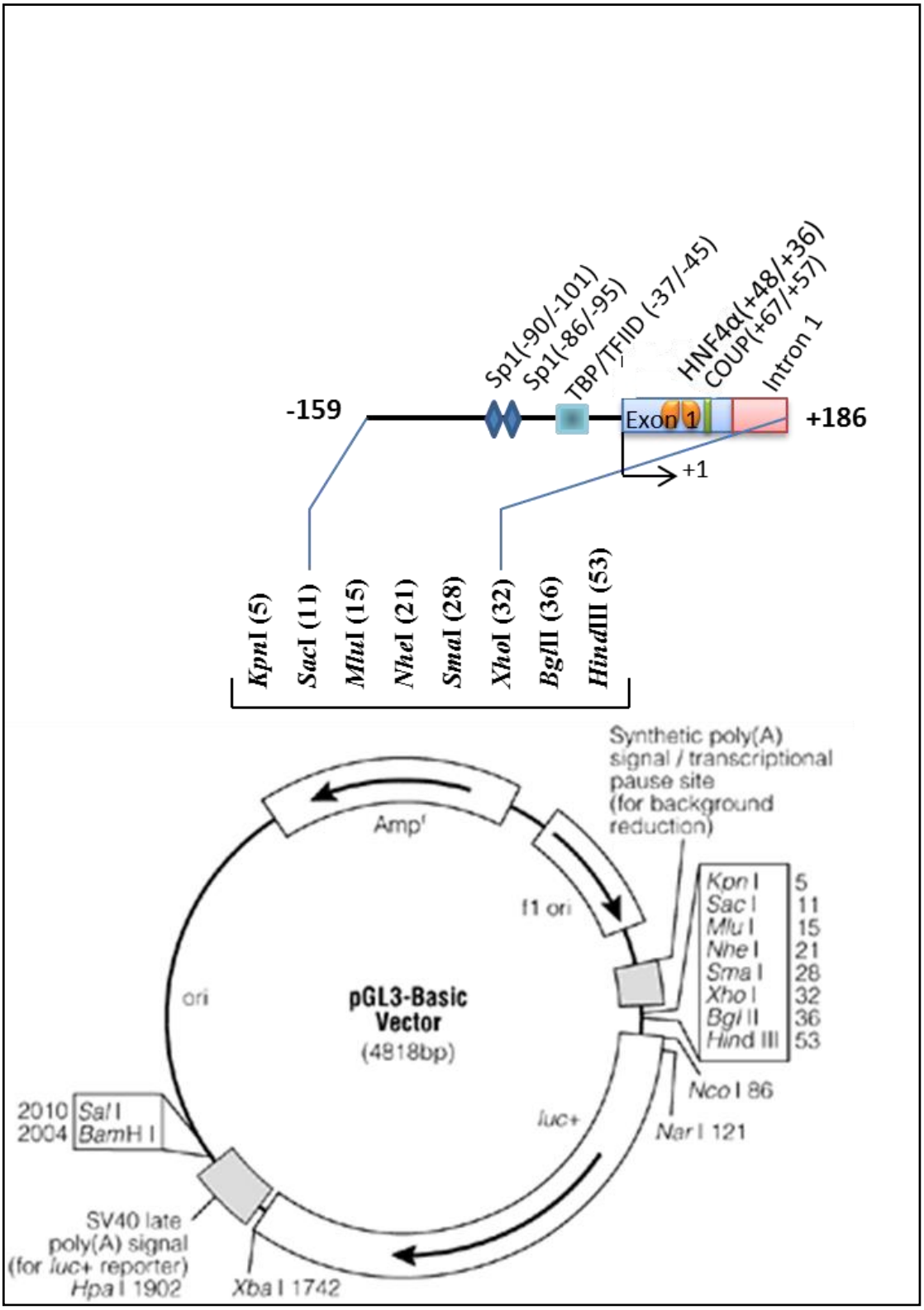

Figure 58. (-159/+186) Ces3b-luc. Mouse Ces3b gene fragment -159/+186 was inserted in pGL3 basic vector using *Sac*I and *Xho*I restriction enzymes. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

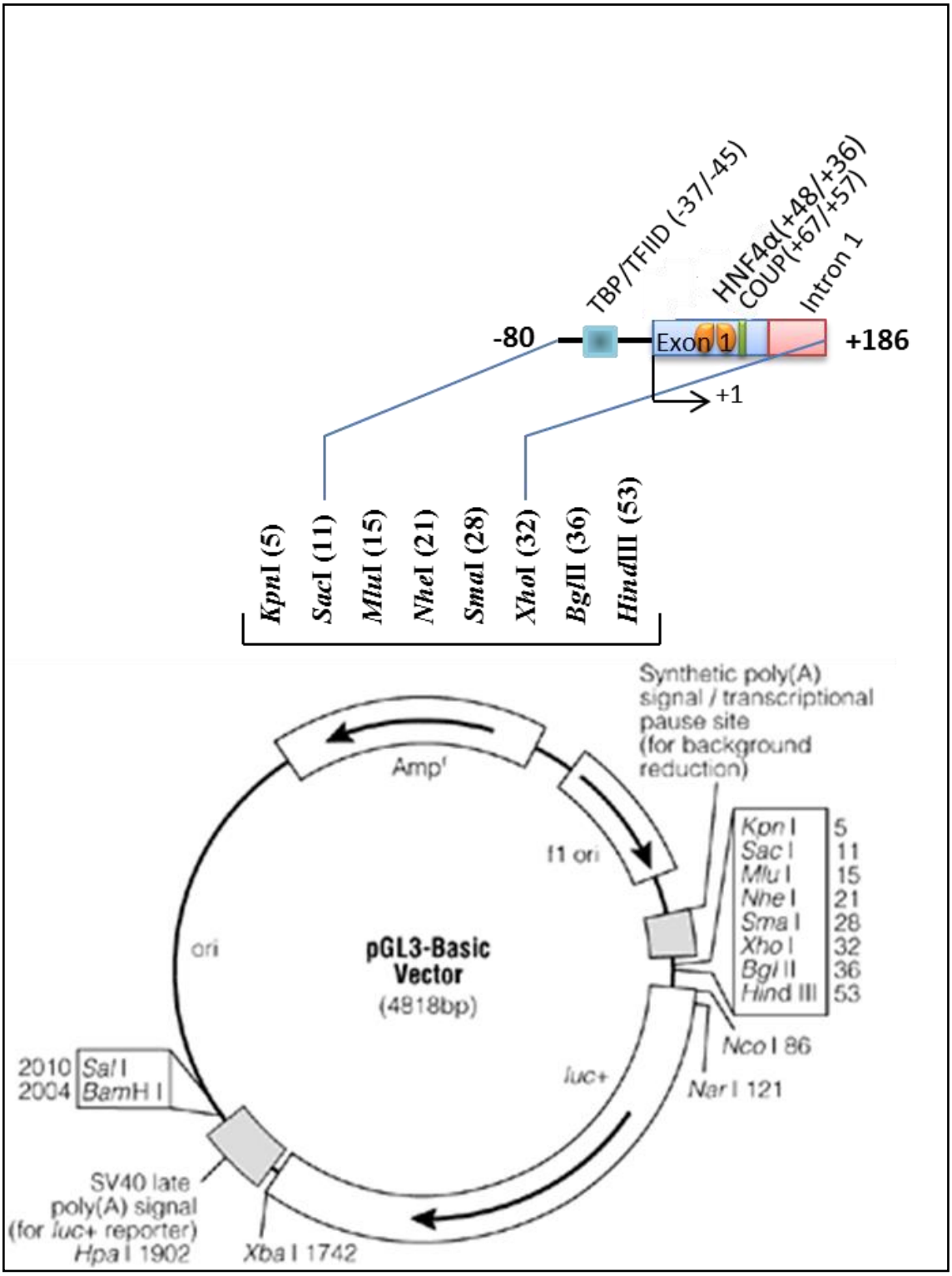

Figure 59. (-80/+186) Ces3b-luc. Mouse Ces3b gene fragment -80/+186 was inserted in pGL3 basic vector using *Sac*I and *Xho*I restriction enzymes. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).

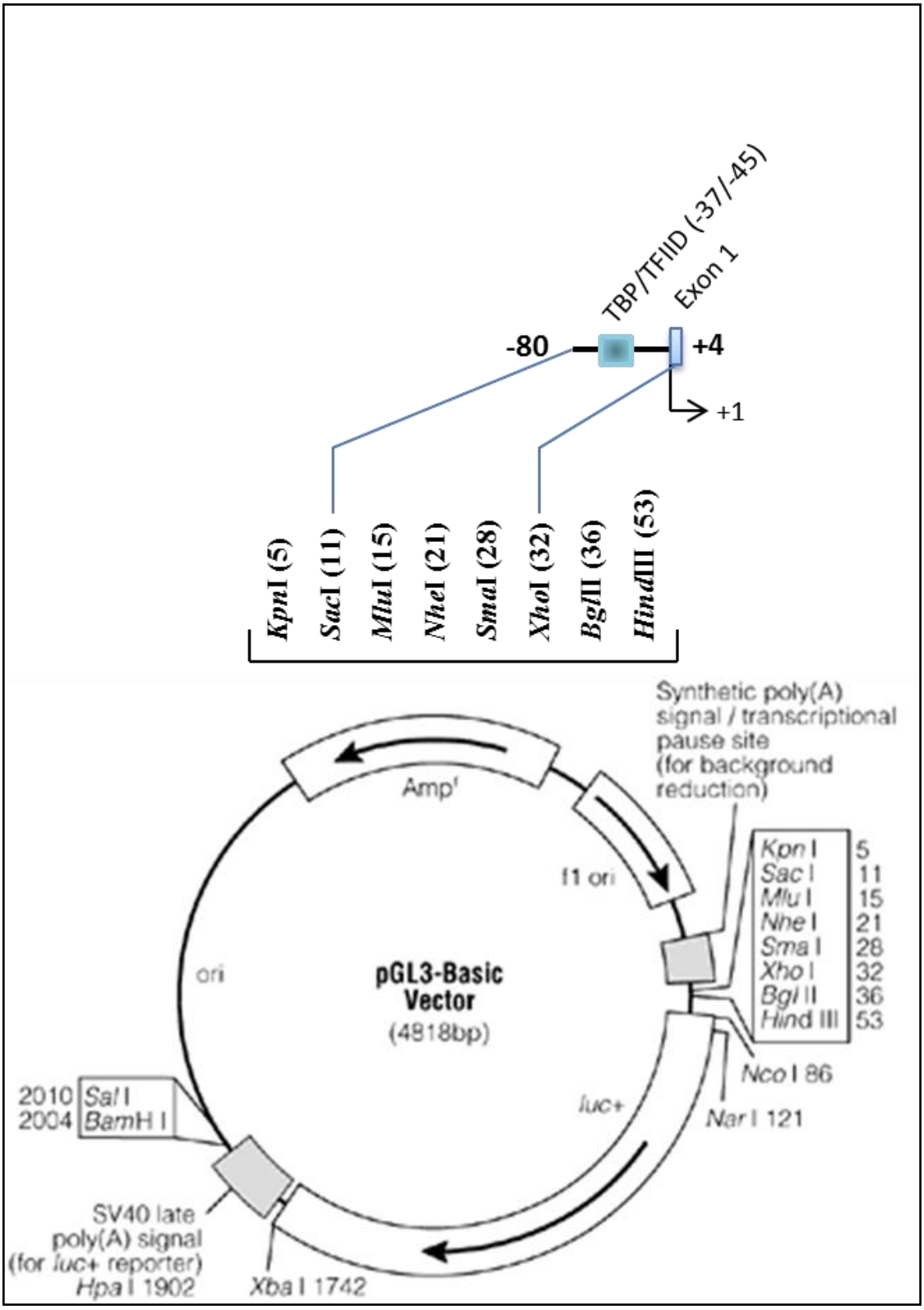

Figure 60. (-80/+4) Ces3b-luc. Mouse Ces3b gene fragment -80/+4 was inserted in pGL3 basic vector using *SacI* and *XhoI* restriction enzymes. Potential transcription factor binding sites are shown, according to Gene-Regulation (http://www.gene-regulation.com/pub/programs/alibaba2/index.html).