Univerzita Karlova v Praze

1. lékařská fakulta

Charles University in Prague

1ˢᵗ Faculty of Medicine

Doktorský studijní program:

Biochemie a patobiochemie

Ph.D. program:

Biochemistry and pathobiochemistry

Mgr. Ondřej Lukšan

MECHANISMY REGULACE EXPRESE GENŮ PRO ORNITIN TRANSKARBAMYLÁZU A BETA-GLUKOCEREBROSIDÁZU A JEJICH VÝZNAM V DIAGNOSTICE

REGULATORY MECHANISMS OF ORNITHIN TRANSCARBAMYLASE AND BETA-GLUCOCEREBROSIDASE GENE EXPRESSION AND THEIR RELEVANCE TO DIAGNOSTICS

Disertační práce / Ph.D. thesis

Vedoucí závěrečné práce / Supervisor: Doc. MUDr. Milan Jirsa, CSc.

Konzultant / Consultant: RNDr. Lenka Dvořáková, CSc.

Praha, 2014

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem řádně uvedl a citoval všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu.

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 14.5.2014

Ondřej Lukšan

....

Doctoral Study Programmes in Biomedicine

Charles University in Prague and Academy of Science of the Czech Republic

PhD program: Biochemistry and pathobiochemistry

Program committee chairman: Prof. MUDr. Stanislav Štípek, DrSc.

Supervising department:

Institute for Clinical and Experimental Medicine, Center for Experimental Medicine, Laboratory of Experimental Hepatology, Prague, Czech Republic

Supervisor: doc. MUDr. et Mgr. Milan Jirsa, CSc.

Institute for Clinical and Experimental Medicine, Prague, Czech Republic

Consultant: RNDr. Lenka Dvořáková, CSc.

Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Czech Republic

Submission of PhD thesis: May, 2014.

The thesis is available at the dean's office of the First Faculty of Medicine, Charles University in Prague

Identifikační záznam:

LUKŠAN, Ondřej. *Mechanismy regulace exprese genů pro ornitin transkarbamylázu a beta-glukocerebrosidázu a jejich význam v diagnostice. [Regulatory mechanisms of ornithin transcarbamylase and beta-glucocerebrosidase gene expression and their relevance to diagnostics]*. Praha, 2014. 115 s., Disertační práce. Univerzita Karlova v Praze, 1. lékařská fakulta, Laboratoř experimentální hepatologie, IKEM. Vedoucí práce: Milan Jirsa

Abstrakt

Konečná diagnóza dědičných metabolických poruch je obvykle určena na základě vyšetření enzymové aktivity (což je mnohdy komplikované) a/nebo za pomoci molekulárně genetického vyšetření. Výsledky standardní mutační analýzy však mohou být někdy zavádějící, zejména v případě rozsáhlých reorganizací na úrovni chromozomu či u mutací v regulačních oblastech. Předkládaná práce je zaměřena na charakterizaci komplexních mutací v genu pro ornitin transkarbamylázu (*OTC*), studium regulačních oblastí *OTC* a genu pro β-galaktozidázu (*GBA*).

V rámci studie provedené na souboru čítajícím 37 pacientů s kongenitální hyperamonémií II. typu (OTCD) bylo identifikováno a charakterizováno 14 nových mutací včetně tří rozsáhlých delecí. Byl zaznamenán i případ symptomatické heterozygotky pro hypomorfní mutaci p.R129H, u níž byl fenotyp OTCD zřejmě podmíněn posunem v X-inaktivaci ve prospěch mutované alely. Nález byl podložen sledováním nevýhodné X-inaktivace ve třech různých periferních tkáních.

Za účelem prověření patogenity promotorové variace c.-366A>G v *OTC* nalezené v jedné rodině s mírným fenotypem OTCD byly identifikovány tři alternativní počátky transkripce (PT) a vymezen promotor lidského *OTC*. Transkripční aktivita promotoru, stejně tak jako aktivita nově nalezeného distálního enhanceru, byla ověřena funkční esejí. Výsledky ukázaly klíčový význam interakce promotoru s enhancerem pro tkáňově specifickou expresi *OTC* v játrech. Přítomnost variace c.-366A>G vedla k padesátiprocentním snížení aktivity mutovaného promotoru v interakci s enhancerem.

Detailní charakterizace promotoru lidského *OTC* odhalila dva silné *cis*-aktivní regulační elementy odpovídající vazebným místům pro HNF-4α. V oblasti proximálního promotoru byl nalezen třetí vysoce konzervovaný motiv rozpoznávaný HNF-4α, přičemž všechna tři vazebná místa se nalézají ve vzdálenosti do 35 bp nad PT. Promotor *OTC* přitom postrádá základní elementy nezbytné pro iniciaci transkripce v konvenčních pozicích. Dosažené výsledky silně podporují úlohu HNF-4α v iniciaci transkripce *OTC* u člověka.

Metodický postup zavedený při studiu *OTC* byl následně aplikován na charakterizaci alternativního promotoru *GBA*. V případě alternativního transkriptu nesoucího dva nepřekládané exony -1 a -2 byly nalezeny tři různé PT. Při porovnání s normálním promotorem vykazoval alternativní promotor nižší transkripční aktivitu, expresní profil napříč několika různými tkáněmi však byl srovnatelný. Byla vyslovena hypotéza, že výrazné rozdíly ve fenotypu u pacientů se stejným genotypem *GBA* mohou být dány variabilní expresí mutovaného alternativního transkriptu. Tuto hypotézu se však nepodařilo experimentálně potvrdit na souboru 20 pacientů s Gaucherovou chorobou.

Výsledky výše uvedených studií rozšiřují možností molekulární diagnostiky OTCD a Gaucherovy choroby.

Klíčová slova: Ornitin transkarbamyláza, kongenitální hyperamonémie II. typu, kyselá β-glukocerebrosidáza, Gaucherova choroba, genová exprese, regulace transkripce, promotor, enhancer

Abstract

Definitive diagnosis of inherited metabolic disorders commonly depends on the measurement of enzyme activity (which is often complicated) and/or molecular genetic testing. Yet even the standard mutation analysis can bring false negative results in the case of gross chromosomal rearrangements or incorrect regulation of gene expression due to the mutations in regulatory regions. In the present study I focused on characterization of complex mutations affecting the gene encoding ornithin transcarbamylase (OTC) followed by studies of regulatory regions of *OTC* and *GBA* (the gene encoding β-glucocerebrosidase).

In the first study we identified 14 novel mutations including three large deletions in a cohort of 37 patients with OTC deficiency (OTCD). Subsequently we evaluated clinical significance of all these mutations. We also found a heterozygote carrying a hypomorphic mutation and manifesting OTCD most likely due to unfavorable X-inactivation which was observed independently in three different peripheral tissues.

In order to evaluate the clinical significance of a promoter variation c.-366A>G found in a family with mild OTCD we identified three alternative transcription start sites (TSSs) of human *OTC* and delimited the promoter. We also found a distal enhancer and performed functional analysis of both regulatory regions. Our results indicate that tissue specific expression of *OTC* in the liver depends on the promoter-enhancer interaction. The variation c.-366A>G decreased the promoter-enhancer transcriptional activity by 50%.

A detailed characterization of human *OTC* promoter revealed two positive *cis*-acting regulatory elements corresponding to HNF-4α binding sites. Both sites, similarly as a third HNF-4α recognition motif found in the proximal promoter, are located within 35 bases upstream of the TSSs. Since the *OTC* promoter lacks general core promoter elements such as TATA-box or initiators on standard positions, our results strongly suggest an important role of HNF-4α in the control of *OTC* transcription in human.

A similar approach as in the *OTC* gene studies was used in studies of an upstream promoter of *GBA*. We identified three alternative TSSs and performed function analysis of the alternative promoter. Its transcriptional activity was lower than that of the normal promoter while expression profiles across multiple tissues were comparable. We hypothesized that phenotypic differences in patients with the same genotype may be caused by variable expression of mutant *GBA*; however, our hypothesis was not confirmed experimentally in a group of twenty Gaucher patients.

In conclusion, our findings extend the possibilities of molecular genetic testing for OTCD and Gaucher disease.

Table of contents

Abbreviations

ACTB	β-actin
AMP, ADP, ATP	adenosin mono-, di- and triphosphate
ARG1	arginase 1
ARG1D	arginase 1 deficiency
ASL	argininosuccinate lyase
ASLD	argininosuccinate lyase deficiency
ASS	argininosuccinate synthetase
ASSD	argininosuccinate synthetase deficiency
ATG	initiation codon
BRE	TFIIB recognition element
C/EBP	CCAAT-enhancer-binding protein
Cis1+, Cis2+	positive *cis*-active regulaytory elements 1 and 2
CLEAR	coordinated lysosomal expression and regulation motif
CNS	central nervous system
COUP-TF	chicken ovalbumin upstream promoter-transcription factor
CPS1	carbamoyl phosphate synthetase 1
CPS1D	carbamoyl phosphate synthetase 1 deficiency
DPE	downstream core promoter element
eIF	eukaryotic initiation factor
EMSA	electromobility shift assay
ER	endoplasmic reticulum
EST	expressed sequence tag
GalCer	galactosylceramide
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GBA	acid β-glucocerebrosidase
GD	Gaucher's disease
GDH	glutamate dehydrogenase
GlcCer	glucosylceramide
GSL	glycosphingolipid
HHH	hyperornithinemia-hyperammonemia-homocitrullinuria
HNF-4	hepatocyte nuclear factor 4
hsc	heat shock cognate protein

|	
Inr	initiator element
LSD	lysosomal storage disease
MLPA	multiplex ligation probe amplification assay
MSF	mitochondrial stimulating factor
MTE	motif ten element
NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
NAG	N-acetylglutamate
NAGS	N-acetylglutamate synthase
NHEJ	non-homologous end joining
ORNT1	mitochondrial ornithin / citrullin antiporter
OTC	ornithin transcarbamylase
OTCD	ornithine transcarbamylase deficiency
P1	acid β-glucocerebrosidase proximal promoter
P2	acid β-glucocerebrosidase alternative upstream promoter
PBF	presequence binding factor
PIC	pre-initiation complex
PolII	RNA polymerase II
PP$_i$	pyrophosphate
ps*GBA*	acid β-glucocerebrosidase pseudogene
RLM-RACE	RNA ligase-mediated rapid amplification of cDNA ends
RNA, DNA	ribonucleic and deoxyribonucleic acid
Sp1	specificity protein 1
TAF	TATA-box-binding protein associated factor
TBP	TATA-box-binding protein
TF	transcription factors
TFII	general transcription factor of RNA polymerase II
TFEB	transcription factor EB
TIS	translation initiation site
TSS	transcription start site
UCD	urea cycle disorder
UTR	untranslated region

1. Introduction

1.1. OTC deficiency

1.1.1. Nitrogen metabolism and urea cycle

Nitrogen is an essential component of many physiologically active compounds, namely amino acids and nucleic acids which represent the basic structural unit of proteins and DNA. Nitrogen intake and metabolism is essential to the function and structure of a large variety of molecules and for survival of all living organisms. Despite its abundance in the atmosphere, nitrogen is relatively inert and has to be fixed by nitrifying bacteria and converted into more reactive forms such as nitrates or ammonia in order to be utilizable by other organisms. Reduced nitrogen in the form of ammonium ions (NH_4^+) can then be used by bacteria and plants to build up more complex nitrogen containing compounds. These products are later used by higher animals thus making a part of nitrogen cycle. In vertebrates, the main sources of reduced nitrogen are dietary free amino acids, protein, and the ammonia produced by intestinal tract bacteria. Metabolism of dietary protein begins in the stomach under action of pepsin A and continues in the duodenum, with the aid of pancreatic proteases. The amino acids and peptides produced are taken up by the mucosal wall enterocytes of the intestines and send to circulation. Unlike carbohydrates or fat, there is no mechanism how to store large amounts of nitrogen in the body. In order to maintain homeostasis of essential proteins the ratio of ingested and excreted nitrogen must be kept in balance. A certain part of free amino acids is constantly stored in the „nitrogen pool", which enables to buffer dynamic requirements of *de novo* proteosynthesis. Without this option, deficiency of any single essential amino acid can inhibit translation.

Ammonia is produced by all tissues during the metabolism of a variety of compounds, e.g. amino acids, biological amines, purines and pyrimidines. The excretion of waste nitrogen can occur through different routes. Most of nitrogen is excreted via urine in the form of urea (5-20 g/day), ammonia (0.4-1.5 g/day), creatinine (0.6-1.8 g/day) and uric acid (0.2-1.0 g/day) (Gropper SAS, et al., 2009). Additional minor routes comprise faeces, other body secrets (such as sweat and saliva), skin, hair and nails. While the level of blood ammonia must be kept very low, because even slightly elevated concentrations (hyperammonemia) are toxic to the central nervous system, it has to be transformed into a neutral molecule. Quantitatively the most important disposal route for ammonia is the conversion to urea. Once synthesized in the liver, urea travels in the blood to the kidneys, where it passes into the glomerular filtrate and is excreted in the urine. The transfer of amino acid nitrogen from peripheral tissues to the liver is

mediated by "transporter" molecules, namely by glutamine or alanine. While glutamine is the principal recipient of amino group in transaminations, alanine acts as an important nitrogen-carrier within the glucose-alanine cycle thus making a part of energy metabolism in skeletal muscles.

In order to keep the overall nitrogen balance, free amino acids, the intermediates of dietary protein metabolism, should be further processed. The first reaction in the breakdown of an amino acid is almost always removal of its α-amino group with the object of excreting excess nitrogen and degrading the remaining carbon skeleton or converting it to glucose. The dominant reactions involved in removing amino acid nitrogen from the body are known as transaminations. During transamination reactions the amino group of an amino acid is transferred to a α-keto acid to yield the α-keto acid of the original amino acid and a new amino acid:

$$\text{Amino acid} + \text{α-ketoglutarate} \rightleftharpoons \text{α-keto acid} + \text{glutamate}$$
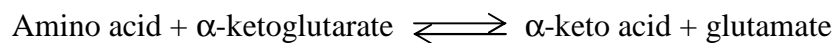
The reactions are catalyzed by aminotransferases (or transaminases) acting with participation of pyridoxal-5'-phosphate as an aldehyde-containing cofactor.

Aminotransferases, which exist for all amino acids except threonine and lysine, differ in their specificity for amino acid substrates in the first stage of the transamination reaction, thereby producing the correspondingly different α-keto acid products. Most aminotransferases, however, accept only α-ketoglutarate or (to a lesser extent) oxaloacetate as the α-keto acid substrate in the second step of the reaction, thereby yielding glutamate or aspartate as their only amino acid products. The amino groups of most amino acids are consequently funneled into the formation of glutamate or aspartate, which are themselves interconverted by glutamate–aspartate aminotransferase (Voet D and Voet JG, 1995). Both, glutamate and aspartate, are the final amino group donors in the synthesis of urea. Because of the participation of α-ketoglutarate in numerous transaminations, glutamate is furthermore considered as a prominent intermediate in nitrogen elimination as well as in anabolic pathways.

While aspartate enters the urea cycle directly, the amino group of glutamate needs to be released by oxidative deamination. The reaction producing a-ketoglutarate and ammonia is catalysed by glutamate dehydrogenase (GDH), the only known enzyme that, in at least some organisms, can accept either NAD^+ or $NADP^+$ as its redox coenzyme. GDH is allosterically inhibited by GTP, NADH and nonpolar compounds such as palmitoyl-CoA and steroid hormones. It is activated by ADP, NAD^+, and leucine, thus contributing to the regulation of the urea cycle (Voet D and Voet JG, 1995). Glutamate can be also alternatively converted into

glutamine by the action of glutamine synthase and transported to kidney tubule cells. There the glutamine is sequentially deamidated by glutaminase and deaminated by kidney glutamate dehydrogenase. The ammonia produced within these reactions is excreted as NH_4^+ directly in the urine, where it plays a role of an acido-basic regulator essential to maintain pH in physiologic range.

As mentioned earlier, in ureotelic organisms, about 80% of nitrogen is excreted in the form of urea which is produced exclusively in the liver following a series of reactions that are distributed between the mitochondrial matrix and the cytosol. The mechanism of urea formation from NH_3 and CO_2 was proposed in 1932 by Krebs and Henseleit to be a cyclic process wherein ornithine, citrulline and arginine acted as carrier compounds (Krebs HA and Henseleit K, 1932). The excess of ammonia is channeled to the mitochondria of liver cells and converted into carbamoyl phosphate. The reaction of HCO_3^- and NH_3 (carrying the first atom of waste nitrogen) is catalyzed by carbamoyl phosphate synthetase 1 (CPS1) and requires co-operation of N-acetyl glutamate (NAG) cofactor. Energy for the reaction is supplied by the hydrolysis of two molecules ATP.

Still in the mitochondrial matrix, carbamoyl phosphate enters the urea cycle. Ornithine transcarbamylase (OTC) transfers the carbamoyl group to ornithine, producing citrulline. The energy for the reaction is provided by the high-energy anhydride of carbamoyl phosphate. Ornithine arising in the cytosol, as well as citrulline synthesized in the mitochondrion, are transported through mitochondrial membrane by the action of ornithine/citrulline antiporter (Bradford N and McGivan J, 1980) encoded by the *ORNT1* gene. ORNT1 is a member of the solute carrier family of transporters and as such is also identified as *SLC25A15* (Camacho JA, et al., 1999).

Once present in the cytosol, citrulline acquires a second atom of waste nitrogen by condensation of citrulline's ureido group with an aspartate amino group. The reaction is catalyzed by argininosuccinate synthetase (ASS). The ureido oxygen atom is activated as a leaving group through formation of a citrullyl–AMP intermediate, which is subsequently displaced by the aspartate amino group. Therefore the reaction requires hydrolysis of one molecule ATP. With formation of argininosuccinate, all of the urea molecule components have been assembled.

The subsequent elimination of fumarate from the argininosuccinate under action of argininosuccinate lyase (ASL) leads to the production of arginine, the urea's immediate precursor. Fumarate is then converted via the fumarase and malate dehydrogenase reactions to oxaloacetate, which is used in gluconeogenesis.

The final reaction of the urea cycle is the hydrolysis of arginine to yield urea and regenerate ornithine. The reaction is catalyzed by arginase 1 (ARG1). Ornithine is then returned to the mitochondrion for another round of the cycle. The urea cycle thereby converts two amino groups, one from NH_3 and one from aspartate, and a carbon atom from bicarbonate to the relatively nontoxic excretion product urea (Voet D and Voet JG, 1995).

The total cost of energy equivalent to 3 molecules ATP and a PP_i consumed in the production of urea is more than recovered by the release of energy formed during the synthesis of the urea cycle intermediates. Ammonia released during the glutamate dehydrogenase reaction is coupled to the formation of NADH. In addition, when fumarate is converted back to aspartate, the malate dehydrogenase reaction used to convert malate to oxaloacetate generates a mole of NADH. These two moles of NADH, thus, are oxidized in the mitochondria yielding 6 moles of ATP.

The capacity of the liver for ureagenesis is sufficiently in excess of usual rates of urea production (Beliveau CG, et al., 1993) so that it can immediately respond to acute increases in waste nitrogen load. In short-term, regulation of urea synthesis is mediated by three factors: substrate availability (Krebs HA, et al., 1973; Cohen NS, et al., 1987), changes in the activity of CPS1 caused by changes in the concentration of its allosteric activator NAG (Shigesada K and Tatibana M, 1971; Meijer AJ and Verhoeven AJ, 1984) and the acid-base status (Bean ES and Atkinson DE, 1984). Between these factors, substrate availability is probably the most important determinant of short-term changes in rates of ureagenesis (Beliveau CG, et al., 1993). The regulation CPS1 via availability of NAG is important even in this aspect, as carbamoyl phosphate plays a role of key substrate at the entry of the urea cycle. The levels of NAG are controlled not only by concentration of its precursors acetyl-CoA and glutamate, but also by the activity of the enzyme catalyzing its synthesis, *N*-acetylglutamate synthase (NAGS). This additional level of control is reciprocally linked to the urea cycle through arginine, the final intermediate of urea synthesis, which acts as a positive allosteric effector of NAGS (Sonoda T and Tatibana M, 1983).

Longer-term changes in activity of the urea cycle enzymes occur in response to alterations in amino acid nitrogen flux resulting from changes in dietary protein intake or catabolism of endogenous protein. Glucagon, insulin, and glucocorticoids play major roles in mediating changes in activity of the urea cycle enzymes under these conditions and altered rates of transcription represent a major portion of changes in expression of the urea cycle enzymes (Morris SM, 2002).

Defects in the urea cycle result in the incapability of wasting nitrogen from the breakdown of protein and other nitrogen-containing molecules from the organism. Reduced wasting capacity of nitrogen commonly leads to hyperammonemia, a metabolic disturbance characterized by an excess of ammonia in the blood.
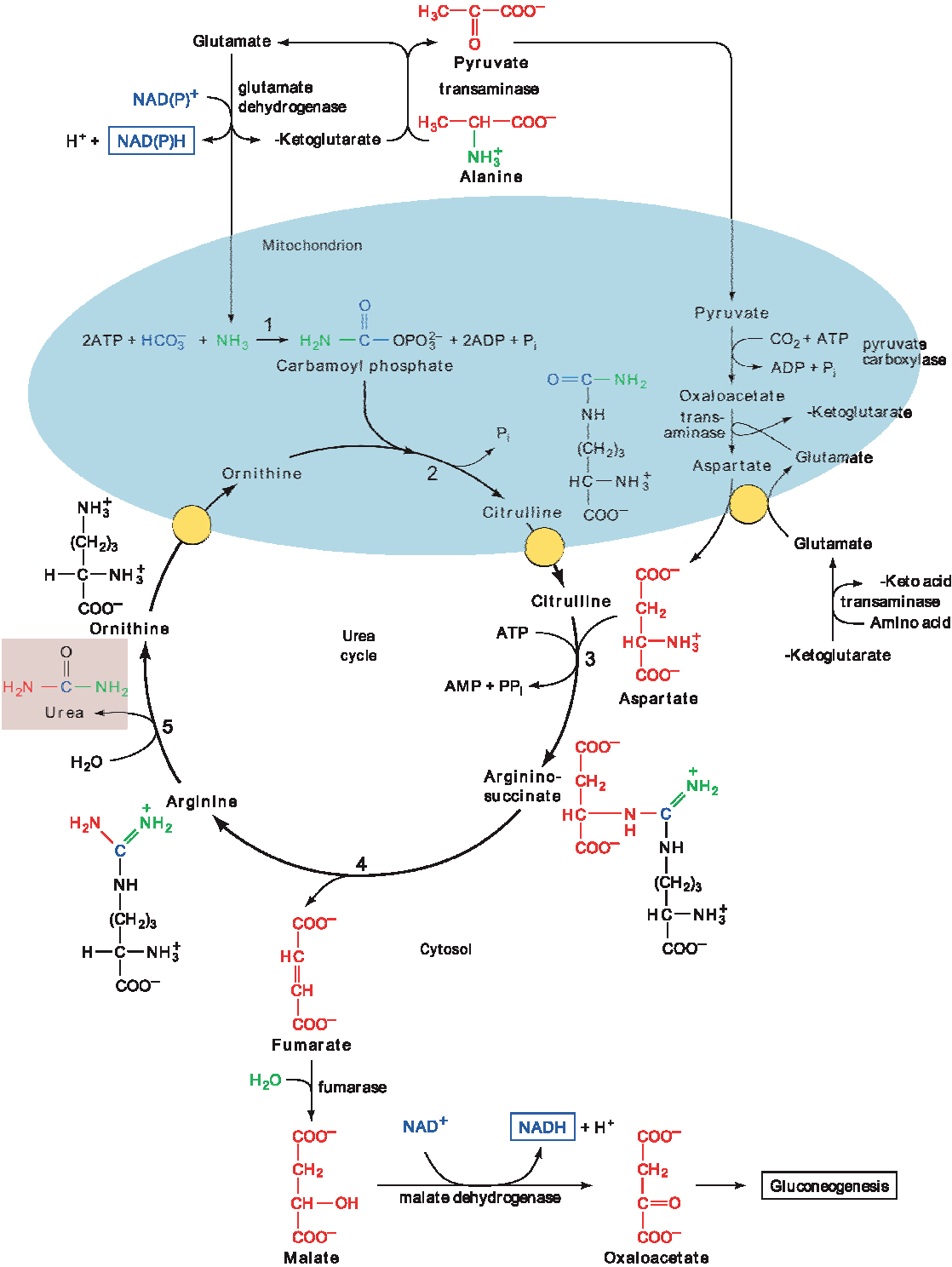
Figure 1: *The urea cycle. Enzymes: (1) carbamoyl phosphate synthetase, (2) ornithine transcarbamoylase, (3) argininosuccinate synthetase, (4) argininosuccinate lyase, and (5) arginase. The reactions occur in part in the mitochondrion (blue ellipse) and in part in the cytosol with ornithine and citrulline being transported across the mitochondrial membrane by specific transport systems (yellow circles). One of the urea amino groups (green) originates as the NH_3 product of the glutamate dehydrogenase reaction (top).The other amino group (red) is obtained from aspartate through the transfer of an amino acid to oxaloacetate via transamination (right).The fumarate product of the argininosuccinase reaction is converted to oxaloacetate for entry into gluconeogenesis via the same reactions that occur in the citric acid cycle but take place in the cytosol (bottom).The ATP utilized in reactions 1 and 3 of the cycle can be regenerated by oxidative phosphorylation from the NAD(P)H produced in the glutamate dehydrogenase (top) and malate dehydrogenase (bottom) reactions (Figure adapted from Voet D and Voet JG, 1995).*

1.1.2. Urea cycle disorders

Urea cycle disorders (UCDs) are inborn errors in the metabolism of waste nitrogen that can lead to brain damage and death due to increased levels of blood ammonia. They involve deficiencies in one of the enzymes required by the urea cycle: carbamoyl phosphate synthetase 1 (CPS1D, MIM#237300), ornithine transcarbamylase (OTCD, MIM#311250), argininosuccinate synthetase (ASSD, MIM#215700), argininosuccinate lyase (ASLD, MIM#207900) and arginase 1 (ARG1D, MIM#207800), each with considerable genetic and phenotypic variability (Table 1). They also encompass deficiencies of N-acetylglutamate synthase (NAGS) (MIM#237310) associated with lack of the N-acetylglutamate (NAG), essential allosteric activator of CPS1, and of the mitochondrial ornithine/citrulline transporter (ORNT1), causing the hyperornithinemia-hyperammonemia-homocitrullinuria (HHH) syndrome (MIM#238970) (Table 1). All listed deficiencies are inherited in an autosomal recessive manner, except for OTCD which is X-linked. The overall prevalence of UCDs varies between 1:8,000 - 1:44,000 (Brusilow SW and Maestri NE, 1996; Dionisi-Vici C, et al., 2002; Wilcken B, 2004); however, these frequencies might be underestimated because of unreliable newborn screening and underdiagnosis of fatal cases.

UCDs are commonly characterized by hyperammonemia followed by encephalopathy, and respiratory alkalosis. Four of these diseases - CPS1D, OTCD, ASSD, ASLD - are characterized by signs and symptoms induced by the accumulation of precursors, principally ammonium and

glutamine. The fifth disease, arginase deficiency, is characterized by a clinical picture consisting of progressive spastic quadriplegia and mental retardation; symptomatic hyperammonemia, which can be life-threatening, occurs neither as severely or as commonly as in the other four diseases (Brusilow SW and Horwich AL, 2004). Infants with an UCD often initially appear normal but the progression symptoms moves rapidly within 24 – 48 from somnolence, through lethargy, to coma accompanied with a loss of thermoregulation resulting in low core temperature and feeding disruption (Summar M, 2001). The hyperammonemia is less severe and the symptoms more subtle in patients with partial enzyme deficiencies. In such cases, ammonia accumulation may be triggered by illness or stress at almost any time of life, resulting in multiple mild elevations of plasma ammonia concentration followed by mild signs of encephalopathy (Bourrier P, et al., 1988). The most common of these late-onset diseases occurs in female carriers of a mutation in *OTC* on one of their X chromosomes.

The diagnosis of a urea cycle disorder is based on clinical suspicion and biochemical and molecular genetic testing. A plasma ammonia concentration of 150 μmol/L or higher associated with a normal anion gap and a normal plasma glucose concentration indicates potential presence of an UCD. Plasma quantitative amino acid analysis can then distinguish between the specific UCDs. Plasma concentration of arginine may be reduced in all urea cycle disorders, except ARG deficiency, in which it is elevated 5-7 fold. Plasma concentration of citrulline helps to discriminate between the proximal and distal urea cycle defects, as citrulline is the product of the proximal enzymes (OTC and CPS1) and a substrate for the distal enzymes (ASS, ASL, ARG) (Lanpher BC, et al., 2003). Urinary orotic acid is measured to distinguish CPSI deficiency and NAGS deficiency from OTC deficiency (Leonard JV and Morris AA, 2002). A definitive diagnosis of a urea cycle defect depends on either results of molecular genetic testing or measurement of enzyme activity in a liver biopsy specimen. Molecular genetic testing is available for all urea cycle defects.

The treatment of UCDs in the acute phase is based on rapid reduction of plasma ammonia concentration by dialysis, pharmacological activation of alternative excretion pathways of excess nitrogen using nitrogen scavengers and on the reduction of neurologic damage related risks. The therapy is accompanied by rigorous dietary modifications in order to limit protein breakdown by discontinuing protein intake and supply sufficient glucose intravenously to limit catabolism. However, complete restriction of protein should not exceed 24 - 48 hours, because

depletion of essential amino acids results in protein catabolism and nitrogen release (Summar M, 2001).

Long-term management requires a decrease of nitrogen load by dietary protein restriction and the use of nitrogen scavengers to provide alternative routes for nitrogen disposal such as sodium phenylbutyrate or sodium benzoate. While sodium phenylbutyrate is converted in the gut to sodium phenylacetate, which is conjugated in the liver with glutamine to form phenylacetylglutamine, sodium benzoate is conjugated in the liver with glycine to make hippuric acid. Both products are easily excreted by the kidneys. Also prompt replacement of citrulline or arginine may be necessary depending on whether the defect is in a proximal or distal urea cycle disorder. In the proximal UCDs, the defects block synthesis of arginine, which, in combination with protein restricted diet, can lead to arginine deficiency. Supplementation is therefore essential. Moreover, additional arginine will facilitate increased production of citrulline and arginosuccinate, in ASSD and ASLD respectively, thus contributing to nitrogen excretion. In CPS1D and OTCD, citrulline supplementation can be substituted for arginine administration. Citrulline offers the advantage of incorporating aspartate into the pathway thus pulling one additional nitrogen molecule into the cycle.

Despite intensive treatment, patients can suffer from recurrent hyperammonemia events, notably during intercurrent illness. In such patients blood tests monitoring levels of plasma ammonia should be performed routinely, and hospitalizations may be considered if levels rise to high.

Table 1: Overview of the urea cycle disordes

	Disease	Gene symbol	Genomic locus	MIM ID	Mechanism
Urea cycle enzyme deficiency	N-acetylglutamate synthase (NAGS) deficiency	NAGS	17q21.31	237310	Enzymatic block within urea cycle
	Carbamoyl phosphate synthetase 1 (CPS1) deficiency	CPS1	2p35	237300	Enzymatic block within urea cycle
	Ornithine carbamoyltransferase (OTC) deficiency	OTC	Xp21.1	311250	Enzymatic block within urea cycle
	Argininosuccinate synthetase (ASS) deficiency	ASS1	9q34.1	215700	Enzymatic block within urea cycle
	Argininosuccinate lyase (ASL) deficiency	ASL	7cen-q11.2	207900	Enzymatic block within urea cycle
	Arginase 1 (AR1) deficiency	ARG1	6q23	207800	Enzymatic block within urea cycle
Urea cycle transporter defect	Hyperammonemia-Hyperornithinemia-Homocitrullinuria syndrome	SLC25A15	13q14	238970	Lack of mitochondrial ornithine as one of the OTC substrates
	Citrullinemia type II	SLC25A13	7q21.3	603471, 605814	Lack of aspartic acid as one of the ASS substrates

1.1.3. The *OTC* gene

Ornithine carbamoyltransferase (OTC, EC 2.1.3.3, MIM#300461) is a mitochondrial matrix enzyme catalyzing synthesis of the citrulline in the second step of the urea cycle in mammals. Human OTC gene is located on the short arm of chromosome X (Xp2l.1) (Lindgren V, et al., 1984). The gene spans 73 kb and comprises 10 exons encoding a 354-amino acid monomer (Hata A, et al., 1988). Mature mitochondrial human OTC is composed of three polypeptides with identical primary structure comprising 322 amino acid residues with total molecular weight of approximately 36 kDa (Kalousek F, et al, 1978). The remaining 32 amino acid residues are related to the leader peptide (Horwich AL, et al., 1984) which is recognized by a complex of molecular chaperons preventing irreversible misfolding involving the heat shock cognate 70 protein (hsc70), a human homolog of bacterial heat shock protein DnaJ called pHSDJ, presequence binding factor (PBF), and mitochondrial stimulating factor (MSF). The chaperon complex is also responsible for targeting the OTC subunit to the mitochondrial matrix

where, after hydrolytic cleavage of leading sequence, homotrimeric mature OTC is assembled by mhsp60, cpn10 and ATP (Snodgrass PJ, 2004).

Genomic variations in the *OTC* coding regions and adjacent intronic sequences were thoroughly investigated in a variety of detailed haplotype analyses (Azevedo L, et al., 2003; Tuchman M, et al., 2002; Yamaguchi S, et al., 2006) and a total of 434 various mutations have been listed in the latest release of HGMD database (HGMD® professional release 2013.2, http://www.hgmd.cf.ac.uk). Using the current molecular screening methods, mutations are found in about 80% of patients. A subset of non-confirmed patients may carry a causative mutation in regulatory domains (Yamaguchi S, et al., 2006). Such mutations, however, have not been reported yet.

The transcription start site (TSS) of human *OTC* was first assigned to the position c.-135 (GenBank K02100; Horwich AL, et al., 1984). Later published data suggest that transcription of human *OTC* does not initiate at a particular site but within a region of about 70 bases, mostly at preferred positions c.-95, c.-120, c.-150, c.-161 and c.-166 (Brusilow SW and Horwich AL, 2004). These facts are in partial correlation with the records of expressed sequence tags (ESTs) library containing several spliced variants: Four records are starting at position c.-92, one at position c.-120, two around position c.-151 and one is transcribed from the position 169 bp upstream of translation initiation site (TIS).

OTC is expressed predominantly in hepatocytes and epithelial cells of the intestinal mucosa (Ryall J, et al., 1985; Hamano Y, et al., 1988). The mechanism of the tissue specific pattern of *Otc* expression was first studied in mouse. It has been shown that the 0.8 kb 5'-flanking region of the mouse gene (Veres G, et al., 1986) and 1.3 kb region of the rat gene (Murakami T, et al., 1989) contained sufficient information to control *Otc* gene expression. However, this 5'-flanking region itself was apparently not sufficient to achieve physiological levels of *Otc* expression in the liver. In transgenic experiments using sparse fur mice, Jones, et al. (1990) showed that the *Otc* promoter was capable of direct liver- and intestine-specific transcription, but the expression level of the introduced gene was much lower than in wildtype mice with endogenous *Otc* expression (Jones SN, et al. 1990). Indeed, a 230 bp positive regulatory element (distal enhancer) situated 11 kb upstream of the TSS has been identified in the rat (Murakami T, et al., 1990).

Two positive *cis*-acting regulatory elements located around the positions -187 and -104 upstream of TIS were identified in the rat and mouse promoter. Both of them can bind an identical set of two members of orphan receptor superfamily: hepatocyte nuclear factor 4 (HNF-4) and chicken ovalbumin upstream promoter-transcription factor (COUP-TF). In co-

transfection experiments, HNF-4 activated transcription of a reporter gene from *Otc* promoter whereas competition with COUP-TF had a repressor effect (Kimura A, et al., 1993). Moreover, HNF-4 seems to contribute to the tissue specific expression of *Otc* as its mRNA is present only in a limited number of tissues such as the liver, kidney and intestine (Ryall J, et al., 1985; Hamano Y, et al., 1988). On the other hand, COUP-TF exhibits rather ubiquitous expression pattern (Miyajima N, et al., 1988) and therefore the competition for binding site can result in repression of *Otc* in other tissues. The HNF-4α is critical for urea homeostasis in mouse (Inoue Y, et al., 2002) since liver-specific HNF4α-null mice exhibited elevated serum ammonia and decreased serum urea, while the only significant decrease in protein expression within the urea cycle enzymes was that of OTC. Similarly as in rat, two HNF-4α recognition motifs were identified within the mouse 5'-flanking region at positions -111 bp and -191 bp upstream of the TIS and HNF-4α binding was confirmed using a specific antibody. Moreover, co-transfection of reporter constructs carrying mouse *Otc* promoter with a rat HNF-4α expression plasmid increased the reporter gene activity in the CV-1 cells (Inoue Y, et al., 2002).

Nishiyori A, et al. (1994) demonstrated that rat *Otc* enhancer is activated by co-operation of two liver-selective transcription factors: HNF-4 and CCAAT-enhancer-binding protein (C/EBP). Furthermore, co-transfection experiments showed that both HNF-4 and C/EBPβ are necessary and neither alone is sufficient for activation of the reconstituted enhancer in non-hepatic cells. Combinatorial operation of these two liver-enriched (but not strictly liver-specific) transcription factors leads to more restricted liver-specific transcription of the *Otc* gene (Nishiyori A, et al., 1994).

These data indicate that the tissue-specific Otc expression in rodents is based on the promoter-enhancer interaction in the hepatocyte-specific transcription factor milieu. No such regulatory elements have been reported in humans.

1.2. Gaucher disease

1.2.1. Lysosomal storage diseases

Lysosomes are subcellular organelles responsible for the physiologic turnover of cell constituents. They contain a variety of catabolic enzymes operating in low pH environment, that degrade biological macromolecules such as proteins, complex carbohydrates, nucleic acids, lipids, sulfates, and phosphates. The end products are either reused by the cell or eliminated from the body. The absence or loss of function of 1 enzyme along the pathway can result in accumulation of an intermediate metabolic product.

Lysosomal storage diseases (LSDs) represent a heterogeneous group of more than 50 rare inherited disorders characterized by accumulation of waste products in the lysosomes (Table 2). Progressive accumulation of these products leads to cellular distortion and dysfunction (Wenger DA, et al., 2003; Wilcox WR, 2004). Widespread cellular destruction eventually causes tissue and organ dysfunction observed in clinical abnormalities. Classically, LSDs encompassed only enzyme deficiencies of the lysosomal hydrolases. More recently, the concept of lysosomal storage disease has been expanded to include deficiencies or defects in proteins necessary for the normal post-translational modification of lysosomal enzymes (which themselves are often glycoproteins), activator proteins, or proteins important for proper intracellular transport between the lysosome and other intracellular compartments (Table 2). Almost all LSDs are inherited as autosomal recessive traits, except for the X-linked Fabry and Hunter (mucopolysaccharidosis type II [MPS II]) diseases and Danon disease (Wilcox WR, 2004; Sugie Ket al., 2002). Individually, the incidence of these inherited diseases is rare, ranging between 1 in 50 000 and 1 in 4×10^6. However, taken together, LSDs are far more common, approximately 1 in 7000 to 8000 births making them a relatively common and significant health problem (Meikle PJ, et al., 1999). Certain populations have a higher incidence of particular LSD compared with the general population. Ashkenazi Jewish descendents are 50 to 60 times more likely to inherit mutations causing Gaucher, Tay-Sachs, Niemann-Pick type A, and mucolipidosis IV (Wilcox WR, 2004; Edelmann L, et al., 2002; Natowicz MR and Prence EM, 1996). Salla disease and aspartylglucosaminuria are more prevalent among those of Finnish descent, and type 3 Gaucher disease is most frequent among individuals with Swedish ancestry (Poorthuis BJ, et al., 1999; Mononen T, et al., 1991; Aula P, et al., 1986). As an increasing number of patients with milder forms are being identified, the current figures may underestimate the actual frequencies of lysosomal storage disorders.

The LSDs are primarily classified according to the nature of the stored material (Table 2). More recently they have tended to be classified by the molecular basis of the defect. Broad categories include MPS, GM_2 gangliosidoses, neutral glycosphingolipidoses, glycoproteinoses, mucolipidoses, leukodystrophies, glycogen storage diseases, disorders of neutral lipids, and disorders of protein transport or trafficking (Hopkin R and Grabowski G, 2004; Wilcox WR, 2004). The emerging classification of diseases based on the recent understanding of the molecular basis LSDs is reported in Table 2 including groups of disorders caused by non-enzymatic lysosomal protein defects, transmembrane protein defects, lysosomal enzyme protection defects, post-translational processing defects of lysosomal enzymes, trafficking defects in lysosomal enzymes and polypeptide degradation defects (Filocamo M and Morrone A, 2011).

Table 2: Overview of lysosomal storage disorders

Disease	Defective protein	Main storage materials	Gene symbol	Genomic locus	MIM ID
Mucopolysaccharidoses (MPSs)					
MPS I (Hurler, Scheie, Hurler/Scheie)	α-Iduronidase	Dermatan sulphate, heparan sulphate	*IDUA*	4p16.3	607014, 607015, 607016
MPS II (Hunter)	Iduronate sulphatase	Dermatan sulphate, heparan sulphate	*IDS*	Xq28	309900
MPS III A (Sanfilippo A)	Heparan sulphamidase	Heparan sulphate	*SGSH*	17q25.3	252900
MPS III B (Sanfilippo B)	Acetyl α-glucosaminidase	Heparan sulphate	*NAGLU*	17q21	252920
MPS III C (Sanfilippo C)	Acetyl CoA: α-glucosaminide N-acetyltransferase	Heparan sulphate	*HGSNAT*	8p11.21	252930
MPS III D (Sanfilippo D)	N-acetyl glucosamine-6-sulphatase	Heparan sulphate	*GNS*	12q14.3	252940
MPS IVA (Morquio A)	Acetyl galactosamine-6-sulphatase	Keratan sulphate, chondroiotin 6-sulphate	*GALNS*	16p24.3	253000
MPS IV B (Morquio B)	β-Galactosidase	Keratan sulphate	*GLB1*	3p22.3	253010
MPS VI (Maroteaux-Lamy)	Acetyl galactosamine 4-sulphatase (arylsulphatase B)	Dermatan sulphate	*ARSB*	5q14.1	253200
MPS VII (Sly)	β-Glucuronidase	Dermatan sulphate, heparan sulphate, chondroiotin 6-sulphate	*GUSB*	7q11.21	253220
MPS IX (Natowicz)	Hyaluronidase	Hyluronan	*HYAL1*	3p21.31	601492
Sphingolipidoses					
Fabry	α-Galactosidase A	Globotriasylceramide	*GLA*	Xp22.1	301500
Farber	Acid ceramidase	Ceramide	*ASAH1*	8p22	228000
Gangliosidosis GM1 (Types I, II, III)	GM1-β-galactosidase	GM1 ganglioside, Keratan sulphate, oligos, glycolipids	*GLB1*	3p22.3	230500, 230600, 230650
Gangliosidosis GM2, Tay-Sachs	β-Hexosaminidase A	GM2 ganglioside, oligos, glycolipids	*HEXA*	15q23	272800
Gangliosidosis GM2, Sandhoff	β-Hexosaminidase A + B	GM2 ganglioside, oligos	*HEXAB*	15q23, 5q13.3	268800
Gaucher (Types I, II, III)	Glucosylceramidase	Glucosylceramide	*GBA*	1q22	230800, 230900, 231000
Krabbe	β-Galactosylceramidase	Galactosylceramide	*GALC*	14p31.3	245200
Metachromatic leucodystrophy	Arylsulphatase A	Sulphatides	*ARSA*	22q13.33	250100
Niemann-Pick (type A, type B)	Sphingomyelinase	Sphingomyelin	*SMPD1*	11p15.4	257200, 607616
Olygosaccharidoses (glycoproteinoses)					
Aspartylglicosaminuria	Glycosylasparaginase	Aspartylglucosamine	*AGA*	4q34.3	208400
Fucosidosis	α-Fucosidase	Glycoproteins, glycolipids, Fucoside-rich oligos	*FUCA1*	1p36.11	230000
α-Mannosidosis	α-Mannosidase	Mannose-rich oligos	*MAN2B1*	19p13.2	248500
β-Mannosidosis	β-Mannosidase	Man(β1 → 4)GlnNAc	*MANBA*	4q24	248510
Schindler	N-acetylgalactosaminidase	Sialylated/ asialoglycopeptides, glycolipids	*NAGA*	22q13.2	609241
Sialidosis	Neuraminidase	Oligos, glycopeptides	*NEU1*	6p21.33	256550
Glycogenoses					
Glycogenosis II/Pompe	α1,4-glucosidase (acid	Glycogen	*GAA*	17q25.3	232300

	maltase)				
Lipidoses					
Wolman/CESD	Acid lipase	Cholesterol esters	*LIPA*	10q23.31	278000
Non-enzymatic lysosomal protein defect					
Gangliosidosis GM2, activator defect	GM2 activator protein	GM2 ganglioside, oligos	*GM2A*	5q33.1	272750
Metachromatic leucodystrophy	Saposin B	Sulphatides	*PSAP*	10q22.1	249900
Krabbe	Saposin A	Galactosylceramide	*PSAP*	10q22.1	611722
Gaucher	Saposin C	Glucosylceramide	*PSAP*	10q22.1	610539
Transmembrane protein defect					
Transporters					
Sialic acid storage disease; infantile form (ISSD) and adult form (Salla)	Sialin	Sialic acid	*SLC17A5*	6q13	269920, 604369
Cystinosis	Cystinosin	Cystine	*CTNS*	17p13.2	219800
Niemann-Pick Type C1	Niemann-Pick type 1 (NPC1)	Cholesterol and sphingolipids	*NPC1*	18q11.2	257220
Niemann-Pick, Type C2	Niemann-Pick type 2 (NPC2)	Cholesterol and sphingolipids	*NPC2*	14q24.3	607625
Structural Proteins					
Danon	Lysosome-associated membrane protein 2	Cytoplasmatic debris and glycogen	*LAMP2*	Xq24	300257
Mucolipidosis IV	Mucolipin	Lipids	*MCOLN1*	19p13.2	252650
Lysosomal enzyme protection defect					
Galactosialidosis	Protective protein cathepsin A (PPCA)	Sialyloligosaccharides	*CTSA*	20q13.12	256540
Post-translational processing defect					
Multiple sulphatase deficiency	Multiple sulphatase	Sulphatides, glycolipids, GAGs	*SUMF1*	3p26.1	272200
Trafficking defect in lysosomal enzymes					
Mucolipidosis IIα/β, IIIα/β	GlcNAc-1-P transferase	Oligos, GAGs, lipids	*GNPTAB*	12q23.2	252500, 252600
Mucolipidosis IIIγ	GlcNAc-1-P transferase	Oligos, GAGs, lipids	*GNPTG*	16p13.3	232605
Polypeptide degradation defect					
Pycnodysostosis	Cathepsin K	Bone proteins	*CTSK*	1q21.3	265800
Neuronal ceroid lipofuscinoses (NCLs)					
NCL 1	Palmitoyl protein thioesterase (PPT1)	Saposins A and D	*PPT1*	1p34.2	256730
NCL 2	Tripeptidyl peptidase 1 (TPP1)	Subunit c of ATP synthase	*TPP1*	11p15.4	204500
NCL 3	CLN3, lysosomal transmembrane protein	Subunit c of ATP synthase	*CLN3*	16p11.2	204200
NCL 5	CLN5, soluble lysosomal protein	Subunit c of ATP synthase	*CLN5*	13q22.3	256731
NCL 6	CLN6, transmembrane protein of ER	Subunit c of ATP synthase	*CLN6*	15q23	601780
NCL 7	CLC7, lysosomal chloride channel	Subunit c of ATP synthase	*MFSD8*	4q28.2	610951
NCL 8	CLN8, transmembrane protein of endoplasmic reticulum	Subunit c of ATP synthase	*CLN8*	8p23.3	600143
NCL 10	Cathepsin D	Saposins A and D	*CTSD*	11p15.5	610127

The spectrum of clinical phenotypes is very broad including organomegaly and central nervous system dysfunction. The patients suffering from LSD are mostly born apparently healthy and the symptoms develop progressively. In several LSDs the infantile, juvenile and adult form of disease is distinguished depending on the age of onset. The severity and extent of the disease are often inconsistent, even within families, and patients with identical genotypes can have significantly different phenotypes (Wilcox WR, 2004). Disease phenotypes may differ in age of onset, complexity of the storage product, rate of substrate accumulation, and tissue distribution (Wilcox WR, 2004). These observations of disease heterogeneity are related, in part, to the complexity of enzyme kinetics. Each tissue type has a set threshold of enzymatic activity, below which pathological changes manifest (Hopkin R and Grabowski G, 2004). Threshold activities are dependent on substrate flux, cellular turnover, and metabolic demand. The residual activity of the defective protein additionally complicates the kinetic model since small variations can greatly influence the rate of substrate accumulation and disease development (Beck M, 2001). Genetic background, environmental factors, and lack of the final metabolic product of the affected pathway can also influence pathological expression (Hopkin R and Grabowski G, 2004; Wilcox WR, 2004; Beck M, 2001).

Many LSDs have central CNS involvement, with or without somatic features. Tissues that normally have a high flux of the accumulating substrate are most affected. Clinical features suggestive of a LSD include developmental delay, progressive regression after a period of normal development, ataxia, seizures, weakness, and dementia (Wenger DA, et al., 2003). A LSD diagnosis should also be considered in the presence of course facial features, bone abnormalities, corneal clouding, unexplained joint stiffness, unexplained bone pain or burning neuropathic pain, psychiatric problems, nonimmune fetal hydrops, or organomegaly (Wenger DA, et al., 2003; Wilcox WR, 2004). The diagnosis is definitively confirmed by the detection of enzymatic deficiency at the molecular level. Lysosomal enzyme activities are usually determined by a fluorometric assay in cultured fibroblasts, leukocytes or sera. Biochemical testing is often accompanied with molecular genetic testing, which also enables prenatal and postnatal testing and allows the provision of genetic counseling. In addition, molecular genetic testing can clarify the type of genetic variation and its impact on the protein and on the presence of residual enzyme activity. This information is crucial in evaluating treatment options (Marsden D and Levy H, 2010).

For a majority of LSDs, therapeutic management consists of symptomatic care of disease manifestations with no possibility for cure. Supportive care measures are disease specific, depending on the organs involved and degree of physical impairment. However, there are no

known cures for LSDs. Some alternative therapeutic approaches such as hematopoietic stem cell transplantation and enzyme replacement therapy are recently available treatment options for patients with certain types of LSDs.

1.2.2. Metabolism of glycosphingolipids and Gaucher disease

Glycosphingolipids (GSLs) are ubiquitous constituents of eukaryotic plasma membranes and the major glycans of the vertebrate brain, where more than 80% of glycoconjugates are in the form of glycolipids. They are composed of a glycan structure attached to a lipid tail that contains the sphingolipid ceramide. The basic structure for a glycosphingolipid is a monosaccharide, usually glucose or galactose, attached directly to a ceramide molecule and resulting in glucosylceramide (GlcCer) or galactosylceramide (GalCer) respectively. The core glycan structure may be extended by additional monosaccharides. This combination structure results in an amphiphilic molecule with a hydrophilic carbohydrate region and a hydrophobic lipid region. In addition to variations in the structure of the glycan, the ceramide structure may also show variations. The fatty acid attached to the sphingosine may contain carbon chain lengths from C_{14} to C_{24} and vary in degree of unsaturation and/or hydroxylation. Depending on the glycan structure and modifications the GSLs are divided into four principal classes: cerebrosides, sulfatides, globosides and gangliosides.

Once incorporated into the cytoplasmatic membrane, glycosphingolipids are clustered in small lateral microdomains of self-associating membrane molecules called "lipid rafts". In that case, the ceramide lipid moiety is embedded in the outer leaflet of the membrane and the oligosaccharide structures are oriented into the extracellular milieu. They form cell type specific profiles which characteristically change in development, differentiation and oncogenic transformation, suggesting their implication in fundamental cellular processes including growth, differentiation, morphogenesis, cell to matrix interaction and cell to cell communication. The functions of membrane-associated glycosphingolipids could be divided into two major categories: mediating cell–cell interactions via binding to complementary molecules on opposing plasma membranes (*trans* recognition) and modulating activities of proteins in the same plasma membrane (*cis* regulation). More complex glycosphingolipids thus contribute to the cell–cell recognition and to the regulation of signal transduction.

The biosynthesis of GSLs occurs in a stepwise fashion, with an individual sugar added first to ceramide and then subsequent sugars transferred by glycosyltransferases from nucleotide sugar donors. Ceramide is synthesized on the cytoplasmic face of the endoplasmic reticulum

(ER); it subsequently equilibrates to the lumenal face and traffics to the Golgi compartment. GlcCer is synthesized on the cytoplasmic face of the ER and early Golgi apparatus; it then flips into the Golgi lumen, where it is typically elongated by a series of glycosyltransferases. In contrast, GalCer is synthesized on the lumenal face of the ER and then traffics through the Golgi, where it may be sulfated to form sulfatide. The process is regulated by the availability of nucleotide sugar donors used by glycosyltransferases in the Golgi lumen (including UDP-Gal, UDP-Glc, UDP-GlcNAc, UDP-GalNAc, and CMP-NeuAc), which ultimately affect the final structure of glycans. Their levels are modulated by synthetic enzymes located in the cytoplasm or nucleus and by the activity of nucleotide sugar transporters in the Golgi membrane. An additional level of regulation may occur via stable association of different glycosphingolipid glycosyltransferases into functional "multiglycosyltransferase" complexes. The multiple enzymes are then thought to act concertedly on the growing glycosphingolipid without releasing intermediate structures, ensuring progression to the preferred end product.

The breakdown of complex GSLs occurs stepwise by sequential removal of sugars by lysosomal hydrolases. Glycosphingolipids on the outer surface of the plasma membrane are internalized, along with other membrane components, in invaginated vesicles that then fuse with endosomes, resulting in the glycan facing the endosome lumen. Glycosphingolipid-enriched areas of the endosomal membrane may then invaginate once again to form multivesicular bodies within the endosome. When endosomes fuse with primary lysosomes, glycosphingolipids become exposed to lysosomal hydrolases. The GSLs are metabolized to the common components monosacharides, sphingosine and fatty acids, which are then available for reuse. Individual defects in GSL hydrolases and other related proteins result in excessive accumulation of specific GSLs in lysosomes leading to the various LSDs (Figure 2).
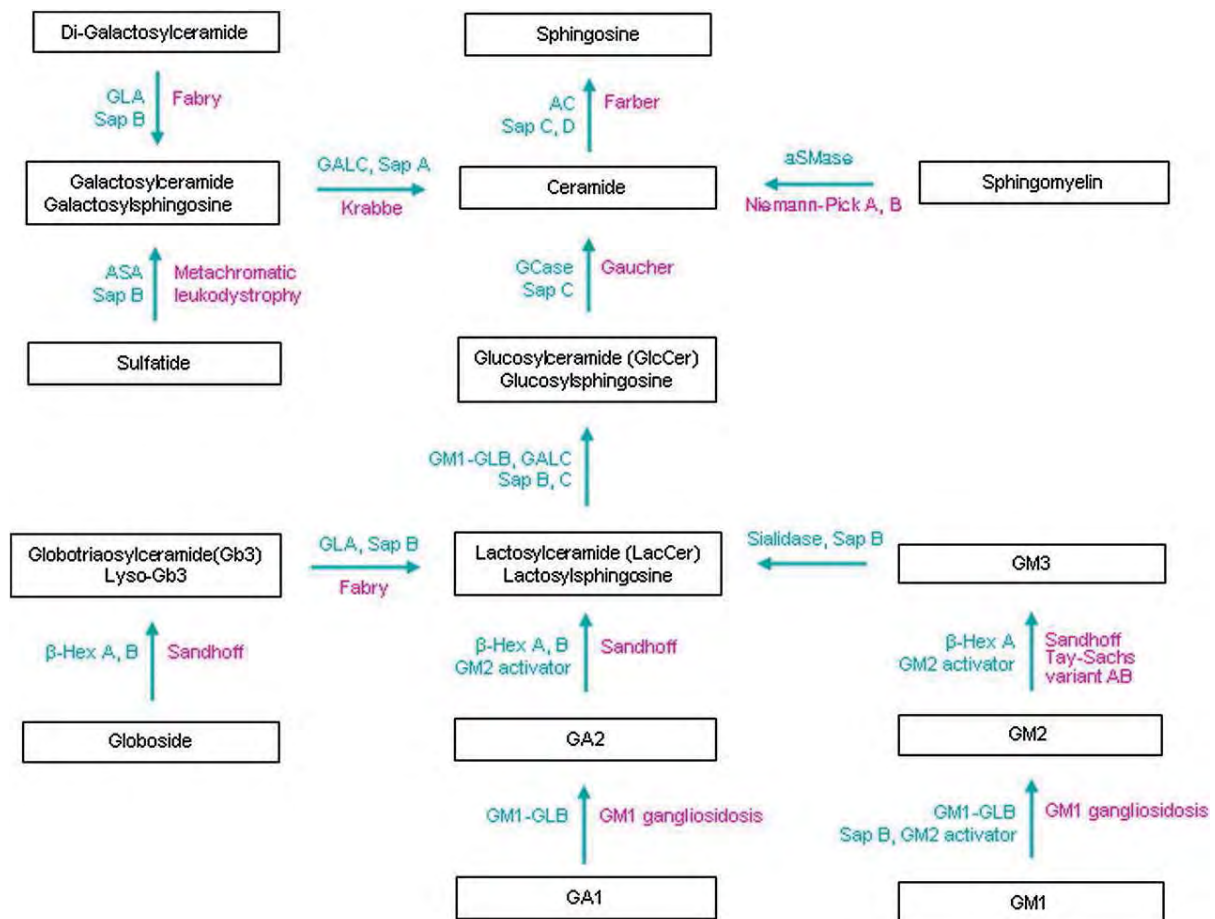
Figure 2: *Disorders of GSL and ganglioside degradation. Inherited diseases (violet) caused by genetic defects of individual hydrolases/proteins (blue) in the GSL and ganglioside degradation pathway. Abbreviations: AC - acid ceramidase; ASA, arylsulphatase A; aSMase, sphingomyelinase; β-Hex, β-hexosaminidase; GALC, β-galactosylceramidase; GCase, glucosylceramidase; GLA, α-galactosidase A; GLB, β-galactosidase; GA1/2 and GM1/2, ganglioside types; Sap, saposin (Figure adapted from Xu YH, et al., 2010).*

The most common glycosphingolipid storage disease is Gaucher's disease (GD), a rare autosomal recessive storage disorder caused by the by a functional deficiency of the acid hydrolase β-glucocerebrosidase (EC.3.2.1.45). The immediate substrates for this enzyme are GlcCer and its nonacylated analog, glucosylsphingosine. These glycosphingolipids arise from the digestion of more complex glucosides and gangliosides present in cell membranes. Rare variant forms of Gaucher disease result from deficiency of the sphingolipid activator protein, saposin C (Beutler E and Grabowski GA, 2001). The storage of substrate primarily occurs in the cells of the reticulo-endothelial system. The classic cellular hallmarks of GD are Gaucher cells, macrophages with substrate-laden lysosomes. These macrophages accumulate namely in the spleen and liver causing organ enlargement, as well as inflammation. Gaucher disease may

occur at any age and in any human population with the overall prevalence of approximately 1:57,000 to 1:75,000 births (Meikle PJ, et al., 1999; Biegstraaten M, et al., 2008). Nevertheless, GD is predominant in some populations, such as Ashkenazi Jews, where it represents the most common inherited disorder with the prevalence of about 1:500 births (Beutler E, et al., 1993).

The GD has been divided into three clinical subgroups reflecting the degree and rate of progression of involvement in the central nervous system (Westbroek W, et al., 2011). The most frequent form is the non-neuronopathic GD type 1 (OMIM#230800) affecting about 90% of patients. Clinical symptoms are restricted to the hematopoietic and skeletal system and visceral organs. Commonly encountered manifestations include hepatosplenomegaly, anemia, thrombocytopenia, and bone involvement. Many subjects have few manifestations and do not reach diagnosis. This form is common in the people of Ashkenazi Jewish ancestry (Hruska KS, et al., 2008; Goker-Alpan O, et al., 2005).

Type 2 or acute neuropathic GD (OMIM#230900) is the rarest and most severe type; it most often presents in the first 6 months of life. Type 2 GD is associated with devastating and progressive neurological manifestations leading to early death. Patients present regression of developmental milestones, brainstem impairment, strabismus and opisthotonus. Seizures, impaired gag reflex, and aspiration are common. Type 2 GD also has a variety of clinical presentations and can be associated with hydrops fetalis, which is characterized by an abnormal accumulation of fluid in the fetus and congenital ichthyosis, a skin disorder characterized by dry, thickened, and flaky skin (Hruska KS, et al., 2008; Goker-Alpan O, et al., 2005).

Type 3 or chronic neuronopathic GD (OMIM#2301000) has a more slowly progressive course and manifests with a spectrum of distinct clinical symptoms. Patients with type 3 GD have a characteristic eye movement abnormality consisting of slowing of the horizontal saccades. Type 3 GD can also be associated with myoclonic epilepsy, cognitive impairment or psychiatric disturbances. Patients within this subgroup develop calcifications of the aorta, hydrocephalus and/or dysmorphic features. Type 3 GD is more common in the Norbottnian region of Sweden (Hruska KS, et al., 2008; Goker-Alpan O, et al., 2005).

Severity and rate of disease progression widely varies, especially in adults, which makes treatment decisions extremely difficult in some patients. Currently there is no effective treatment for type 2 Gaucher patients. Therapeutic management of type 1 and type 3 GD include enzyme replacement therapy using intravenously applied artificial enzymes or bone marrow transplantation.

Enzyme replacement therapy is based on the provision of sufficient exogenous enzyme to overcome the block in the catabolic pathway thus contributing to the clearance of the stored substrate. The treatment is frequently effective in people with type 1 GD. Treatment results in breakdown of stored GlcCer, reduction in liver and spleen size, amelioration or resolution of anemia and thrombocytopenia, decreased bone pain, and increased bone mineralization and remodeling over a period of several years. Unfortunately it remains unclear whether this therapy is effective for the neurological problems of Gaucher disease.

Bone marrow transplantation has been undertaken in individuals with severe GD, primarily those with chronic neurologic involvement (type 3 GD). Successful engraftment can correct the metabolic defect, improve blood counts, and reduce increased liver volume. In a few individuals, stabilization of neurologic and bone disease has occurred. However, the morbidity and mortality associated with this technique limit its use in individuals with type 1 and type 3 GD. Therefore, this procedure has been largely superseded by enzyme replacement therapy.

1.2.3. The *GBA* gene

The gene encoding the acid β-glucocerebrosidase (GBA, EC.3.2.1.45, MIM#606463) is located on the chromosome 1 at a gene-rich locus 1q21 (Entrez Gene) or 1q22 (Ensembl, HGNC). It comprises 11 exons and 10 introns, spanning a sequence of 7.8 kb. A highly homologous pseudogene (ps*GBA*), resulting from a tandem duplication event involving ancestral *GBA* and neighbouring metaxin genes, is located 16 kb downstream (Horowitz M, et al., 1989) and spans 5.5 kb in length, retaining the same organization of exons and introns as the functional gene. Despite their different lengths, caused by several *Alu* insertions in introns of *GBA*, there is 96% sequence identity between the functional *GBA* gene and ps*GBA*. Although the pseudogene is beeing consistently transcribed and the level of transcription in some cases seems to be equal to that of the functional gene, its mRNA is degraded and no functional enzyme is synthesized. An important feature useful in molecular diagnostic applications is the microdeletion of 55 bp located in exon 9 of ps*GBA*. The presence of a conserved pseudogene in the same locus is responsible for a subset of Gaucher mutations originating from recombination events between *GBA* and ps*GBA* (Eyal N, et al., 1990; Latham TE, et al., 1991).

Beside the major transcript of *GBA* spanning approximately 2.6 kb (Horowitz M, et al., 1989), several alternative mRNA sequences have been identified, including variants with additional non coding exons -1 and -2. Human *GBA* cDNA contains 2 in-frame ATG start

codons (Sorge J, et al., 1985), with the upstream ATG resulting in a protein with a 39-residue signal peptide and the downstream ATG resulting in a protein with a 19-residue signal peptide. The corresponding signal peptides differ in their hydrophobicity and either of them could function to produce active enzyme in cultured fibroblasts (Sorge J, et al., 1987). Both precursor variants are sharing the 536-residue sequence of functional glucocerebrosidase monomer. Only one ATG translation initiation signal is present in the mouse sequence which shares 86% sequence identity (O'Neill RR, et al., 1989).

Promoter regions of genes encoding lysosomal enzymes commonly share characteristic features of housekeeping genes. The promoter of the glucocerebrosidase functional gene is unusual in this respect since it contains TATA and CAAT boxes but no Sp1 binding site (Horowitz M, et al., 1989). Moreover, the *GBA* mRNA is expressed differentially and glucocerebrosidase activity varies widely between different cell types, thus indicating that regulation of transcription is an important factor in the control of glucocerebrosidase expression. On the other hand, other regulatory mechanisms should be considered as well as the ratio between *GBA* mRNA and enzyme activity differs between cell lines. (Doll RF and Smith FI, 1993). Analyses of the human proximal promoter point to the presence of both regulatory enhancers and suppressors in the region delimited by nucleotide positions c.-354 and c.267 according to the TIS. Some of these regulatory regions function constitutively, while others appear to function in a cell-specific fashion (Doll RF, et al., 1995). Protein binding on some of these elements was later confirmed by gel-shift and DNase I footprinting assays in a more restricted region ranging c.-106 to c.82 (Moran D, et al., 1997). The 250 bp region upstream of the initiation codon was recently identified as a highly conserved cluster of transcriptional regulatory elements by computational comparisons of high quality genomic sequences originating from human and 8 non-human vertebrates (Blech-Hermoni YN, et al., 2010). Another highly conserved cluster of transcription factor binding sites, a downstream enhancer spanning from the 3'UTR of *GBA* through intron 3 of neighboring *MTX1*, was found by the same method and transcriptional activity of both regions was confirmed in a set of reporter assays (Blech-Hermoni YN, et al., 2010).

More than 300 different mutations have been identified in patients with GD (Koprivica V, et al., 2000; Hruska KS, et al., 2008) and, up to date, 403 variations are recorded in the HGMD database (HGMD® professional release 2013.2, http://www.hgmd.cf.ac.uk). These mutations are scattered through *GBA*, and include missense, nonsense, and frame-shift mutations as well as insertions, deletions, and complex rearrangements. Unfortunately, successful identification

of mutant *GBA* alleles is often complicated by interference with the highly homologous pseudogene.

Genotype–phenotype correlations bring only a limited number of relevant conclusions. There is significant phenotypic variation not only among patients with the same disease type, but also among patients with identical genotypes (Sidransky E, et al. 1994; Grabowski GA, 1997). Most of mutations are found in patients with clinical signs of more than one type of GD (Beutler E, 1992; Grabowski GA, 1997; Koprivica V, et al., 2000; Sidransky E, et al., 1994). The most common mutations are p.N370S (c.1226 A > G), 84GG (c.84-85insG), p.L444P (c.1448 T > C), and IVS2+1 (c.115+1G>A). In a study of Ashkenazi Jews, p.N370S was found in 78% of patients whereas 95% of patients carried at least one of these common mutations. In non-Jewish populations, the most common mutation was p.L444P (36%), followed by p.N370S (29%) (Beutler E, 2006). The international registry of GD patients was created in 1991; in 1998, it included 1698 patients, of whom 84% had the p.N370S mutation in at least one allele and 23% in both alleles; 30% had the p.L444P mutation in at least one allele and 8% in both alleles (Charrow J, et al., 2000).

1.3.   Regulation of gene expression

1.3.1. Gene expression and control

Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. The multi-level process consisting of chromatin remodeling, transcription, RNA processing, translation and post-translational modification of proteins is controlled at different stages.
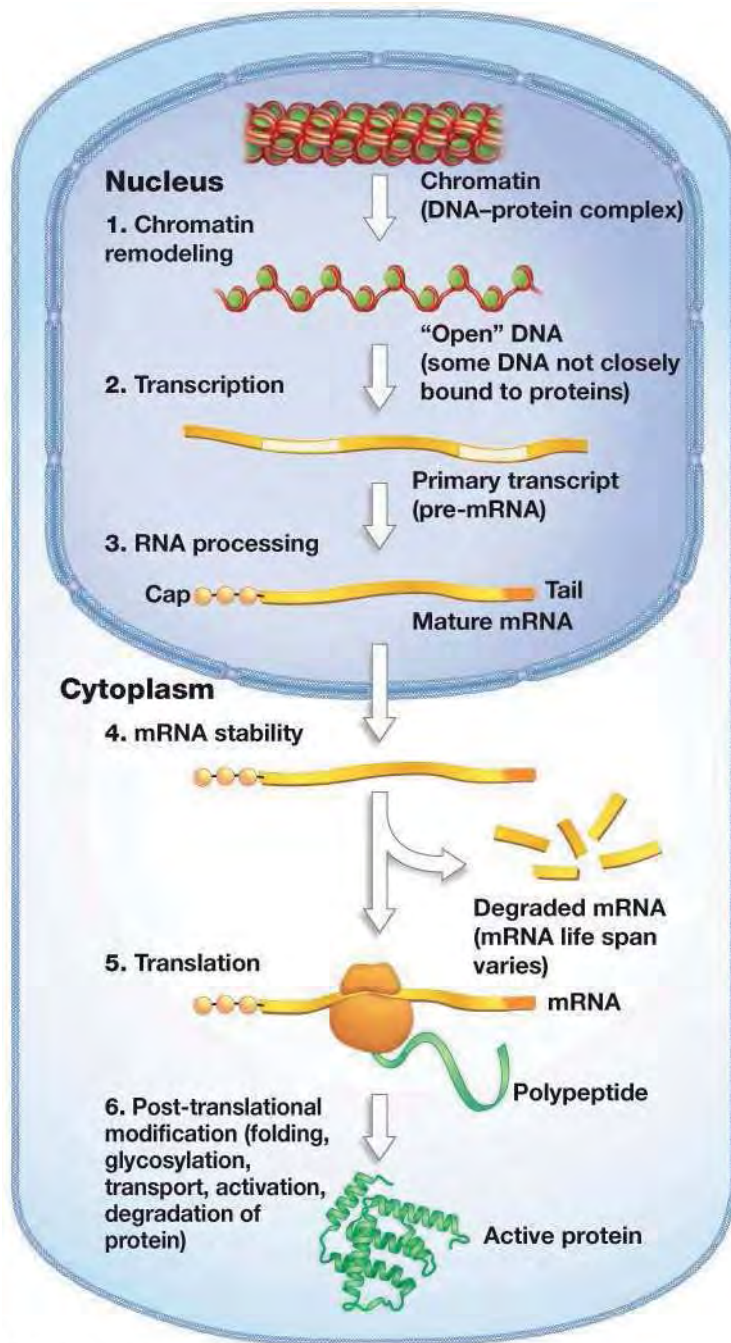
Figure 3: *Regulation of eukaryotic gene expression. Gene expression is a complex process towards the formation of a functional protein. It involves numerous events at the level of*

chromosome, DNA, RNA and protein. These, include chromatin remodeling (1), transcription (2), RNA processing (3, 4), translation (5) and post-translation modifications (6). Some of events take place in the nucleus (1, 2, 3),whereas protein formation and modification (5, 6) occurs in the cytoplasm. The process is controlled by complex regulatory mechanisms at all levels. Figure by Pearson Education, Inc., 2011.

On the largest scale, the expression of genetic information from chromosomes is performed by changes in chromatin structure. The local structure during interphase depends on the genes present on the DNA - the regions containing genes that are actively transcribed are packaged more loosely and are found associated with RNA polymerases (euchromatin) while regions carrying inactive genes are found associated with structural proteins and are packaged more tightly (heterochromatin). In eukaryotes, chromatin organization plays a critical role in transcriptional regulation (Figures 3 and 4). In the condensed form, chromatin can inhibit access of transcription factors to the DNA and can thereby repress gene expression. In eukaryotic organisms, with their very large number of genes, most of chromatin is condensed into a "coiled coil" structure to prevent non-controlled over-expression of related genes.

In order for transcription to occur, the area around a prospective transcription zone needs to be unwound. This is a complex process requiring the coordination of histone modifications, transcription factor binding and other chromatin remodeling activities. Once the DNA is open, specific DNA sequences are then accessible for specific proteins to bind. Many of these proteins are activators, while others are repressors; in eukaryotes, all such proteins are often called transcription factors (TFs). Each TF has a specific DNA binding domain that recognizes a 6-10 base-pair motif in the DNA sequence, as well as an effector domain. For an activating TF, the effector domain recruits RNA polymerase II (Pol II), the eukaryotic mRNA-producing polymerase, to begin transcription of the corresponding gene. Some activating TFs even turn on multiple genes at once. All TFs bind at the promoters just upstream of eukaryotic genes, similar to bacterial regulatory proteins. However, they also bind at regions called enhancers, which can be oriented forwards or backwards and located upstream or downstream or even in the introns of a gene, and still activate gene expression.

Most human genes are divided into exons and introns, and only the exons carry information required for protein synthesis. Most primary transcripts are therefore processed by splicing to remove intron sequences and generate a mature transcript or mRNA that only contains exons. Another processes, important for stability and transport of mature mRNA, are modifications on both, 5'- and 3'- termini. While the 5'- end capping by 7-methylguanosine occurs immediately

after formation of first 25 bases within the elongation phase, cleavage the 3' terminus followed by the addition of a poly A tail represents the termination of RNA transcription. As RNA transcription occurs in the nucleus and only mature mRNA can translocate through nuclear membrane, functional transcripts need to be modified before they are exported to the cytoplasm for translation. Transcript processing thus provides an additional level of regulation in eukaryotes. Moreover, due to the control of RNA splicing, eukaryotic cells can differentially express genes in a variety of alternative transcripts performing various functions in the organism.

Upon successful transcription, processing and transport to the cytoplasm, the mRNA is finally competent to perform its function - translation of the sequence encoded by the gene into the protein. The regulation at the level of translation is characterized by differential utilization of preexisting mRNAs. Compared to transcriptional regulation, translational control can ensure more rapid changes in cellular concentrations of critical gene products. Most regulation is exerted at the level of initiation where the pre-initiation complex (PIC) comprising the small ribosomal subunit with loaded Met-tRNAi scans the 5'-untranslated region (UTR) for the initiation codon. Recruitment of the PIC is modulated by mRNA secondary structure, antisense RNA binding and eukaryotic initiation factors (eIFs) that recognize the cap structure or the poly(A) tail. Other mechanisms responsible for specific translational control include alterations in translation elongation or poly(A) tail-length modulation. Initiation rate changes involve mainly the interaction of *trans*-acting factors (proteins or occasionally RNAs) with *cis*-acting elements located in the 5′ or 3′ untranslated regions of a particular mRNA (Sonenberg N and Hinnebusch AG, 2009; Jackson RJ, et al., 2010).

1.3.2. Regulation of transcription

As described in previous paragraph, the expression of eukaryotic protein-coding genes can be regulated at several steps, including transcription initiation, mRNA processing, transport and translation. Most regulation, however, is believed to occur at the level of transcription initiation. The control of gene expression at the level of transcription is performed by altering transcription rate – it controls when transcription occurs and how much mRNA is created. The dynamic process is mediated through interaction of TFs with DNA sequences commonly recognized as *cis*-acting regulatory elements (alternatively called boxes, DNA modules, initiator elements, or response elements). Genes transcribed by Pol II typically contain two distinct families of *cis*-acting transcriptional regulatory DNA elements: a promoter, which is

composed of a core promoter, and nearby (proximal) regulatory elements and distal regulatory elements, which can be enhancers, silencers, insulators, or locus control regions (Figure 4). While promoter is usually located directly in the 5'-flanking region, distal regulatory units can be found in any location of the chromosome. Folding, looping, or bending of the chromosomal DNA may bring such DNA regions close to the promoter. Also, TFs that bind to certain enhancers or response elements may form a chain of proteins and create bridges from one DNA site to another (Figure 4).
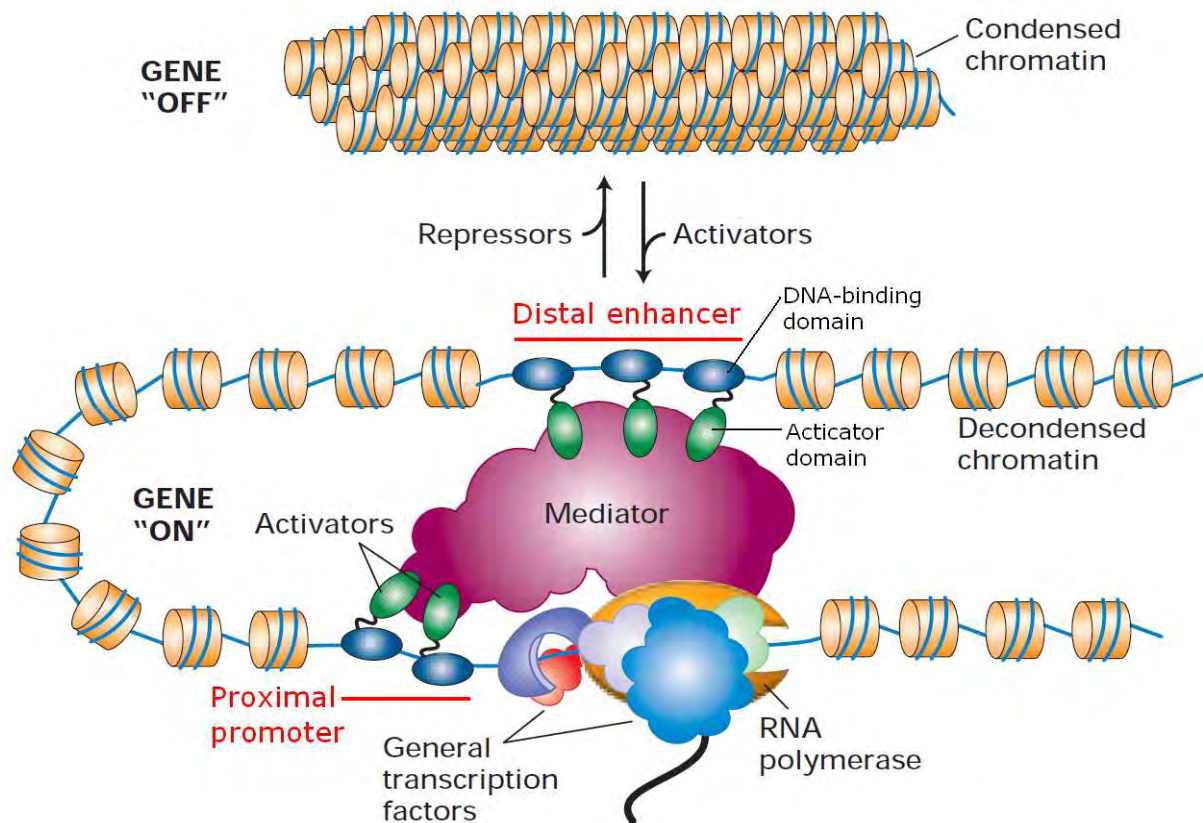
Figure 4: *Transcription control in eukaryotes. Activator proteins bind to specific DNA response elements and interact with proteins of co-activators, such as mediator complex, to decondense chromatin and assemble the pre-initiation complex of Pol II on promoters. Inactive genes are assembled into regions of condensed chromatin where they are protected against undesirable expression. The action of activators could be inhibited by repressor proteins or modulated by insulators bound to other control elements. Repressors can further interact with multiprotein co-repressor complexes to condense chromatin. The DNA chain loops and presence of a mediator complex enable the interaction with distant regulatory regions. The ability of different mediator subunits to interact with specific activation domains may contribute to the integration of signals from several activators at a single promoter. (Figure adapted from Lodish H, et al., 2000).*

The transcription mediated by the multi-subunit enzyme Pol II initiates at the core promoter in the close proximity of TSS (Orphanides G, et al., 1996; Roeder RG, 1996; Hampsey M, 1998; Woychik and Hampsey 2002). Accurate and efficient transcription from the core promoter requires the assembly of transcription PIC containing polymerase along with auxiliary factors that are commonly termed the general transcription factors, which include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. The first step in PIC assembly is binding of TFIID, a multisubunit complex consisting of a TATA-box-binding protein (TBP) and a set of 13 associated factors (TAFs). Transcription then proceeds through a series of steps, including promoter melting, clearance, and escape, before a fully functional RNA polymerase II elongation complex is formed. Interaction of TBP with core promoter could be facilitated and stabilized by TFIIA, which also acts as a co-activator for some transcriptional activators, assisting with their ability to increase transcription. The rest of PIC is assembled in following order: TFIIB, complex RNA polymerase II-TFIIF, TFIIE and TFIIH. It appears that factors TFIID and TFIIB have a critical role in the recognition of core promoter motifs.

The core promoter is considered as the minimal contiguous DNA sequence that is sufficient to direct accurate initiation of transcription by the Pol II machinery. Typically, the core promoter encompasses the site of transcription initiation and extends either upstream or downstream for other additional 35-40 nucleotides. There are several sequence motifs that are commonly found in core promoters such as the TATA box, initiator element (Inr), TFIIB recognition element (BRE), motif ten element (MTE) and downstream core promoter element (DPE) (Figure 5).
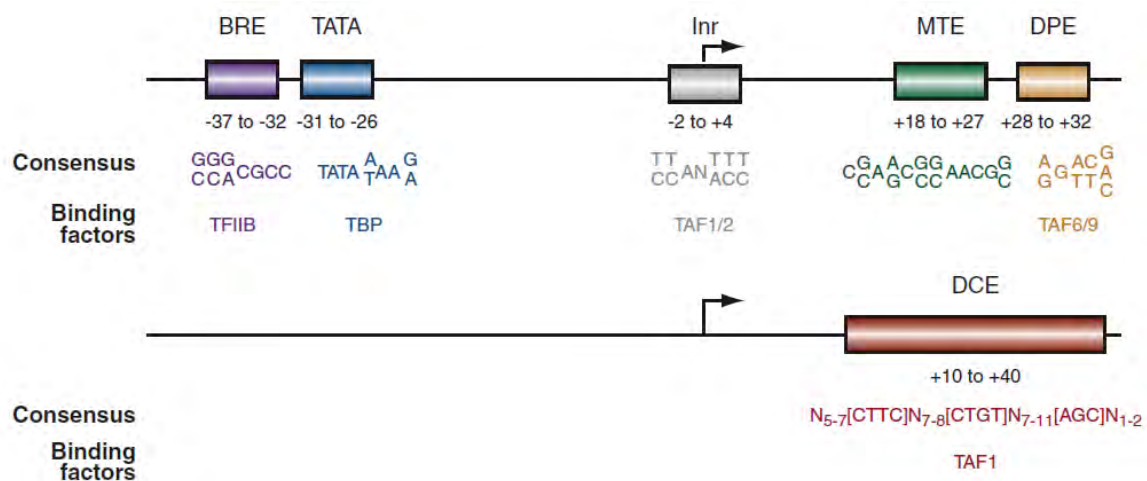
Figure 5.: *General core promoter elements. General core promoter elements are shown as color boxes with marked positions according to the TSS (black arrow). Consensus matrix*

The major core promoter-binding factor is TFIID. The TATA-box is directly recognized and bound by TBP, while the TAFs interact with sequences upstream and downstream to the TATA-box. In certain promoters, the TATA-box cooperates with one or more elements to direct efficient transcription initiation. Two TFIIB recognition elements BRE located either upstream (BREu) or downstream (BREd) of the TATA-box (Deng W and Roberts SG, 2005; Lagrange T, et al., 1998) function only together with the TATA-box. Similarly, the contribution to promoter strength of the TAF1 recognition element DCE that is located downstream relative to the TSS is also dependent on the presence of a TATA-box (Lee DH, et al., 2005).

Another element conserved in most eukaryotes and strictly located around the TSS is the Inr (Figure 5). The Inr can be weakly bound by Pol II itself or more strongly by PIC consisting of TFIIB, TFIID, TFIIF and Pol II (Carcamo J, et al., 1990) In that case the interaction is mediated via TAF1 and TAF2 subunits of TFIID (Smale ST and Baltimore D, 1989; Kaufmann J and Smale ST, 1994; Chalkley GE and Verrijzer CP, 1999). The Inr can function alone, together with the TATA-box or in conjunction with two specific downstream core promoter elements, the DPE and the MTE. Both elements are recognized by TFIID through the TAF6 and TAF9 subunits (Burke TW and Kadonaga JT, 1997; Burke TW, et al., 1998; Lim CY, et al., 2004; Theisen JW, et al., 2010).

It is important to note that each of these core promoter elements is found in some but not all core promoters. Bioinformatic studies revealed that about 46% of human promoters lack both TATA-like and consensus Inr elements and TATA-like motifs are present in only about 30% of human core promoters (Yang C, et al., 2007; Moshonov S, et al., 2008; Kim TH, et al., 2005). These are specially tissue specific genes and genes involved in development, response to wounding, response to external stimulus, inflammatory response and chromatin assembly (Moshonov S, et al., 2008; Schug J, et al., 2005). In general, TATA-like motifs and related elements are enriched genes that are highly regulated, whereas TATA-less genes are frequently involved in basic "housekeeping" processes (Yang C, et al., 2007). Most of the mammalian TATA-less promoters are associated with multiple CpG islands situated in the proximal promoter (Bajic VB, et al., 2006; FitzGerald PC, et al. 2004; Schug J, et al., 2005). Methylation at CpG dinucleotides is believed to repress transcription by blocking the ability of transcription factors to bind their recognition sequences, thus providing more complex control of gene

expression. Another poorly investigated phenomenon characteristic to a large number of TATA-less and Inr-less promoters is transcription initiation from multiple sites, as opposed to the single major site that is characteristic of promoters driven by a TATA-box or initiator (Frith MC, et al., 2008). In such cases the Pol II machinery is probably recruited by a transcription factor bound to a proximal promoter element. In the absence of a direct docking site, the Pol II is more flexible and can initiate transcription at favorable nucleotides in the vicinity of the element. Another possibility might be the presence of a number of docking sites on the same promoter to which the general machinery can weakly bind and direct transcription initiation.

In addition to the core promoter, other *cis*-acting DNA sequences that regulate RNA polymerase II transcription include the proximal promoter, enhancers, silencers, and boundary/insulator elements (Figure 5) (Blackwood EM and Kadonaga JT, 1998; Bulger M and Groudine M, 1999). These elements contain recognition sites for a variety of sequence-specific DNA-binding factors that are involved in transcriptional regulation. Enhancers and silencers can be located hundreds or thousands of nucleotides from the TSS and act either to activate or to repress transcription. Boundary/insulator elements appear to prevent the spreading of the activating effects of enhancers or the repressive effects of silencers or heterochromatin.

On the whole, the regulation of transcription is a complex phenomenon. Some target genes have a number of different response elements, with each one capable of responding to a particular signal that is sufficient to initiate transcription. In these cases, the structural gene can be activated both in different cell types and by diverse signals at different times during the life cycle of an organism. There are also instances where a unique transcription factor regulates the expression of a single structural gene. As well, proteins interact with a protein(s) of the transcription complex to block transcription, either before initiation or during the elongation process.

2. Aims of the study

Our main goal was to extend the possibilities of molecular genetic testing for OTCD and GD. We focused on the following specific aims:

1) To perform a detailed molecular characterization of *OTC* in 37 OTCD patients originating from 29 families, including description of novel disease alleles, precise breakpoint mapping of large rearrangements and genotype–phenotype correlation.

2) To define the 5'UTR, identify human *OTC* regulatory regions and investigate the functional impact of a novel unique variation in the 5'-flanking region in a female patient with clinical and biochemical signs of OTCD.

3) To perform detailed characterization of human *OTC* promoter and enhancer and identify key elements responsible for transcriptional regulation of *OTC* expression.

4) To identify transcription start site(s) of human *GBA* non-coding exon -2, investigate the upstream region for potential promoter activity and examine the expression pattern of alternative transcripts originating from the two alternative promoters in a group of 20 Ashkenazi Jewish Gaucher patients homozygous for the common mild mutation p.N370S.

3. Methods

3.1. General methods

Nucleic acid and protein isolation from human and animal cells and tissues

Polymerase chain reaction

Quantitative real-time polymerase chain reaction

DNA electrophoresis and restriction fragment length polymorphism analyses

Reverse transcription of total RNA (generation of cDNA)

DNA sequencing using the Sanger method

Molecular cloning

Cell culture techniques

3.2. Specific methods

Mutation analysis

Genomic DNA was isolated from peripheral leukocytes using QIAamp columns (Qiagen, Valencia, CA). Exons of human *OTC*, including the intron-exon boundaries, and four upstream non-coding regions, i.e. promoter, enhancer and two highly conserved regions, were analyzed by direct sequencing using the ABI Prism 3100 – Avant and ABI 3500 Genetic analyzers (Applied Biosystems, Foster City, CA). For details, primer sequences and reaction conditions see Ref I and Ref II.

Copy number analysis

In male patients, the approximate extent the deletions was assessed by inspection of presence/absence of PCR products. In females, the copy number of *OTC* coding exons was determined by the multiplex ligation-probe amplification assay (MLPA, MRC Holland, Amsterdam, The Netherlands) according to the manufacturers instructions as decribed in Ref I and Ref II.

Analysis of large deletions and gross chromosomal rearrangements

The analysis of gross chromosomal rearrangements was performed using the Affymetrix Human SNP 6.0 array (Santa Clara, CA). Data analysis was performed within the Affymetrix Genotyping Console 3.0.2 using CEU HapMap samples as the reference set. Breakpoints were located by sequencing of PCR products overlapping the deletion boundaries. For more details, see Ref I.

Determination of transcription start sites

The 5′-terminal mRNA sequence was analyzed by RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) using the First-Choice™ RACE ready liver kit with total human liver or placenta cDNA (Ambion, Austin, TX). Alternativelly, the GeneRacer™ Kit (Invitrogen) with RNAs isolated from human leukocytes, liver, and placenta was carried out. Both assays were performed according to the manufacturer's instructions as described in Ref II and Ref IV. Major PCR products were gel-purified, cloned into the pCR2.1- or pCR4-TOPO® vector (Invitrogen, Carlsbad, CA) and sequenced.

Construction of reporter plasmids

The fragments with corresponding overhangs were ligated into the *XhoI* or *KpnI* (promoter variants) and *SalI* (enhncer) restriction sites of pGL3 or pGL4. Different combinations of inserts in both sense and anti-sense orientation were prepared in order to confirm the specificity of their regulatory function. Serial deletions of promoter-containing constructs were created by site-directed mutagenesis using the QuickChange XL site-directed mutagenesis kit (Stratagene, La Jolla, USA). All resulting constructs were verified by sequencing. For complete list of constructs and other details see Ref II, Ref III and Ref IV.

Transient transfections and reporter gene assays

Transfection using Tfx-20 and FuGene HD reagents (Promega, Mannheim, Germany) were performed in Opti-MEM media (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. The pGL3- or pGL4-derived experimental constructs were cotransfected with the *Renilla* luciferase containing plasmid pRL-TK in order to normalize reaction yields. Reporter assays were performed using Dual Luciferase Reporter assay (Promega, Mannheim, Germany) according to the manufacturer instructions with few modifications as described in Ref II, Ref III and Ref IV. Emitted light was detected using a Berthold Sirius luminometer (Bundoor, Australia) and Synergy 2 automatic microplate reader (BioTek, Winooski, VT).

Nuclear extract preparation

Nuclear and cytoplasmatic extracts were prepared from Hep-G2 cells using the NE-PER nuclear extraction kit (Thermo Scientific, Waltham, MA) according to the manufacturer's instructions. Extracts were dialyzed according to Dignam, et al. (1983). Protein concentrations were determined using the Pierce BCA Protein Assay Kit (Thermo Scientific, Waltham, MA).

DNase I footprinting

DNase I footprinting was used to analyse the interaction of nuclear proteins with the regulatory regions. The method, based on the separation of fluorescently labeled products of DNase I cleavage using capillary electrophoresis (Wilson DO, et al., 2001; Zianni M, et al., 2006), was modified as described in Ref III. Cleaved DNA fragments were purified by phenol-chloroform extraction followed by ethanol precipitation and separated by capillary electrophoresis on the Applied Biosystems 3130 Genetic Analyzer.

Electromobility shift assays

Double-stranded biotin-labeled probes were prepared by standard PCR or by hybridization of complementary oligonucleotides. The binding reactions with nuclear proteins were incubated for 20 min at room temperature, separated using non-denaturing poly-acrylamide gel electrophoresis and electro-blotted on the Amersham Hybond-N nylon membrane (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The membranes were fixed by baking or UV crosslinking, processed using the Chemiluminiscent Nucleic Acid Detection Module (Thermo Scientific, Rockford, IL) according to the manufacturer´s instructions and analyzed with the FUJI LAS 3000 system (Fujifilm, Tokyo, Japan). A non-labeled DNA probe derived from the consensus HNF-4α recognition sequence was used in the competitor assays; A goat anti-HNF-4α polyclonal antibody (Santa Cruz, Dallas, TX) was used in supershift assays. For detail list of DNA probes, buffer compositions and reaction conditions see Ref III.

Real-time PCR

In case of *GBA*, utilization of the transcription initiation sites was studied by quantitative RT-PCR. Primers and TaqMan probes were designed using Primer Express 3.0 software (Applied Biosystems, Foster City, CA). All reactions were performed using StepOne™ Real-time PCR system (Applied Biosystems). The cDNA was obtained by reverse transcription of Human Placenta Total RNA or FirstChoice® Human Total RNA Survey Panel (both from Ambion) using a High Capacity RNA-to-cDNA Kit (Applied Biosystems). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β-actin (ACTB) were used as endogenous controls for relative quantification. The Ct values were transformed to relative quantities using delta Ct method and the expression stability was evaluated using geNorm (version 3.5) and NormFinder software packages.

Analysis of DNA methylation

The X-inactivation ratio in *OTC* patients was determined by analysis of methylation status of the human androgen-receptor locus (HUMARA) as described elsewere (Dobrovolny R, et al., 2005). The methylation status of the CpG island overlapping with P2 promoter of *GBA* gene was studied using bisulphite sequencing. Bisulphite-modified genomic DNA of three control individuals served as a template for amplification of the CpG-rich sequence in the vicinity of P2 promoter in three fragments. Gel-purified products were sequenced. For more details see Ref I and Ref IV.

In-silico predictions and other bio-informatic analyses

Human *OTC* promoter and enhancer regions were predicted on the basis of orthologous regions in rat and mouse. Homology was assessed using the

NCBI Blast tool http://blast.ncbi.nlm.nih.gov/Blast.cgi) and

UCSC genome browser Multiz Alignment&Conservation utility (http://genome.ucsc.edu).

Prediction of transcription factor binding sites in was performed with

MatInspector (http://www.genomatix.de/shop/evaluation.html),

Match (http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi),

TFSEARCH (http://www.cbrc.jp/research/db/TFSEARCH.html) and

Alibaba2 (http://www.gene-regulation.com/pub/programs.html#alibaba2).

All reference DNA and transcript sequences were obtained from

NCBI (http://www.ncbi.nlm.nih.gov/refseq/)

The *GBA* ESTs containing sequence homologous to exons −2 and −1 were identified using BLAST and UCSC Genome Browser. The multiple alignment of exon −2 and P2 was modified from 44-way alignment of assembled genomes

available from Galaxy (http://main.g2.bx.psu.edu/).

Statistical analysis

Descriptive statistics, parametric and non-parametric tests of statistical hypotheses, analyses of variance (ANOVA) and multiple comparisons were performed using SigmaPlot software (Systat Software, Chicago, IL). All values are given as mean ± SD of replicates. The results were evaluated with Kruskal-Wallis's one way ANOVA on ranks, multiple comparisons were set up using the Dunn's method. Differences at $p \leq 0.01$ were considered as statistically significant.

4. Results

4.1. Ornithine carbamoyltransferase deficiency: molecular characterization of 29 families

4.1.1. Commentary

The OTCD is the most common inherited defect of the urea cycle. Severity of clinical symptoms is highly variable, apparently depends on genotype and in heterozygous females also on X-inactivation pattern in hepatocytes. Although males with hemizygous null mutations present with severe form of OTCD in neonatal period, males with hypomorphic mutations develop first symptoms at any age from the neonatal period to late adulthood, or rarely remain asymptomatic during their entire lifetimes (Brusilow SW and Horwich AL, 2004). About 20% of female carriers develop clinical symptoms similar to those in males with late onset variants (Tuchman M, et al., 2002).

This report presents results of molecular analysis in a cohort of OTCD patients containing 11 males with severe neonatal form, 17 males with late onset and 9 symptomatic female heterozygotes coming from 29 families. We focused on novel disease alleles, precise breakpoint mapping of large rearrangements and genotype–phenotype correlation.

We have identified 25 different mutations, 14 of which were novel. Three novel missense mutations (p.Ala102Pro, p.Pro158Ser, p.Lys210Glu) and a novel deletion of the Leu43 are not directly involved either in the enzyme active site or in the intersubunit interactions; however, the mutations include conserved residues involved in intramolecular interaction network essential for the function of the enzyme.

In one female and two male patients we identified 3 novel large deletions. A 444 kb deletion g.1276_445398delins63 encompassed the entire locus containing 15 out of 19 exons of *RPGR* and the entire *OTC* and *TSPAN7* genes was accompaigned with 63 nucleotide insertion. The other two large deletions affected only parts of *OTC*: the 10 kb-deletion g.42940_52876delinsT encompassing exons 5 and 6 and the 24.5 kb-deletion g.57467_82002 delinsCCT removing *OTC* exons 9 and 10. All three rearrangements have probably been initiated by double strand breaks at recombination-promoting motifs with subsequent non-homologous end-joining repair.

Moreover, we presented a manifesting heterozygote carrying a hypomorphic mutation p.Arg129His in combination with unfavorably skewed X-inactivation in three peripheral tissues.

To our best knowledge, no further similar study of *OTC* mutations in a patient cohort has been published to date.

Short Report

Ornithine carbamoyltransferase deficiency: molecular characterization of 29 families

Ornithine carbamoyltransferase deficiency is the most common inherited defect of the urea cycle. We examined 28 male and 9 female patients from 29 families and identified 25 distinct mutations in *OTC*, 14 of which were novel. Three novel missense mutations (p.Ala102Pro, p.Pro158Ser, p.Lys210Glu) and a novel deletion of the Leu43 are not directly involved either in the enzyme active site or in the intersubunit interactions; however, the mutations include conserved residues involved in intramolecular interaction network essential for the function of the enzyme. Three novel large deletions – a 444 kb deletion affecting *RPGR*, *OTC* and *TSPAN7*, a 10 kb-deletion encompassing *OTC* exons 5 and 6 and a 24.5 kb-deletion encompassing *OTC* exons 9 and 10 – have probably been initiated by double strand breaks at recombination-promoting motifs with subsequent non-homologous end joining repair. Finally, we present a manifesting heterozygote carrying a hypomorphic mutation p.Arg129His in combination with unfavorably skewed X-inactivation in three peripheral tissues.

Conflict of interest

The authors declare that they have no conflict of interest.

G Storkanova[a,†], H Vlaskova[a,†], N Chuzhanova[b], J Zeman[c], V Stranecky[a], F Majer[a], K Peskova[a], O Luksan[d], M Jirsa[d], M Hrebicek[a] and L Dvorakova[a]

[a]Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague, and General University Hospital in Prague, Czech Republic, [b]School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK, [c]Department of Paediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University in Prague, and General University Hospital in Prague, Czech Republic, and [d]Laboratory of Experimental Hepatology, Institute for Clinical and Experimental Medicine (IKEM), Prague, Czech Republic

[†]These authors contributed equally to the manuscript.

Ornithine carbamoyltransferase (OTC, OMIM *300 461) is a mitochondrial enzyme which plays a key role in the elimination of the toxic ammonia in protein catabolism. Human *OTC* gene, located on Xp21.1, is expressed principally in the liver and is subject to X-inactivation. Ornithine carbamoyltransferase deficiency (OTCD; OMIM #311250) is the most common inherited defect of the urea cycle. Severity of clinical symptoms of OTCD is highly variable, apparently depends on genotype, and in heterozygous females also on X-inactivation pattern in hepatocytes. Although males with hemizygous null mutations

1

present with severe form of OTCD in neonatal period, males with hypomorphic mutations develop first symptoms at any age, from the neonatal period to late adulthood, or rarely remain asymptomatic during their entire lifetimes (1). About 20% of female carriers develop clinical symptoms similar to those in males with late onset variants (2).

The aim of this report is a detailed molecular characterization of 29 OTCD families including description of novel disease alleles, precise breakpoint mapping of large rearrangements and genotype–phenotype correlation.

Patients and methods

Thirty-five patients from 27 families were Czechs and Slovaks, one was Hungarian and one was Japanese (Table 1). All were born at term with uneventful post-natal adaptation. Twelve patients with neonatal OTCD onset including one heterozygous female (patient 33) presented in the first week of life with poor feeding, tachypnea, apneic spells, progressive hypotonia or tonic–clonic seizures and multiorgan failure due to severe hyperammonemia. Blood levels of glutamine and glutamic acid and urinary excretion of orotic acid were markedly elevated. All boys (Patients 1–11) died in early neonatal period.

Twenty-three symptomatic patients with late onset including eight heterozygous females presented between 8 months and 20 years of age with headaches, abdominal pain, vomiting, diarrhea and progressive apathy due to hyperammonemia. Three males with late onset (patients 12, 16 and 26) died due to hyperammonemia at the age of 1, 13 and 6 years, respectively.

Genotyping in the families revealed 33 clinically asymptomatic heterozygotes. All of them were healthy with no history of clinical symptoms associated with OTCD. Of the 15 examined heterozygous females, all but one had urinary excretion of orotic acid within the reference range. Markedly increased urinary orotic acid was observed in 9 of these 15 heterozygotes during the test with allopurinol.

The study was approved by the institutional review board of the General University Hospital in Prague and conducted in agreement with institutional guidelines. Written informed consent for DNA investigation was obtained from the study participants.

Mutation analysis was performed as described (3). Potential impact of mutations on OTC function and/or folding was assessed using multiple alignments of orthologous protein sequences and human OTC and structural data from Protein Data Bank (1C9Y and available orthologs).

In male patients, the approximate extent of the deletions was assessed by inspection of presence/absence of PCR products. In females, the deletions were determined using the SALSA multiplex ligation probe amplification (MLPA) KIT P079 OTC (MRC-Holland, Amsterdam, the Netherlands) and the Affymetrix Human SNP 6.0 array (Santa Clara, CA). Data analysis was performed within the Affymetrix Genotyping

Console 3.0.2 using CEU HapMap samples as the reference set. Breakpoints were located by sequencing of PCR products overlapping the deletion boundaries.

Sequence spanning 38,211,736 – 38,300,703 bp region on chromosome X (GRCh37) and including *OTC* was scanned for motifs CCTCCCT, CCTCCTT, CCTCCCTT, CCCCACCCC, CCNCCNTNNCCNC, GGNGGNAGGG and their complements known as being associated with recombination hotspots (4, 5). In addition, direct, inverted and mirror repeats capable of non-B DNA structure formation implicated in double strand breaks (DSBs) (6) were sought using complexity analysis (7). The X-inactivation ratio was determined by analysis of methylation status of the human androgen-receptor locus (HUMARA) as described earlier (8).

Results and discussion

Eleven male patients with severe neonatal form and 17 males with late onset coming from 22 families are reported in this study. Two out of nine symptomatic heterozygotes were siblings of male probands, whereas other seven were index patients in their families.

In 37 patients we have identified 25 different mutations, 14 of them being novel (Table 2). A novel 3 bp-deletion resulting in elimination of a leucine residue (p.Leu43del) was identified in patient 2. The leucine 43 residue (underlined) is situated within a stretch -Gly-Arg-Asp-Leu-Leu-Thr-Leu- fully conserved in eight vertebrate orthologs. Its deletion may abolish conserved hydrophobic interactions of Leu43 with helix 4 and impair folding of the enzyme.

None of the three novel missense mutations p.Ala102Pro, p.Pro158Ser and p.Lys210Glu was indicated to affect splicing by the SKIPPY program (http://research.nhgri.nih.gov/skippy/) (9). Although not directly involved either in the enzyme active site or in the intersubunit interactions (10), the mutated residues show critical intramolecular interactions with other spatially neighboring highly conserved residues. The mutation p.Ala102Pro possibly disrupts structurally conserved helix 2. Both Pro158 and Lys210 are at the surface of the active trimer. Pro158 is located in an edge strand of a beta sheet. The side chain of Lys210 shows polar interactions involving backbone carbonyls of highly conserved residues in helices H1a, H6a and a side chain of Tyr236 in H7 (10).

Sequencing of cDNA obtained from the liver specimens of Patients 2, 5 and 16 did not reveal any abnormally spliced transcripts caused by a deletion c.126_128delTCT or substitutions c.304G>C and c.829C>T, respectively. On the other hand, analysis of cDNA of patient 26 showed that the transition c.867G>A, which did not lead to amino acid change (p.Lys289Lys), abolished the natural splicing site and resulted in the usage of an alternative exonic splicing site and an in-frame deletion of 12 nucleotides in the transcript.

Large deletions were identified in two males and one female. In patient 1 a large deletion of 444 kb

2

Table 1. Clinical, biochemical and molecular characterization of 37 patients with OTCD

| Patient | Gender, family member | Age of onset | Biochemical data at the time of clinical diagnosis[a] | | | Molecular findings | |
			Plasma ammonia (µmol/l)	Urinary orotic acid (mmol/mol creatinine)	Serum glutamate + glutamine (µmol/l)	Mutation (predicted effect on protein)[b]	Nature of mutation
1	Male	Neonatal	1700	104	3678	g.1276_445398 delins63 Contiguous deletion incl. *OTC* gene	Inherited
2	Male	Neonatal	5000	397	3276	p.Leu43del	*De novo*
3	Male	Neonatal	477	449	5959	p.Tyr73*	*De novo*
4	Male	Neonatal	1200	97	6316	p.Arg92*	Inherited
5	Male	Neonatal	1355	15	6131	p.Ala102Pro	Inherited
6	Male	Neonatal	1200	1646	8242	p.Glu154Alafs*18	Inherited
7	Male	Neonatal	1500	351	Elevated	p.Ser164*	Inherited
8	Male	Neonatal	500	614	N.D.	Exons 5/6 deleted (p.Val129Glyfs*134)	Inherited
9	Male	Neonatal	2800	700	N.D.	p.Tyr183Cys	Inherited
10	Male	Neonatal	1100	228	8079	p.Ser207Asn	*De novo*
11	Male	Neonatal	1096	207	2921	p.Ser340Pro	*De novo*
12	Male	8 months	800	99	1294	p.Arg129His (donor splicing site affected)	Inherited
13	Male	1 year	260	1186	Elevated	p.Pro158Ser	*De novo*
14	Male	18 years	N.D.	17	524	p.Arg277Trp	Inherited
15	Male, brother of P14	20 years	N.D.	30	1086	p.Arg277Trp	Inherited
16	Male	13 years	400	928	657	p.Arg277Trp	Inherited
17	Male, cousin of P16	9 years	29	2	1209	p.Arg277Trp	Inherited
18	Male, cousin of P16	8 years	298	114	1740	p.Arg277Trp	Inherited
19	Male, cousin of P16	asymptomatic at 5 years	128	2	677	p.Arg277Trp	Inherited
20	Male, grandfather of P16	asymptomatic at 62 years	32	3	553	p.Arg277Trp	Inherited
21	Male	6 years	N.D.	332	1619	p.Arg277Trp	Inherited
22	Male	4 years	33	69	1202	p.Arg277Trp	Inherited
23	Male	9 years	197	59	1445	p.Arg277Trp	Inherited
24	Male, brother of P23	9 years	100	355	1317	p.Arg277Trp	Inherited
25	Male	2 years	148	135	1108	p.Arg277Gln	Inherited
26	Male	1 year	4500	N.D.	N.D.	p.Val286_Lys289del	*De novo*
27	Male	6 months	396	538	1142 (2351 in cerebrospinal fluid)	p.Glu310del	Inherited
28	Male	2 years	499	855	2717	p.*355Cysext*14	Inherited
29	Female, sister of P12	2 years	65	480	1350	p.Arg129His (donor splicing site affected)	Inherited
30	Female, sister of P7	4 years	39	7	1392	p.Ser164*	Inherited
31	Female	6 years	N.D.	27	1021	p.Gly195Arg	Inherited
32	Female	8 months	364	168	N.D.	p.Lys210Glu	*De novo*
33	Female	Neonatal	830	319	2776	p.Lys210Gln	*De novo*
34	Female	6 years	163	383	Elevated	c.717+1G>T (donor splicing site affected)	*De novo*
35	Female	2 years	489	1681	Elevated	p.Gln285*	Inherited
36	Female	3 years	198	256	1672	Exons 9/10 deleted (p.Thr290_Phe354del)	*De novo*
37	Female	2 years	76	53	1104	c.1043delA (p.Gln348Argfs*47)	*De novo*

[a]Reference values: Plasma ammonia: neonates <90 µmol/l, other <60 µmol/l. Urinary orotic acid: 0–1 year <6.6 mmol/mol creatinine, 1–10 years <3.5 mmol/mol creatinine, over 10 years <2.4 mmol/mol creatinine. Serum glutamate + glutamine: 0–1 month 200–1200 µmol/l, 1 month–1 year 200–1100 µmol/l, 1 year–18 years 200–900 µmol/l, over 18 years 200–800 µmol/l.
[b]Characterization of the mutations is given in Table 2.

3

Table 2. Characterization of the *OTC* gene mutations found in the reported cohort of patients

Location (exon)	Mutation[a]	Probability of deleterious mutation			Disease presentation reported in this study	Disease presentation reported elsewhere	Conservation of the changed amino acid residue[b]
		MutPred	Mutation taster	Poly Phen 2			
1–10	g.1276_445398 delins63 / Contiguous deletion includes entire *OTC* gene	–	–	–	Neonatal – male	Neonatal (23), breakpoints not characterized	
2	c.126_128 delTCT / p.Leu43del	–	1	–	Neonatal – male		Conserved in 10 species, exon missing in *Gallus gallus, Xenopus tropicalis* and *Gasterosteus aculeatus*
3	c.219T>G / p.Tyr73*	–	–	–	Neonatal – male		
3	c.247C>T / p.Arg92*	–	–	–	Neonatal – male	Neonatal (24)	
4	c.304G>C[c] / p.Ala102Pro	0.715	0.999	0.988	Neonatal – male	Neonatal (25), reported mutation p.Ala102Glu	Conserved in nine species, exon missing in *Monodelphis domestica, G. gallus, X. tropicalis* and *G. aculeatus*
4	c.386G>A / p.Arg129His (+donor splicing site affected)	0.943	0.999 donor motif lost	0.973	Late – male/ Manifesting heterozygote	Late (26)	Conserved in nine species, exon missing in *M. domestica, G. gallus, X. tropicalis* and *G. aculeatus*
5 and 6	g.42940_52876 delinsT (deleted exons 5 and 6) / p.Val129Glyfs*134	–	–	–	Neonatal – male		
5	c.461_471del / p.Glu154Alafs*18	–	–	–	Neonatal – male		
5	c.472C>T[c] / p.Pro158Ser	0.954	0.999	1.000	Late -male		Conserved in 13 species
5	c.491C>A / p.Ser164*	–	–	–	Neonatal – male/ Manifesting heterozygote		
6	c.548A>G / p.Tyr183Cys	0.777	0.999	1.000	Neonatal – male	Neonatal (27)	Conserved in 13 species
6	c.583G>C / p.Gly195Arg	0.986	0.999	1.000	Manifesting heterozygote	Manifesting heterozygote, diagnosed at 1 year of age (28)	Conserved in 13 species
6	c.620G>A / p.Ser207Asn	0.889	0.869	0.868	Neonatal – male	Neonatal (29)	Conserved in 12 species, Thr in *G. aculeatus*
6	c.628A>G[c] / p.Lys210Glu	0.855	0.999	0.990	Manifesting heterozygote		Conserved in 13 species

Table 2. Continued

Location (exon)	Mutation[a]	Probability of deleterious mutation			Disease presentation reported in this study	Disease presentation reported elsewhere	Conservation of the changed amino acid residue[c]
		MutPred	Mutation taster	Poly Phen 2			
6	c.628A>C p.Lys210Gln	0.940	0.998	0.730	Neonatal – female: described in Ref. (17)		Conserved in 13 species
Intron 7	c.717+1G>T (IVS7+1G>T) donor splicing site affected	–	–	–	Manifesting heterozygote	Manifesting heterozygote (2)	
8	c.829C>T p.Arg277Trp	0.892	0.999	1.000	Late – male	Late (30)	Conserved in 13 species
8	c.830G>A p.Arg277Gln	0.924	0.999	1.000	Late – male	Late (26)	Conserved in 13 species
8	c.853C>T p.Gln285*	–	–	–	Manifesting heterozygote		
8	c.867G>A r.856_867del GTTACAATGAAG p.Val286_Lys289del (donor splicing site affected)	–	Donor motif lost	–	Late – male		
9	c.929_931 delAAG p.Glu310del	–	–	–	Late – male	Late (26)	Conserved in 13 species
9 and 10	g.57467_82002 delinsCCT (deleted exons 9 and 10) p.Thr290_Phe354del	–	–	–	Manifesting heterozygote		
10	c.1018T>C p.Ser340Pro	0.887	0.982	0.997	Neonatal male	Manifesting heterozygote (31)	Conserved in 11 species, Ala in *X. tropicalis*, exon missing in *Loxodonta africana*
10	c.1043delA p.Gln348Argfs*47	–	–	–	Manifesting heterozygote		
10	c.1065A>T p.*355Cysext*14	–	–	–	Late – male		

[a]Reference sequence: NC_000023.10 (CON 10-JUN-2009), REGION: 38211736–38280703; large deletion in patient 1: NC_000023.10 (CON 10-JUN-2009), REGION: 38139000–38585000; and large deletion in patient 35: NC_000023.10 (CON 10-JUN-2009), REGION: 38211736–38295000.

[b]Conservation compared in *Homo sapiens*, Rhesus macaque (*Macaca mulatta*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), dog (*Canis lupus*), cattle (*Bos taurus*), elephant (*Loxodonta africana*), opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), lizard (*Anolis carolinensis*), frog (*Xenopus tropicalis*), stickleback (*Gasterosteus aculeatus*). Data from http://genome.ucsc.edu/ and http://www.ncbi.nlm.nih.gov/gene/5009.

[c]Novel missense mutations. None of them has been found in 300 control alleles. *Bse*LI (Fermentas, Burlington, ON, Canada) was used for restriction analysis of c.472C>T and c.628A>G. Direct sequencing was used for screening of c.472C>T and c.628A>G.
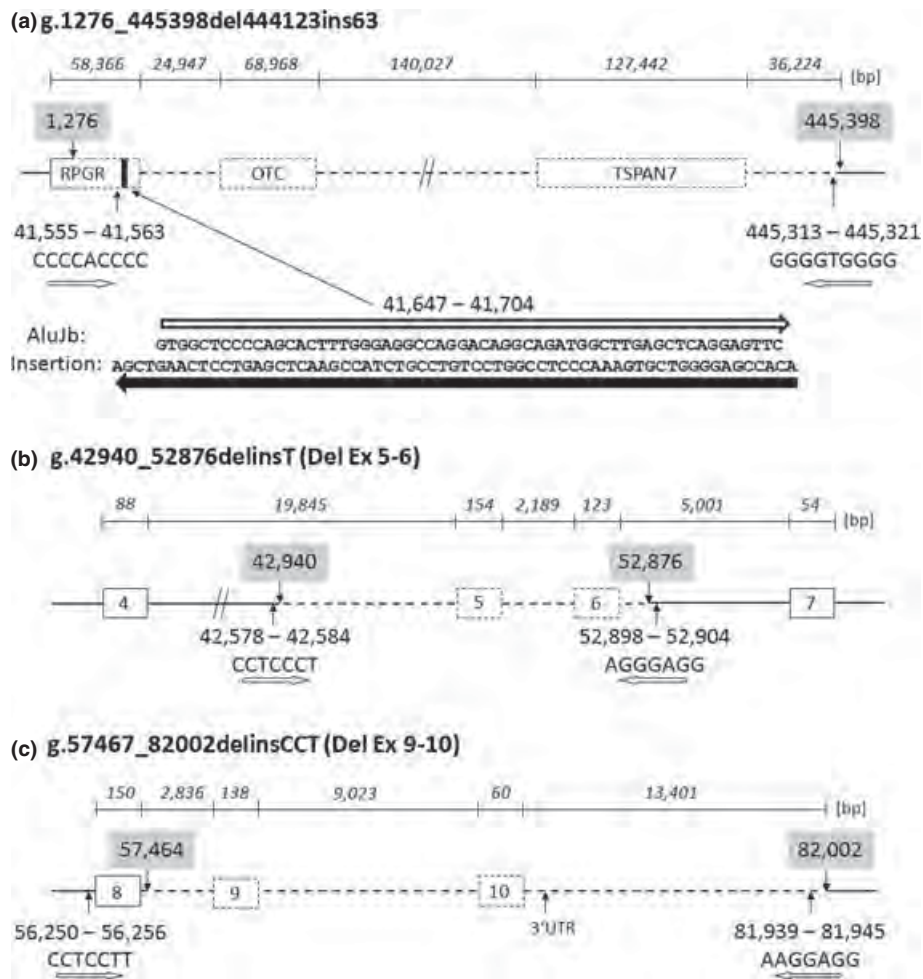
5

Fig. 1. Genomic region encompassing large deletions in *OTC*. The deleted segments are shown by broken line. The positions of deletion breakpoints are shaded. Positions of recombinogenic motifs are indicated by vertical arrows. Orientation of fragments is indicated by horizontal arrows. (**a**) A deletion identified in patient 1: g.1276_445398delins63 (Reference sequence: NC_000023.10, REGION: 38139000–38585000). A 63 bp insertion and *AluJb* from deleted segment of the *RPGR* gene show reverse orientation. (**b**) A deletion identified in patient 8: g.42940_52876delinsT (Reference sequence: NC_000023.10, REGION: 38211736–38280703). (**c**) A deletion identified in patient 36: g.57467_82002delinsCCT (Reference sequence: NC_000023.10, REGION: 38211736–38295000).

encompassed 15 out of 19 exons of *RPGR* and the entire *OTC* and *TSPAN7* genes. An insertion of 63 nucleotides was identified between the breakpoints. A part of the inserted fragment was found to be identical with the region g.41,647–41,704 that coincides with the *AluJb* retroelement (Fig. 1a). This complex mutation was likely originated through a two-step insertion/deletion process (11) with the insertion occurring first followed by a deletion. The insertion may have occured through misalignment between two inexact inverted repeats that occur upstream of the 5′ breakpoint and the intermediate breakpoint and consist of *AluJb* repeats in opposite orientation with subsequent strand switching via interchromosomal serial replication slippage (12). Recombinogenic motif CCCCACCCC and its complement found in close proximity to the intermediate breakpoint (position g.41,555) and the 3′ breakpoint (position g.445,313), respectively, may have been responsible for DSBs and non-homologous end joining (NHEJ) resulting in deletion.

The other two large deletions affected only parts of *OTC*: the 10 kb-deletion encompassing exons 5 and 6 was identified in patient 8, whereas patient 36 was heterozygous for the 24.5 kb-deletion removing exons 9 and 10. Unlike the large deletions described in the *OTC* gene by Engel (13) and Quental (14), there are no repetitive elements in the vicinity of the breakpoints of the deletions in patients 8 and 36. Recombinogenic motifs CCTCCCT and its complement AGGGAGG were found in close proximity to the breakpoints of mutation in patient 8 (Fig. 1b). A similar recombinogenic motif CCTCCTT and its complement were also found in close proximity to the breakpoints in patient 36 (Fig. 1c). The deletions most likely have been mediated by DSBs and subsequent NHEJ (15). Small insertions are common for this type of recombination; they were observed in both cases. To our knowledge, this mechanism has not been previously described in *OTC* gene deletions.

6

In accordance with the published data (16), we observed that hypomorphic mutations may cause a broad spectrum of phenotypes even among the male members of the same family: patient 16 developed the symptoms of the disease and died at the age of 13 years, whereas his grandfather (patient 20) remained asymptomatic till the age of 62 years.

Heterozygous females have even a wider spectrum of disease manifestation due to variable X-inactivation pattern in the liver. Patient 33 has been symptomatic from the third day of her life (17). Patient 31 underwent successful liver transplantation at the age of 14 years indicated due to 2–3 hyperammonemic episodes per year during puberty. During the 2 years after transplantation she did not manifest any symptoms.

Our data, as well as the previously published findings (18), showed that symptomatic heterozygous OTCD females carry mutations almost exclusively associating with the severe phenotype, whereas heterozygous females carrying less deleterious mutations remain generally asymptomatic. Only two unrelated heterozygotes carrying a hypomorphic mutation were reported to manifest the symptoms of OTCD (19, 20). In our study, we identified a third case of a symptomatic carrier of a hypomorphic mutation (patient 29), who presented with episodes of vomiting and nausea.

It is usually accepted that X-inactivation in peripheral blood cells may not fully reflect the pattern in the liver. This concept is based on analysis of a single OTCD patient (21). By contrast, Bittel (22) and Minks (unpublished results) evaluated X-inactivation patterns in multiple tissues, including the liver and blood cells from 31 females unaffected by OTCD and found that the differences in X-inactivation skewing in tissues from the same individual, while present, are not prominent.

To our knowledge, X-inactivation pattern has not been analyzed in any manifesting carrier with a hypomorphic mutation. Patient 29 had skewed X-inactivation in favor of the mutant allele (82:18 in peripheral leukocytes, 73:27 in urinary cells and 80:20 in salivary cells). Interestingly, her asymptomatic heterozygous mother showed the opposite pattern – slightly preferential usage of the wild type allele (22:78 in peripheral leukocytes, 16:84 in urinary cells and 23:77 in salivary cells). On the basis of these findings we hypothesize that manifestation of symptoms caused by hypomorphic mutation might depend on highly skewed X-inactivation in the liver, which is reflected by X-inactivation status in peripheral tissues.

In conclusion, we identified three novel missense mutations affecting residues of highly conserved OTC architecture, described three large deletions and suggested their mutational mechanisms, and identified a manifesting heterozygote for a hypomorphic mutation with unfavorable X-inactivation skewing in three peripheral tissues.

Acknowledgements

References

1. Brusilow SW, Horwich AL. Urea cycle enzymes. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. The online metabolic and molecular bases of inherited disease. New York, NY: McGraw Hill, Inc, 2004.
2. Tuchman M, Jaleel N, Morizono H, Sheehy L, Lynch MG. Mutations and polymorphisms in the human ornithine transcarbamylase gene. Hum Mutat 2002: 19: 93–107.
3. Luksan O, Jirsa M, Eberova J et al. Disruption of OTC promoter–enhancer interaction in a patient with symptoms of ornithine carbamoyltransferase deficiency. Hum Mutat 2010: 31: E1294–E1303.
4. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. Science 2005: 310: 321–324.
5. Zheng J, Khil PP, Camerini-Otero RD, Przytycka TM. Detecting sequence polymorphisms associated with meiotic recombination hotspots in the human genome. Genome Biol 2010: 11: R103.
6. Wells RD. Non-B DNA conformations, mutagenesis and disease. Trends Biochem Sci 2007: 32: 271–278.
7. Gusev VD, Nemytikova LA, Chuzhanova NA. On the complexity measures of genetic sequences. Bioinformatics 1999: 15: 994–999.
8. Dobrovolny R, Dvorakova L, Ledvinova J et al. Relationship between X-inactivation and clinical involvement in Fabry heterozygotes. Eleven novel mutations in the alpha-galactosidase A gene in the Czech and Slovak population. J Mol Med 2005: 83: 647–654.
9. Woolfe A, Mullikin JC, Elnitski L. Genomic features defining exonic variants that modulate splicing. Genome Biol 2010: 11: R20.
10. Shi D, Morizono H, Yu X, Tong L, Allewell NM, Tuchman M. Human ornithine transcarbamylase: crystallographic insights into substrate recognition and conformational changes. Biochem J 2001: 354: 501–509.
11. Chuzhanova NA, Anassis EJ, Ball EV et al. Meta-analysis of indels causing human genetic disease: mechanisms of mutagenesis and the role of local DNA sequence complexity. Hum Mutat 2003: 21: 28–44.
12. Chen JM, Chuzhanova N, Stenson PD, Krawczak M, Cooper DN. Intrachromosomal serial replication slippage in trans gives rise to diverse genomic rearrangements involving inversions. Hum Mutat 2005: 26: 362–373.
13. Engel K, Nuoffer JM, Muhlhausen C et al. Analysis of mRNA transcripts improves the success rate of molecular genetic testing in OTC deficiency. Mol Genet Metab 2008: 94: 292–297.
14. Quental R, Azevedo L, Rubio V, Diogo L, Amorim A. Molecular mechanisms underlying large genomic deletions in ornithine transcarbamylase (OTC) gene. Clin Genet 2009: 75: 457–464.
15. Lieber MR, Ma Y, Pannicke U, Schwarz K. Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol 2003: 4: 712–720.
16. Matsuda I, Matsuura T, Nishiyori A et al. Phenotypic variability in male patients carrying the mutant ornithine transcarbamylase (OTC) allele, Arg40His, ranging from a child with an unfavourable prognosis to an asymptomatic older adult. J Med Genet 1996: 33: 645–648.
17. Valik D, Sedova Z, Starha J, Zeman J, Hruba E, Dvorakova L. Acute hyperammonaemic encephalopathy in a female newborn caused by a novel, de novo mutation in the ornithine transcarbamylase gene. Acta Paediatr 2004: 93: 710–711.
18. McCullough BA, Yudkoff M, Batshaw ML, Wilson JM, Raper SE, Tuchman M. Genotype spectrum of ornithine transcarbamylase deficiency: correlation with the clinical and biochemical phenotype. Am J Med Genet 2000: 93: 313–319.
19. Pinner JR, Freckmann ML, Kirk EP, Yoshino M. Female heterozygotes for the hypomorphic R40H mutation can have ornithine transcarbamylase deficiency and present in early adolescence: a case report and review of the literature. J Med Case Reports 2010: 4: 361.

7

20. Numata S, Harada E, Maeno Y et al. Paternal transmission and slow elimination of mutant alleles associated with late-onset ornithine transcarbamylase deficiency in male patients. J Hum Genet 2008: 53: 10–17.

21. Yorifuji T, Muroi J, Uematsu A et al. X-inactivation pattern in the liver of a manifesting female with ornithine transcarbamylase (OTC) deficiency. Clin Genet 1998: 54: 349–353.

22. Bittel DC, Theodoro MF, Kibiryeva N, Fischer W, Talebizadeh Z, Butler MG. Comparison of X-chromosome inactivation patterns in multiple tissues from human females. J Med Genet 2008: 45: 309–313.

23. Arranz JA, Madrigal I, Riudor E, Armengol L, Mila M. Complete deletion of ornithine transcarbamylase gene confirmed by CGH array of X chromosome. J Inherit Metab Dis 2007: 30: 813.

24. Grompe M, Muzny DM, Caskey CT. Scanning detection of mutations in human ornithine transcarbamoylase by chemical mismatch cleavage. Proc Natl Acad Sci USA 1989: 86: 5888–5892.

25. Tuchman M, Morizono H, Rajagopal BS, Plante RJ, Allewell NM. Identification of 'private' mutations in patients with ornithine transcarbamylase deficiency. J Inherit Metab Dis 1997: 20: 525–527.

26. Tuchman M, Plante RJ, McCann MT, Qureshi AA. Seven new mutations in the human ornithine transcarbamylase gene. Hum Mutat 1994: 4: 57–60

27. Reish O, Plante RJ, Tuchman M. Four new mutations in the ornithine transcarbamylase gene. Biochem Med Metab Biol 1993: 50: 169–175.

28. Kim GH, Choi JH, Lee HH, Park S, Kim SS, Yoo HW. Identification of novel mutations in the human ornithine transcarbamylase (OTC) gene of Korean patients with OTC deficiency and transient expression of the mutant proteins in vitro. Hum Mutat 2006: 27: 1159.

29. Yamaguchi S, Brailey LL, Morizono H, Bale AE, Tuchman M. Mutations and polymorphisms in the human ornithine transcarbamylase (OTC) gene. Hum Mutat 2006: 27: 626–632.

30. Finkelstein JE, Hauser ER, Leonard CO, Brusilow SW. Late onset ornitine transcarbamylase deficiency in male patients. J Pediatr 1990: 117: 897–902.

31. Oppliger Leibundgut E, Liechti-Gallati S, Colombo JP, Wermuth B. Ornithine transcarbamylase deficiency: ten new mutations and high proportion of de novo mutations in heterozygous females. Hum Mutat 1997: 9: 409–411.

8

4.2. Disruption of OTC promoter-enhancer interaction in a patient with symptoms of ornithine carbamoyltransferase deficiency

4.2.1. Commentary

The regulatory regions of *Otc* have been studied in mice and rats but not in humans. A 0.8 kb promoter sequence located in the 5'-flanking region has been shown to modulate *Otc* transcription in mouse (Veres G, et al., 1986) and a homologous 1.3 kb region was described in rat (Murakami T, et al., 1989). The same group identified later a 230 bp distal enhancer located 11 kb upstream of the initiation codon (Murakami T, et al., 1990); however no such region was described in human *OTC*.

The following article deals with an OTCD family carrying a variation in 5'-flanking region which was excluded from the previous cohort and published separately. In this study we identified and delimited the promoter and yet unknown enhancer of human *OTC*. Moreover, we explored the functional impact of the novel unique variation in the promoter found in a two generation family with mild clinical and biochemical signs of OTCD.

We determined 3 alternative transcription start sites of the *OTC* gene located at positions c.-95, c.-119 and c.-169 upstream of the translation origin. Based on analogy with rat and mouse we predicted and examined the promoter region and identified a distal enhancer located 9 kb upstream of the translation origin. Using the dual luciferase reporter assay we found a significant increase in transcriptional activity of the promoter when compared with the control plasmid. Moreover, three times higher expression of the reporter gene was observed when the enhancer was subcloned into the plasmid containing the promoter.

In a female patient with mild phenotype of OTCD, we evaluated the influence of single nucleotide substitution c.-366A>G in the promoter. No statistically significant difference in the transcriptional activity was observed between wildtype and mutated promoter, however, a 50% difference in the reporter gene expression was found when promoter interacted with the subcloned enhancer. This finding indicated pathogenicity of the c.-366A>G variation.

Our results show that expression of OTC is controlled by a weak promoter and a strong enhancer. The promoter – enhancer interaction contributes to tissue specific expression of OTC in the liver. Single nucleotide substitutions in the regulatory regions such as the c.-366A>G variation may cause OTC deficiency.

As mentioned earlier, mutation analysis of coding regions in patients with disruption of UCD-related genes often fails. Special attention is therefore paid to the role of regulatory regions and other factors affecting expression control of these genes. About two dozens of *OTC*

promoter polymorphisms are listed in the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/snp), 7 of witch are relatively common, but none of them has clinical significance. Just one more mutation, c.-981G>A, associated with Alzheimer disease is reported in the HGMD database, however, this variation is located out of the promoter region. To our best knowledge, we report the first disease-causing mutation in a regulatory region of one of urea cycle genes. A second similar mutation, the enhancer variation c.-3063C>A in *NAGS* encoding the NAG synthase, was reported one year later in an article citing our manuscript (Heibel SK, et al., 2011).

In addition to the regions located upstream of the initiation codon, an interest is paid to 3′ UTR region which is responsible for mRNA processing. Recently, two expressed mRNA isoforms differing in the 3′ UTR and exhibiting tissue-dependent heterogeneity at the cleavage sites have been isolated (Lopes-Marques M, et al., 2012).

Disruption of OTC Promoter-enhancer Interaction in a Patient with Symptoms of Ornithine Carbamoyltransferase Deficiency

Ondrej Luksan[1], Milan Jirsa[1], Jitka Eberova[1], Jakub Minks[2,3], Helena Treslova[2], Michaela Bouckova[2], Gabriela Storkanova[2], Hana Vlaskova[2], Martin Hrebicek[2], and Lenka Dvorakova[2]*

[1] Laboratory of Experimental Hepatology, Institute for Clinical and Experimental Medicine, Prague; [2] Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague, and General University Hospital in Prague, all Czech Republic. [3]Current address: Dept. of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver B.C., Canada

*Correspondence to Lenka Dvorakova, Institute of Inherited Metabolic Disorders (Building D), Ke Karlovu 2, 128 08 Praha 2, Czech Republic. E-mail: lenka.dvorakova@lf1.cuni.cz

ABSTRACT: In a female patient with signs of ornithine carbamoyltransferase deficiency (OTCD), the only variation found was a heterozygous single nucleotide substitution c.-366A>G. Determination of transcription start sites of human *OTC* 95, 119 and 169 bp upstream of the initiation codon located the variation upstream of the 5'-untranslated region. We predicted the human promoter and enhancer elements from homology with rat and mouse, performed function analysis of both regulatory regions and assessed the impact of the promoter variation in functional studies using dual luciferase reporter assay. Our data indicate that: (i) Full transcriptional activity of human *OTC* promoter depends on an upstream enhancer, as do the rodent promoters. (ii) The promoter variation c.-366A>G does not affect the function of the promoter alone but it disrupts the interaction of the promoter with the enhancer. (iii) The promoter – enhancer interaction contribute to tissue specific expression of *OTC* in the liver. We conclude that mutations in the regulatory regions of *OTC* can lead to OTCD and should be included in genetic testing. ©2010 Wiley-Liss, Inc.

KEY WORDS: ornithine carbamoyltransferase, OTC deficiency, promoter, enhancer, expression, dual luciferase reporter assay

INTRODUCTION

Ornithine carbamoyltransferase (OTC, EC 2.1.3.3; MIM# 300461) is a homotrimeric enzyme of the mitochondrial matrix catalyzing synthesis of citrulline from ornithine and carbamoyl phosphate. *OTC* is located on the short arm of the X chromosome (Xp21.1) and expressed principally in the liver (Brusilow and Horwich, 2004).

OTC deficiency (OTCD; MIM# 311250) is the most common inherited urea cycle disorder. Clinical manifestation of OTCD comprises a spectrum of severity ranging from neonatal hyperammonemic coma and death to later onset disease with milder symptoms (episodic metabolic crises, behavioral abnormalities, mental retardation) or an asymptomatic course. The disease is transmitted as a partially dominant X-linked trait, 15 to 20% of carrier females are symptomatic. The phenotype of heterozygous females depends on the type of mutation and at least partially on the inactivation of X chromosome. Some heterozygotes develop symptoms only during catabolic stress such as severe infections or surgery (Brusilow and Horwich, 2004).

Molecular basis of the disease is usually clarified by mutation analysis of the coding sequence of *OTC*. Almost 400 various mutations are listed in the HGMD database (HGMD® professional release 2009.2, https://portal.biobase-international.com/hgmd). Using the current molecular screening methods, mutations are found in about 80% of patients. A subset of non-confirmed patients may carry a causative mutation in regulatory domains (Yamaguchi, et al., 2006). Such mutations, however, have not been reported yet.

The regulatory regions of *OTC* have been studied in mice and rats but not in humans. It has been shown that the 0.8 kb 5' flanking region of the mouse gene (Veres, et al., 1986) and the 1.3 kb region of the rat gene (Murakami, et al., 1989) contain sufficient information to control *Otc* gene expression. However, this 5' flanking region itself was apparently not sufficient to maintain physiological levels of *Otc* expression in the liver. Indeed, a 230 bp positive regulatory element (enhancer) situated 11 kb upstream of the transcription start site (TSS) has been identified in rat (Murakami, et al., 1990). Nishiyori, et al. (1994) showed that the enhancer was activated by at least two liver-selective transcription factors: HNF-4 and C/EBP. This data indicates that the tissue-specific *Otc* expression in rodents is based on the promoter-enhancer interaction in the hepatocyte-specific transcription factor milieu. No such enhancer has been reported in humans to the best of our knowledge.

In this study, we describe the promoter and enhancer of human *OTC* and explore the functional impact of a novel unique variation in the promoter found in a female patient with clinical and biochemical signs of OTCD.

PATIENTS AND METHODS

Case report

The girl was born from a physiological first pregnancy. Both parents of Czech origin were healthy. The patient's early medical history and development were unremarkable. At 9 months of age, she presented with pyelonephritis and ureterocele. Before the ureterocele surgery ALT and AST activities were mildly elevated (ALT 1.14 μkat/l, AST 1.66 μkat/l). Following the surgery, significant elevation of AST and ALT activities in serum was noted (AST 25.5 μkat/l, ALT 17.8 μkat/l) and the liver enzyme activities gradually decreased to 7.71 μkat/l and 5.23 μkat/l within a week. Plasma LDH was elevated (31.6 μkat/l, normal range 2.55 - 7.68 μkat/l) with high activities of LDH4 and LDH5 isoenzymes (15 and 29% of total activity, respectively). Other routine biochemical tests were within normal limits and screening tests for common liver diseases - serology for hepatitis A, B, C, and E, CMV, EBV, HSV and alpha-1 antitrypsin level - were negative. Liver ultrasound was unremarkable.

Ammonia was not measured and plasma amino acid levels were within the normal ranges, only glutamine and glutamic acid levels were borderline (glutamine + glutamic acid = 833 μmol/l, controls up to 800 μmol/l). Urine orotic acid level was elevated (30.5 mmol/mol creatinine, normal range up to 3.5 mmol/mol creatinine). The levels dropped to 7.7 mmol/mol of creatinine within a week after the surgery and remained mildly elevated (4.5 – 8.3 mmol/mol creatinine). The results of allopurinol loading test were abnormal (orotic acid - pre-load: 5.8 mmol/mol creatinine, 0-8h: 19.2, 8-16h: 50.6, and 16-24h: 47.6 mmol/mol creatinine). Because of suspicion of OTC deficiency mild low-protein diet (2 g/kg/day) was introduced, nonetheless, mild elevation of aminotransferase activities up to twice upper limit of the normal range persisted.

At 5 years of age during a febrile illness the patient became somnolent and developed seizures with transient loss of consciousness; 10 fold elevation of ALT and AST was observed. Ammonia and orotic acid were not measured during this episode. Laboratory findings 3 months later, when the patient was asymptomatic, showed normal levels of blood ammonia, mild elevation of urine orotic acid (6.4 mmol/mol creatinine) and serum aminotransferases (ALT 2.12 μkat/l, AST 1.64 μkat/l), and normal plasma amino acid levels.

The suspicion of mild form of OTCD was considered while other causes of orotic aciduria were unlikely. Other urea cycle disorders and Hyperornithinemia-Hyperammonemia-Homocitrulinemia syndrome were excluded by

amino acid analysis. Orotic aciduria due to uridine monophosphate synthetase deficiency was not likely because of absence of megaloblastic anemia and normal levels of uracil in the urine (Davies, et al., 1997). The levels of urine orotate were up to 2.5 fold of the upper limit of normal range while in patients with benign persistent orotic aciduria the levels were 8 - 26.6 times elevated (Carpenter, et al., 1997). The fact that the peak levels of orotate concentration and aminotransferase activities were associated with increased catabolism (surgery, febrile disease) also denied benign persistent orotic aciduria. Finally, benign orotic aciduria is not associated with liver injury.

The patient's father had mild orotic aciduria (7.1 mmol/mol creatinine), mild elevation of plasma glutamine and glutamic acid (total 889 μmol/l, normal range up to 800 μmol/l). His allopurinol loading test was also abnormal (orotic acid - pre-load: 5.2, 0-8h: 20.1, 8-16h: 26.7, 16-24h: 16.1 mmol/mol creatinine) while urinary uracil was not elevated. Except for eczema he had no complaints. The patient mother's allopurinol loading test was negative and the patient's younger brother had normal levels of plasma ammonia and urine orotic acid.

Based on the data described above, the diagnosis of OTC deficiency was not proven beyond all doubts, however, the clinical course, orotic aciduria and positive results of allopurinol loading made the diagnosis of mild OTCD highly suspect.

Mutation analysis

The study was approved by the institutional review board of the General University Hospital in Prague and conducted under institutional guidelines. Written informed consent was obtained from the study participants. Genomic DNA was isolated from peripheral leukocytes of the patient, her family members and healthy control individuals using QIAamp columns (Qiagen, Valencia, CA). Ten coding exons of human *OTC* (GenBank NG_008471.1; NM_000531.5), including intron-exon boundaries, were analyzed by direct sequencing using ABI Prism 3100 - Avant Genetic Analyzer (Applied Biosystems, Foster City, CA). Primer sequences used for mutation analysis are shown in Table 1. The copy number of *OTC* coding exons was determined by multiplex ligation probe amplification assay (MLPA, MRC Holland, Amsterdam, The Netherlands).

Table 1. Primers and conditions used for mutation analysis of OTC

Primer ID	Primer sequence	T^{ann} (°C)	Product size	Amplified region
OTC-E1U	5'-ATCATCTTGCACCCCCAAAATAGC-3'	55	705 bp	Exon 1
OTC-E1L	5'-ATGATTCCTAAATCAAACCCAAGT-3'			
OTC-E2U	5'-CACCATAGTACATGGGTCTTTTCT-3'	53	287 bp	Exon 2
OTC-E2L	5'-AAAAAGGGGACTGGTAGTAACTGG-3'			
OTC-E3U	5'-ACACTTATTTGGGGGCTAGTTATT-3'	55	292 bp	Exon 3
OTC-E3L	5'-CCTCCAAAGTCTTCACCTTCAATC-3'			
OTC-E4U	5'-CTTGAATAAGAGGGTGGAGAAATG-3'	55	407 bp	Exon 4
OTC-E4L	5'-GATTCTGAAATCAGCTTTGGA-3'			
OTC-E5U	5'-GAACATCTGGCTTATGTAATGAAT-3'	55	433 bp	Exon 5
OTC-E5L	5'-AAAGCAAGTCAGGAATTATTTGAT-3'			
OTC-E6U	5'-GTTGCTCTGTGTAATGCAAAAAAAT-3'	55	453 bp	Exon 6
OTC-E6L	5'-AGTAGGGCTGGTAACGTAACCTAA-3'			
OTC-E78U	5'-AAGGCACTATTGAGTTTTGGAATT-3'	55	582 bp	Exons 7, 8
OTC-E78L	5'-TGCTCCCCCTAAAGTGAATAAGTG-3'			
OTC-E9U	5'-AGGTGTCTTGAGATAAAACCATCA-3'	55	458 bp	Exon 9
OTC-E9L	5'-AAAATAGATGTTACATGAGCAAGT-3'			
OTC-E10U	5'-AAATTCAGTCGTGGCTACCA-3'	55	548 bp	Exon 10
OTC-E10L	5'-TTTCACAATGGCAAAGCATA-3'			
pGI3_EnhU	5'-GGTGAACTTTGAACCTCTGTGA-3'	65	452 bp	Enhancer
pGI3_EnhL	5'-CCTCCCCATTGGAAGTACAG-3'			

Determination of transcription start sites and prediction of OTC regulatory regions

The 5′-terminal sequence of *OTC* was analyzed by RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) using the First-Choice™ RACE ready liver kit containing total liver cDNA enriched by a specific 5'- adapter (Ambion, Austin, TX). The kit was used as suggested in the protocol supplied by the manufacturer. Nested PCR was performed with the supplied 5'- adapter primers and two gene-specific primers: 5'-GCCCTTCAGCTGCACTTTATT-3' (outer primer) in exon 2 and 5'-AGCTGCATTGTTTAACAGGATCCTCAG-3' (inner primer) in exon 1. Major PCR products were gel-purified, cloned into the pCR2.1-TOPO® vector (Invitrogen, Carlsbad, CA) and sequenced.

The promoter and enhancer regions were predicted on the basis of the data published for rat and mouse (Veres, et al., 1986; Murakami, et al., 1989). Homology was assessed using NCBI Blast tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and UCSC genome browser Multiz Alignment&Conservation utility (http://genome.ucsc.edu).

Reporter plasmid constructs

Fourteen different plasmid constructs were generated by cloning the predicted promoter and enhancer fragments into pGL3-basic vector (Promega, Mannheim, Germany) containing the luciferase gene from *Photinus pyralis*. Two different variants of the promoter, both originating from the studied family, were tested: the wildtype variant (WT) and the variant with a single base substitution c.-366A>G (M or mutated promoter). The haplotype of both variants was otherwise the same within the studied region. Both wildtype and mutated promoter variants were subcloned into the *Xho*I restriction site 55 bp upstream of the luciferase transcription origin. The fragment predicted as an enhancer was subcloned into the *Sal*I restriction site 1923 bp downstream of the luciferase gene TSS. All combinations of inserts were prepared in both sense and anti-sense orientation in order to confirm the specificity of their regulatory function. The complete list of constructs is presented in Table 2.

Table 2. List of plasmids used for dual luciferase reporter assay

Plasmid	Promoter, orientation	Enhancer	Plasmid	Promoter, orientation	Enhancer
pGL3_WT+	wildtype, sense	-	pGL3_M+	c.-366A>G, sense	-
pGL3_WT-	wildtype, antisense	-	pGL3_M-	c.-366A>G, antisense	-
pGL3_WT+/Enh+	wildtype, sense	sense	pGL3_M+/Enh+	c.-366A>G, sense	sense
pGL3_WT-/Enh+	wildtype, antisense	sense	pGL3_M-/Enh+	c.-366A>G, antisense	sense
pGL3_WT+/Enh-	wildtype, sense	antisense	pGL3_M+/Enh-	c.-366A>G, sense	antisense
pGL3_WT-/Enh-	wildtype, antisense	antisense	pGL3_M-/Enh-	c.-366A>G, antisense	antisense
pGL3_Enh+	-	sense	pGL3_Enh-	-	antisense

Cell cultures, transient transfections and reporter gene assays

Four different cell lines were used: Human hepatoblastoma Hep-G2 (ECACC 85011430), human hepatocellular carcinoma HuH-7 (JCRB 0403), human hepatocellular carcinoma PLC/PRF/5 (ATCC CRL 8024) and human embryonic kidney cells HEK293 (ECACC 85120602). Hep-G2 cells were used as the main experimental cell line, HuH-7 and PLC/PRF/5 were used to confirm the results obtained in Hep-G2 cells and HEK293 cells were used as a control. All media and reagents for cell cultures were purchased from Sigma (Sigma-Aldrich, St. Louis, MO) unless otherwise indicated. Hep-G2, HEK293 and PLC/PRF/5 cell lines were cultured in minimal essential Eagle's medium, HuH-7 cells in Dulbecco's modified Eagle's medium. Both media contained L-glutamine, 10% fetal bovine serum, 100 units/ml penicillin, and 100 mg/ml streptomycin. The cultures were maintained in humidified atmosphere at 37 °C and 5% CO_2. Transfection was performed with Tfx-20 reagent (Promega, Mannheim, Germany) in Opti-MEM media (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol 48 hours after seeding. The cells were cotransfected with pGL3-derived experimental plasmids (0.5 μg per well) and a reporter

plasmid pRL-TK (0.05 μg per well) containing a luciferase gene from *Renilla reniformis.* Cell lysis and determination of the luciferase activity was performed using Dual Luciferase Reporter assay (Promega, Mannheim, Germany). Cells were harvested 48 h after transfection and lysed in 200 μl of passive lysis buffer. Ten microliters of the cell lysate was mixed with 50 μl of Luciferase Assay Reagent II in a luminometer tube. The light emission for the firefly luciferase was recorded immediately for 10 seconds after one second premeasurement delay using a Berthold Sirius luminometer (Bundoor, Australia). Subsequently, in another tube, 10 μl of cell lysate was quickly mixed with 50 μl of Stop&Glo^TM reagent and the *Renilla* luciferase light emission was measured for 2 seconds after one second premeasurement delay. To decrease the scattering of the *Photinus* luciferase luminescence caused by variations in cell number per well and transfection efficiency, the data were normalized to the *Renilla* luciferase signal. All transfections were done in triplicates and the results were calculated from 5 separate experiments.

Statistical analysis

SigmaPlot software (Systat Software, Chicago,IL) was used for statistical data processing. All values are given as mean ± SD of replicates. The results were evaluated with Kruskal-Wallis's one way ANOVA on ranks, multiple comparisons were set up using Dunn's method. Differences at $p \leq 0.01$ were considered as statistically significant.

RESULTS AND DISCUSSION

Mutation analysis

No predictably pathogenic mutation was found in the coding region of *OTC* with both direct sequencing and MLPA. The only detected sequence variation was a heterozygous single nucleotide substitution NG_008471.1:g.4849A>G (c.-366A>G; nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, according to journal guidelines, www.hgvs.org/mutnomen). This substitution, confirmed by sequencing of a second PCR product amplified from another sample of genomic DNA, has also been detected in the patient's father but was found neither in the patient's brother and mother. It was not identified by sequencing of either 300 control alleles of Czech origin or DNA samples from Portugal and Mozambique (Azevedo, et al., 2003). The nucleotide A is conserved in mammals including rat and mouse and the substitution has not been published in SNP databases (NCBI, http://www.ncbi.nlm.nih.gov; Ensembl, http://www.ensembl.org/index.html).

Determination of the transcription start site

Determination of the *OTC* TSS was essential for identification of the 5'-untranslated region (5'UTR) and localization of the c.-366A>G variation with regard to the *OTC* transcription region. Analysis of the 5'-RACE nested PCR reaction products by agarose electrophoresis revealed three bands. The sequence of all three PCR products, verified by direct DNA sequencing, corresponded to the sequence of the 5'UTR. The data indicate the presence of three TSSs located 95, 119 and 169 bp upstream of the initiation codon (Figure 1A).

The human *OTC* TSS was first assigned to the position c.-135 (GenBank K02100; Horwich, et al., 1984). Later published data suggested that transcription of human *OTC* does not initiate at a particular site but within a region of about 70 bases, mostly at preferred positions c.-95, c.-120, c.-150, c.-161 and c.-166 (Brusilow and Horwich, 2004). These findings correspond well with our results. The situation in human is different from that in mouse and rat where only one 5'UTR has been reported so far: The transcription origin of the mRNA was assigned to the position c.-136 bp in the mouse (Veres, et al., 1986) and c.-87 bp in the rat (Takiguchi, et al., 1987). Two sequences located 53 - 49 bp and 192 - 188 bp upstream of the TSS were previously suggested as potential binding site for TATA binding protein in the promoter of rat (Takiguchi, et al., 1987). Both these sequences are conserved in human promoter as GATAA at positions 142 - 138 bp and TATAA located 286 - 282 bp upstream of the initiation codon (Figure 1A). The *in-silico* sequence analysis of the human promoter region using Matinspector (Genomatix Software GmbH, München, Germany) confirmed the latter, however, this motif is not located at the expected position up to 50 nucleotides upstream of the TSSs found in human and rat. We therefore suggest that the TATA-like sequences CATAA, located 26 - 30 bp upstream of the c.-169 TSS, and GATAA, located 19 - 23 bp

upstream of the c.-119 TSS (see Figure 1A), may act as TATA-boxes in human. Our results together with the previously published data show that the novel variation c.-366A>G is located upstream of the *OTC* 5'UTR.
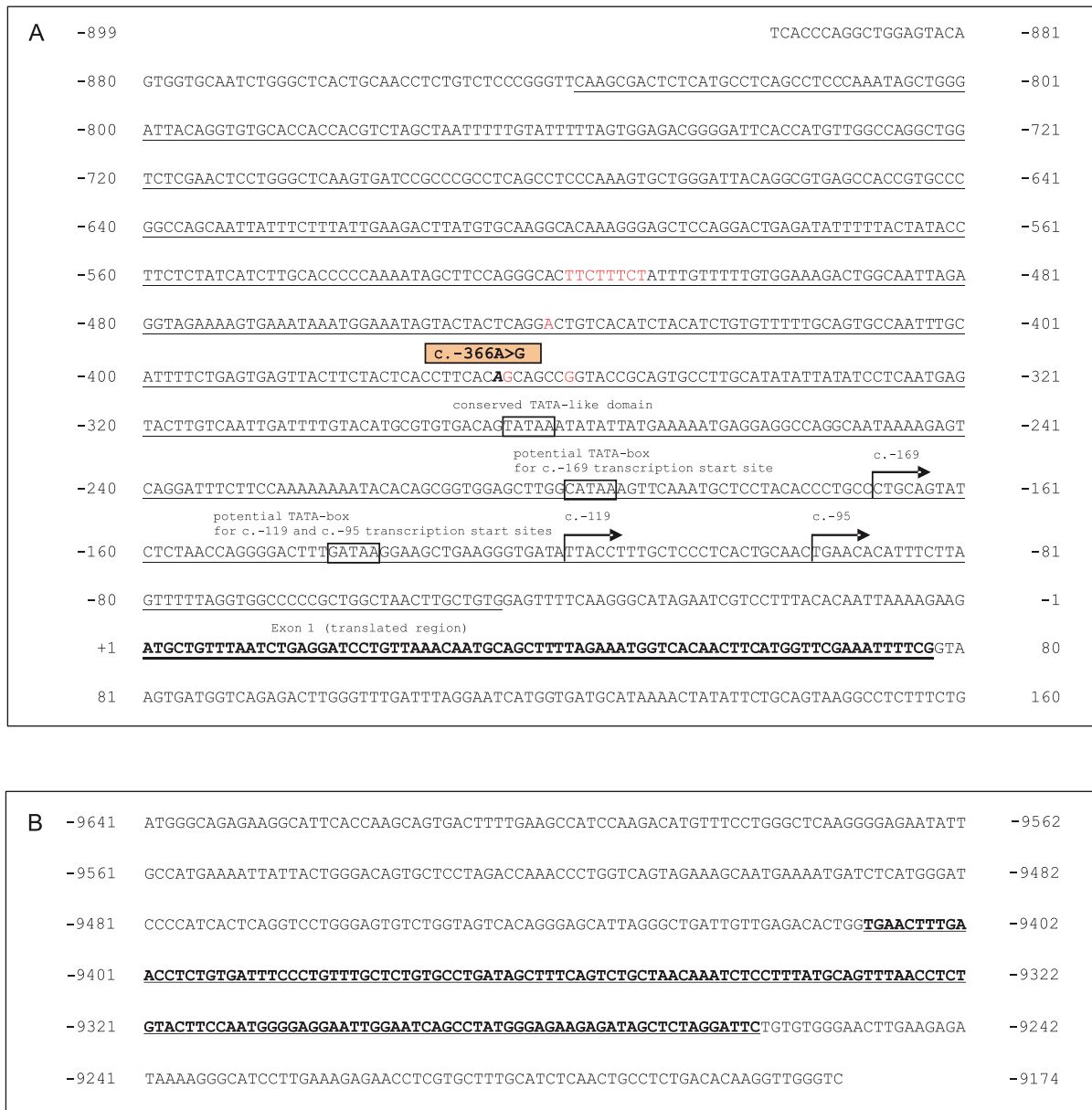
Figure 1. Sequence of the 5' flanking region of human *OTC* gene (**A**) and the distal enhancer (**B**). Panel **A** shows the sequence of the protein coding part of exon 1 (boldfaced & underlined), 5'UTR and the predicted proximal promoter. The sequence used for functional studies is underlined, the transcription start sites are marked with horizontal arrows, the single nucleotide substitution c.-366A>G is in bold and the TATA-like domains are depicted in boxes. Positions of known polymorphisms are in red. Panel **B** shows the fragment used in the functional studies containing the predicted distal enhancer (boldfaced and underlined).

Prediction and functional analysis of the OTC promoter

A 793 bp fragment ranging from -838 to -46 bp upstream of the *OTC* translation start was chosen for functional testing (WT promoter). The genotype in four polymorphic loci was the same as in the reference sequence NG_008471.1 (rs57752938 dupTTCT; rs5917572 A; rs5963030 G; rs5917573 G, see Figure 1A). The fragment revealed 72% sequence homology with both rat and mouse.

The level of the luminescence signal recorded after transfection of Hep-G2 cells with the construct pGL3_WT+ was 3 times higher than that of pGL3-basic (Figure 2A). Similar results were observed in the other two hepatocellular lines HuH-7 (441±79%) and PCL/PRF/5 (317±203%) and also in HEK293 cells (Figure 2A). Statistical analysis using Kruskal-Wallis test showed a significant difference (p≤0.01) between the activity of the construct pGL3_WT+ and both pGL3-basic and pGL3_WT-. The signal recorded from pGL3_WT- transfected cells was always lower than that of pGL3-basic. This can be explained by insertion of a long sequence lacking *cis*-acting elements and shifting the f1 ori transcriptional origin of the pGL3 plasmid.

No difference between the liver and kidney derived cells, together with the low positive regulatory effect on the reporter gene expression, can hardly explain the tissue specific expression pattern of *OTC*. The previous studies in transgenic mice revealed that introduction of the rat *Otc* controlled by the 1.3 kb rat promoter (Murakami, et al., 1989) and the human *OTC* controlled by the 0.7 kb mouse promoter (Jones, et al., 1990) can conduct small intestine-selective gene expression at levels comparable with those of the endogenous *Otc*. However, the mRNA levels of the introduced genes were low in the liver. Together with our data this suggests the presence of one or more liver-specific *cis*-acting element(s).

Prediction and functional analysis of the enhancer

A sequence homologous in 83% with rat enhancer described by Murakami et al. (1990) was found 9 kb upstream of the human *OTC* gene using the NCBI Blast tool. A 465 bp long fragment located 9177 to 9641 bp upstream of the initiation codon and containing the sequence homologous to rat enhancer (Figure 1B) was tested for its enhancer activity. No statistical differences were observed between the signal levels recorded after transfection of the HepG2 cells with pGL3_WT+/Enh- and pGL3_WT+. Also the signal obtained with both constructs pGL3_Enh+ and pGL3_Enh- containing the enhancer alone did not differ from the pGL3-basic signal. On the other hand, when using the construct pGL3_WT+/Enh+, the luminescence dramatically increased (Figure 2B). Similar results were observed in the other two cell lines derived from the liver (data not shown). In contrast, the signal levels in all experiments with pGL3_WT+, pGL3_WT+/Enh- and pGL3_WT+/Enh+ remained the same in HEK293 (Figure 2B). This indicates a role for the promoter-enhancer interaction in liver-specific expression of *OTC*.

Orientation of the enhancer relative to the promoter may be required for formation of the DNA-protein interaction complex. Our results may also indicate a role for such directionality, however, extrapolation to the situation *in vivo* without additional experimental verification is hardly possible.

Functional analysis of the variant promoter

The influence of the substitution c.-366A>G was studied by comparing the transcriptional activity of the plasmid variants containing the wildtype (pGL3_WT+, pGL3_WT+/Enh+) and mutated (pGL3_M+ and pGL3_M+/Enh+) promoter. No significant difference was observed between the transcriptional activity of the wildtype and mutated promoter in constructs without the enhancer (Figure 2C). The interaction of the promoter with the enhancer in Hep-G2 cells increased the activity of the reporter enzyme considerably in both wildtype and mutated promoter variants (Figure 2C). However, the difference between pGL3_M+/Enh+ and pGL3_WT+/Enh+ variants was not statistically significant (p=0.061) with regard to the low difference in means and high data dispersion when results from a series of independent experiments were evaluated. The high data dispersion typical for this experimental setting and documented by Figure 3 is caused by numerous known external factors affecting reproducibility of the dual luciferase reporter assay such as plasmid purity, pipetting error, transfection efficiency, lysis efficiency, substrate quality for enzymatic reaction, measurement errors and others. To avoid the interassay variation, a single assay with a high number of replications per group (n=22) was performed. In this experimental setting the activity of luciferase generated in pGL3_M+/Enh+ transfected Hep-G2 cells was only about 50% of the activity obtained with the pGL3_WT+/Enh+ construct (Figure 2D). By contrast, in the kidney-derived control cell line HEK293 the signal level of both construct variants did not differ from those in which the enhancer was not

present. The results observed in Hep-G2 cells were concordant with those obtained in HuH-7 (pGL3_WT+/Enh+: 1796±278%, pGL3_M+/Enh+: 1491±114%, single assay with 7 replicates, p≤0.01) and PCL/PRF/5 cells (pGL3_WT+/Enh+: 1081±304%, pGL3_M+/Enh+: 564±278%, single assay with 7 replicates, p≤0.01).

The results of our functional studies strongly suggest the role of the enhancer-dependent mechanism for regulation of *OTC* expression in the liver and molecular pathogenicity of the c.-366A>G promoter variation.
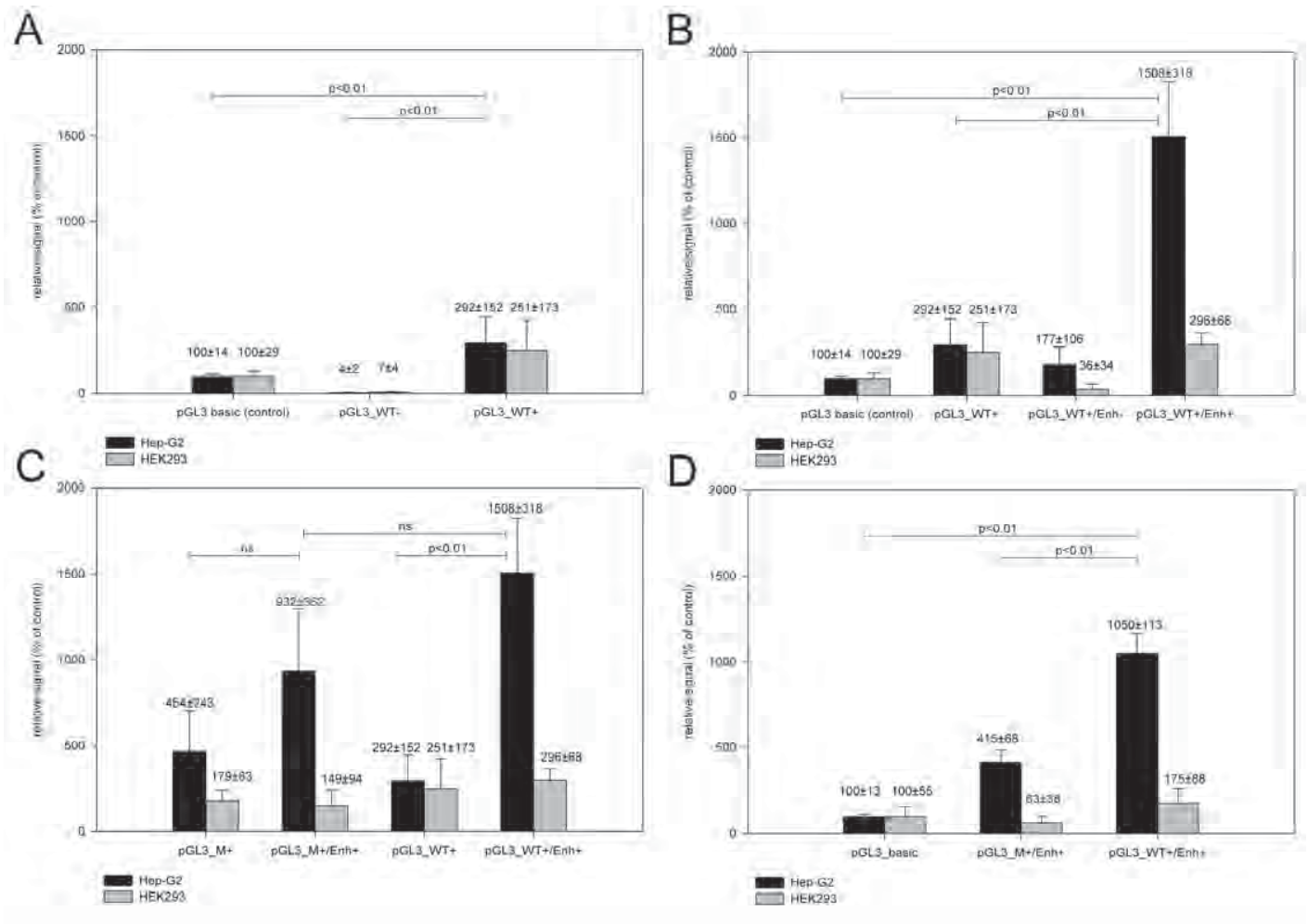
Figure 2. Functional analysis of *OTC* wildtype promoter, enhancer and variant promoter carrying the substitution c.-366A>G with the dual luciferase reporter assay. **A** – transcriptional activity of the predicted promoter; **B** – transcriptional activity of the promoter in the interaction with the predicted enhancer; **C, D** - influence of the single nucleotide substitution c.-366A>G on the activity of the promoter. Data from five independent experiments are shown in panel **C**. Panel **D** shows the intra-assay variation in 22 replications per group. Horizontal bars indicate statistical significance (Dunn's method).

Figure 3. Reporter luciferase activity in Hep-G2 and Hek293 cells transfected with 12 constructs containing the promoter sequence. Black points represent the luciferase activity of up to 15 measurements originating from 5 independent experiments. The box represents the range between the 25th and 75th percentile, horizontal line within the box shows the median. Whiskers indicate the 90th and 10th percentiles.

Conclusions

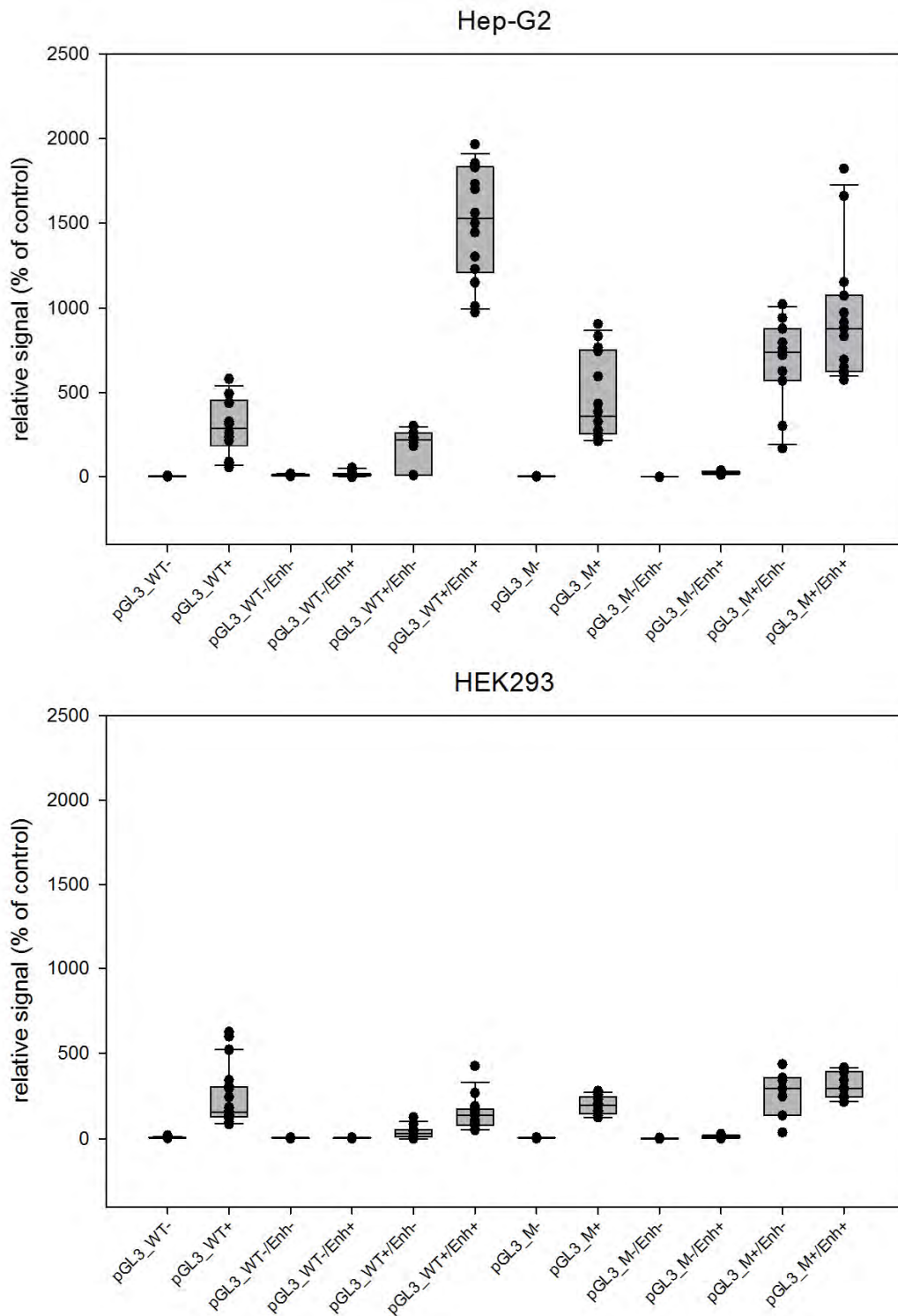In conclusion, our data indicate that transcriptional activity of the human *OTC* promoter depends on an upstream enhancer as do the rodent promoters. The promoter – enhancer interaction contributes to tissue specific expression of *OTC* in the liver. The unique promoter variation c.-366A>G does not affect the function of the promoter alone but it disrupts the interaction of the promoter with the enhancer. The c.-366A>G variation cosegregates with symptoms of OTCD in one family. Therefore, analysis of the regulatory regions of *OTC* should be considered in patients with OTCD and negative finding in the coding region of the gene.

ACKNOWLEDGMENTS

REFERENCES

Azevedo L, Stolnaja L, Tietzeova E, Hrebicek M, Hruba E, Vilarinho L, Amorim A, Dvorakova L. 2003. New polymorphic sites within ornithine transcarbamylase gene: population genetics studies and implications for diagnosis. Mol Genet Metab 78(2):152-7.

Brusilow S, W, Horwich A, L. 2004. Urea Cycle Enzymes. In: Scriver C, R., Beaudet A, L., Sly W, S. , Valle D, editors. Metabolic and Molecular Bases of Inherited Diseases -- OMMBID. www.ommbid.com. New York, McGraw Hill, Inc. Chap. 85.

Carpenter KH, Potter M, Hammond JW, Wilcken B. 1997. Benign persistent orotic aciduria and the possibility of misdiagnosis of ornithine carbamoyltransferase deficiency. J Inherit Metab Dis 20(3):354-8.

Davies PM, Fairbanks LD, Duley JA, Simmonds HA. 1997. Urinary uracil concentrations are a useful guide to genetic disorders associated with neurological deficits and abnormal pyrimidine metabolism. J Inherit Metab Dis 20(3):328-30.

Horwich AL, Fenton WA, Williams KR, Kalousek F, Kraus JP, Doolittle RF, Konigsberg W, Rosenberg LE. 1984. Structure and expression of a complementary DNA for the nuclear coded precursor of human mitochondrial ornithine transcarbamylase. Science 224(4653):1068-74.

Jones SN, Grompe M, Munir MI, Veres G, Craigen WJ, Caskey CT. 1990. Ectopic correction of ornithine transcarbamylase deficiency in sparse fur mice. J Biol Chem 265(24):14684-90.

Murakami T, Nishiyori A, Takiguchi M, Mori M. 1990. Promoter and 11-kilobase upstream enhancer elements responsible for hepatoma cell-specific expression of the rat ornithine transcarbamylase gene. Mol Cell Biol 10(3):1180-91.

Murakami T, Takiguchi M, Inomoto T, Yamamura K, Mori M. 1989. Tissue- and developmental stage-specific expression of the rat ornithine carbamoyltransferase gene in transgenic mice. Dev Genet 10(5):393-401.

Nishiyori A, Tashiro H, Kimura A, Akagi K, Yamamura K, Mori M, Takiguchi M. 1994. Determination of tissue specificity of the enhancer by combinatorial operation of tissue-enriched transcription factors. Both HNF-4 and C/EBP beta are required for liver-specific activity of the ornithine transcarbamylase enhancer. J Biol Chem 269(2):1323-31.

Takiguchi M, Murakami T, Miura S, Mori M. 1987. Structure of the rat ornithine carbamoyltransferase gene, a large, X chromosome-linked gene with an atypical promoter. Proc Natl Acad Sci U S A 84(17):6136-40.

Veres G, Craigen WJ, Caskey CT. 1986. The 5' flanking region of the ornithine transcarbamylase gene contains DNA sequences regulating tissue-specific expression. J Biol Chem 261(17):7588-91.

Yamaguchi S, Brailey LL, Morizono H, Bale AE, Tuchman M. 2006. Mutations and polymorphisms in the human ornithine transcarbamylase (OTC) gene. Hum Mutat 27(7):626-32.

4.3. HNF-4α regulates expression of human ornithin transcarbamylase through interaction with two positive *cis*-acting regulatory elements located in the proximal promoter

4.3.1. Commentary

The *OTC* is expressed predominantly in hepatocytes and epithelial cells of the intestinal mucosa (Hamano, et al., 1988, Ryall, et al., 1985). In the previous study we demonstrated the contribution of distal enhancer to the tissue specific expression of *OTC* in liver. The mechanism of tissue specific expression of *Otc* was partially elucidated in rat. The rat promoter carries two positive *cis*-acting regulatory elements recognized by a set of two transcription factors HNF-4 and COUP-TF. While liver-enriched factor HNF-4 induces *Otc* transcription, ubiquitous COUP-TF acting as competitive repressor is probably responsible for silencing the expression in other tissues (Kimura, et al., 1993). Two ortologous HNF4α binding sites were later identified in the mouse *Otc* promoter and HNF4α binding was confirmed using a specific antibody (Inoue Y, et al., 2002). The same group has also shown that HNF4α-null mice exhibited a phenotype closely similar to UCD (Inoue Y, et al., 2002).

Another two HNF-4 binding sites and two sites recognized by C/EBP family members were identified within the 110 bp rat minimal enhancer (Nishiyori A, et al., 1994). In co-transfection experiments the authors have shown that the enhancer activity could be reconstituted by concatemerization of a fragment containing one HNF-4 site and one C/EBP site, but not by concatemerization of either single site suggesting the importance of interaction between enhancer bound C/EBP and promoter bound HNF-4.

The present study conveys the topic of *OTC* regulatory regions and aims to deeper characterization of elements responsible for promoter and enhancer functions in human. We identified the key elements responsible for regulation of *OTC* transcription in human and we suggested a regulatory mechanism which is mediated directly by HNF-4α.

We delimited the minimal promoter encompassing the three alternative transcription start sites. We identified two positive *cis*-acting regulatory elements within the minimal promoter, both essential for the transcriptional activity of the promoter itself and for the interaction with the enhancer. We predicted the motifs recognized by appropriate transcription factors *in-silico* and confirmed protein binding at the corresponding sites by DNase I footprinting and electromobility shift assay (EMSA). With a modified EMSA protocol using specific competitor we identified three HNF-4α binding sites within the minimal promoter. Moreover, we

identified a series of C/EBPβ recognition motifs within the proximal enhancer which likely co-operate with promoter bound HNF-4α.

The manuscript was accepted for publication and stills in press.

HNF-4α regulates expression of human ornithin transcarbamylase through interaction with two positive *cis*-acting regulatory elements located in the proximal promoter

O. Lukšan[1], L. Dvořáková [2] and M. Jirsa[1]*

[1]Laboratory of Experimental Hepatology, Institute for Clinical and Experimental Medicine and Institute of Medical Biochemistry and Laboratory Diagnostics, First Faculty of Medicine, Charles University in Prague, Czech Republic
[2]Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague, and General University Hospital in Prague, Czech Republic

(ii)   **Corresponding author:** Milan Jirsa, Institute for Clinical and Experimental Medicine (IKEM), Laboratory of Experimental Hepatology, Videnska 1958/9, Praha 4 - 140 21, Czech Republic
E-mail: milan.jirsa@ikem.cz

(iii)   **Abbreviations:** C/EBP, CCAAT-enhancer-binding protein; *Cis+*, *cis*-acting regulatory element; COUP-TF, chicken ovalbumin upstream promoter-transcription factor; EMSA, electromobility shift assay; HNF, hepatocyte nuclear factor; Inr, initiator element; OTC, ornithine carbamoyltransferase; OTCD, ornithine carbamoyltransferase deficiency; PIC, pre-initiation complex; Sp1, specificity protein 1; TBP, TATA-box binding protein; TFII, general transcription factor for RNA-polymerase II; TIS, translation initiation site; TSS, transcription start site.

Abstract

OTC encodes ornithine carbamoyltransferase, the mitochondrial matrix enzyme involved in synthesis of urea. The tissue specific expression of *OTC* in the liver and intestine is dependent on the interaction of *OTC* promoter with an upstream enhancer. Two factors, HNF-4 and C/EBPβ, seem to be crucial for this interaction in rat and mouse. In the present study we focused on characterization of elements involved in regulation of *OTC* transcription in human.

Using a set of 5'-deleted promoter mutants in a reporter assay we identified two positive *cis*-acting regulatory elements located at c.-105 and c.-136 within the human *OTC* promoter. Both are essential for the transcriptional activity of the promoter itself and for the interaction with the enhancer. Protein binding at the corresponding sites was confirmed by DNase I footprinting. Electromobility shift assay with a specific competitor and anti-HNF-4α antibody identified the DNA-protein binding sites as HNF-4α recognition motifs. A third HNF-4α binding site has been found at the position c.-187. All three HNF-4α binding sites are located within 35 bp upstream of the transcription start sites at positions c.-95, c.-119 (major) and c.-169 (minor). A series of C/EBPβ recognition motifs was identified within the enhancer. Involvement of C/EBPβ and HNF-4α in the promoter-enhancer interaction is further supported by a massive DNA-protein interaction observed in the footprinting and EMSA assays.

Since the *OTC* promoter lacks general core promoter elements such as TATA-box or initiators on standard positions, HNF-4α most likely plays an essential role in initiation of *OTC* transcription in human.

Introduction

Ornithine carbamoyltransferase (OTC, EC 2.1.3.3; MIM #300461) is a mitochondrial matrix enzyme catalyzing the synthesis of citrulline - the second step of urea biosynthesis. Human *OTC* gene, located on chromosome Xp2l.1, comprises 10 coding exons (Hata et al., 1988, Lindgren et al., 1984). Ornithine carbamoyltransferase deficiency (OTCD; OMIM #311250) is the most common inherited urea cycle disorder (Brusilow and Horwich, 1995). Though its incidence is not high, the severity of clinical manifestations requires an early diagnosis, especially in males. Variations in the *OTC* coding regions were thoroughly investigated (Storkanova et al., 2013, Tuchman et al., 2002, Yamaguchi et al., 2006,); however, molecular diagnosis still fails in 20-25% of cases. Molecular basis of the semi-dominant disease in some of these cases might be related to a genetic defect in the regulatory elements: Recently we have reported a first patient manifesting mild form of OTCD caused by a point mutation in the promoter (Luksan et al., 2010).

The human *OTC* transcription start site (TSS) was first assigned to the position c.-135 (GenBank K02100; Horwich et al., 1984). Later published data suggested that transcription of human *OTC* does not initiate at a particular site but within a 70 bp region, mostly at preferred positions c.-95, c.-120, c.-150, c.-161 and c.-166 (Brusilow and Horwich, 1995). Three alternative TSS located at the positions c.-95 c.-119 and c.-169 upstream of the initiation codon were identified using rapid amplification of cDNA ends in total liver RNA (Luksan et al., 2010).

OTC is expressed predominantly in hepatocytes and epitelial cells of the intestinal mucosa (Hamano et al., 1988, Ryall et al., 1985). The mechanism of the tissue specific pattern of *Otc* expression was first studied in mouse. In transgenic experiments using sparse fur mice, Jones et al. (1990) showed that the *Otc* promoter was capable of direct liver- and intestine-specific transcription, but the expression level of the introduced gene was much lower than in the case of wild type mouse with endogenous *Otc* expression. In our previous study (Luksan et al., 2010), we investigated the 793 bp human *OTC* promoter localized 46 bp upstream of the translation initiation site (TIS). In line with the findings by Jones et al. (1990) we observed a significant increase of the reporter luciferase activity in Hep-G2 cells, but not in HEK296 cells, achieved by subcloning a 465 bp sequence of the distal enhancer localized 9 kb upstream of the TIS. This demonstrates the importance of the promoter-enhancer interaction for the regulation of human *OTC* expression.

Two positive *cis*-acting regulatory elements located around the positions -187 and -104 upstream of TIS were identified in the rat and mouse promoter. Both can bind an identical set of two members of orphan receptor superfamilly: Hepatocyte nuclear factor 4 (HNF-4) and chicken ovalbumin upstream promoter-transcription factor (COUP-TF). In co-transfection experiments,

HNF-4 activated the transcription of reporter gene from *Otc* promoter whereas competition with COUP-TF had a repressor effect (Kimura et al., 1993). Moreover, HNF-4 seems to contribute to the tissue specific expression of *Otc* as its mRNA is present only in a limited number of tissues such as the liver, kidney and intestine (Drewes et al., 1996). On the other hand, COUP-TF exhibits rather ubiquitous expression pattern (Miyajima et al., 1988) and therefore competition for the binding site can result in repression of *Otc* in other tissues.

The HNF-4α is critical for urea homeostasis in mouse (Inoue et al., 2002) since liver-specific HNF4α-null mice exhibited elevated serum ammonia and decreased serum urea, while the only significant decrease in protein expression within the urea cycle enzymes was that of OTC. Similarly as in rat, two HNF-4α recognition motifs were identified within the mouse 5' flanking region at positions -111 bp and -191 bp upstream of the TIS and HNF-4α binding was confirmed using specific antibody. Moreover, co-transfection of reporter constructs carrying mouse *Otc* promoter with a rat HNF-4α expression plasmid increased the reporter gene activity in the CV-1 cells (Inoue et al., 2002).

Nishiyori et al. (1994) demonstrated that rat *Otc* enhancer is activated by the co-operation of two liver-selective transcription factors: HNF-4 and CCAAT-enhancer-binding protein (C/EBP). Furthermore, co-transfection experiments showed that both HNF-4 and C/EBPβ are necessary and neither alone is sufficient for activation of the reconstituted enhancer in non-hepatic cells. Combinatorial operation of these two liver-enriched (but not strictly liver-specific) transcription factors leads to more restricted liver-specific transcription of *Otc* (Nishiyori et al., 1994).

In the present study we identified key elements responsible for regulation of *OTC* transcription in human. We suggest that initiation of *OTC* transcription in the absence of the general core promoter elements such as TATA-box or initiators on standard positions is mediated directly by HNF-4α.

Materials and Methods

Reporter plasmid constructs

Fourteen reporter constructs were generated from pGL3-basic vector (Promega, Mannheim, Germany) by subcloning the 465 bp *OTC* enhancer (9 constructs) and promoter variants (12 constructs) in *Sal*I and *Xho*I restriction site respectively. Six promoter variants 685, 458, 223, 173, 133 and 81 bp in length were prepared by a series of deletions from the 5'-end of the full 793 bp promoter (Table 1). All sequences and nucleotide positions used in the study are in accordance with GenBank NG_008471.1 and NM_000531.5.

Transient transfections and reporter gene assays

Human hepatoblastoma Hep-G2 (ECACC 85011430) cells were cultivated in minimal essential Eagle's medium supplemented with L-glutamine, 10% fetal bovine serum, 100 units/ml penicillin, and 100 µg/ml streptomycin, all purchased from Sigma-Aldrich, St. Louis, MO. The cultures were maintained in humidified atmosphere at 37°C and 5% CO_2. Transfection was performed 48 hours after seeding using FuGene HD reagent (Promega, Mannheim, Germany) in the reagent to DNA ratio 5:1 according to the manufacturer's protocol. The cells cultivated in 96-well plates were co-transfected with pGL3-derived experimental plasmids (0.1 µg per well) and a *Renilla* luciferase containing reporter plasmid pRL-TK (0.01 µg per well). Cell lysis and determination of the luciferase activity was performed according to the manufacturer's instructions using the Dual Luciferase Reporter assay (Promega). All measurements were performed on a Synergy 2 automatic microplate reader (BioTek, Winooski, VT).

Signal of the *Photinus* luciferase was normalized by *Renilla* luciferase luminiscence in all measurements. All transfections were done in 6 replicates and the results were calculated from at least 6 separate experiments. Because of the high interassay variation, the transcriptional activity of each promoter variant is presented relative to the activity of pGL3_793/Enh (positive control, 100%). Statistical evaluations were performed after logarithmic transformation using the SigmaPlot software (Systat Software, Chicago, IL). All groups passed the test for normality and analysis of variance using the one-way ANOVA. Multiple comparisons were performed using the Bonferroni method.

Nuclear extract preparation

Nuclear and cytoplasmatic extracts were prepared from Hep-G2 cells using the NE-PER nuclear extraction kit (Thermo Scientific, Waltham, MA) according to the manufacturer's instructions. Extracts were dialyzed for 4 hours against 50 volumes of buffer containing 20 mM HEPES (pH 7.9), 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF, and 0.5 mM DTT. Protein concentrations were determined using the Pierce BCA Protein Assay Kit (Thermo Scientific, Waltham, MA).

DNase I footprinting

The DNase I footprinting method based on the separation of fluorescently labeled products using capillary electrophoresis was derived from the literature (Wilson et al., 2001, Zianni et al., 2006). Fluorescent probe spanning the 223 bp *OTC* proximal promoter, 5'- untranslated region and a part of exon 1 was prepared by polymerase chain reactions using 6-carboxyfluorescein (6-FAM)-labeled reversed primer 5'-(6-FAM)-TTTCGAACCATGAAGTTGTGACCA-3' and the forward

primer 5'-AAATGAGGAGGCCAGGCAAT-3'. The 200 µl reaction mixtures containing the 100 µl of the protein binding buffer (2x), 2 µg of poly(dI-dT), 133 µg (sample) or 0 µg (control) of Hep-G2 nuclear extract and BSA up to 400 µg of total protein were incubated with 600 fmoles of fluorescently labeled probe for 20 minutes on ice. The DNA cleavage was performed with 10 µl of DNase I at the concentration 4 U/ml (Fermentas – Thermo Scientific, Pittsburgh, PA) after addition of 20 µl of Mg^{2+}/Ca^{2+} solution (10x). The cleavage was inhibited after 5 minutes by addition of the DNase stop solution and supplemented with GeneScan™ 500 ROX™ (Applied Biosystems, Foster City, CA) serving as a size standard and assay control. The DNA fragments were purified by phenol-chloroform extraction followed by ethanol precipitation, dissolved in 15 µl of Hi-Di formamide (Applied Biosystems) and separated by capillary electrophoresis on the Applied Biosystems 3130 Genetic Analyzer. The composition of the buffers and solutions used was as follows: Protein binding buffer (2x) – 50 mM Tris-HCl, pH 8.0, 100 mM KCl, 12.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 20 % (v/v) glycerol; Mg^{2+}/Ca^{2+} solution (10x) - 50 mM $MgCl_2$, 25 mM $CaCl_2$; DNase stop solution – 200 mM NaCl, 50 mM EDTA, 1 % (v/v) SDS.

Electromobility shift assays
Double-stranded biotin-labeled probes were prepared by standard PCR or by hybridization of complementary oligonucleotides. The binding reaction was carried out in 20 µl of a mixture containing 10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM MgCl2, 0.5 mM EDTA, 0.5 mM DTT, 4% glycerol 1 µg of poly(dI-dT), 25 µg of nuclear extract and 300 fmoles of biotin-labeled probe. In the case of competition assays, 0, 1, 4 and 12 pmoles of the non-labeled probe [-98..-126] or a 200-fold excess of a double-stranded probe AGTGAGGGTCAAAGTTCATATCAC, derived from the consensus HNF-4α recognition sequence was used as a specific competitor. The samples were incubated for 20 min at room temperature, separated using non-denaturing poly-acrylamide gel electrophoresis (6% polyacrylamide gel in 0.5x TBE, 100 V, 50 minutes) and electro-blotted at 300 mA, for 30 minutes on the Amersham Hybond-N nylon membrane (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The DNA – protein complexes were fixed by baking the membrane for 2 hours in a dry oven at 80°C. The membrane was processed using the Chemiluminiscent Nucleic Acid Detection Module (Thermo Scientific, Rockford, IL) according to the manufacturer´s instructions and analyzed with the FUJI LAS 3000 system (Fujifilm, Tokyo, Japan). In supershift assays, a goat anti-HNF-4α polyclonal antibody (Santa Cruz, Dallas, TX) was pre-incubated with the nuclear extract for 20 min prior to addition of labeled probes. Detection of the anti-HNF-4α antibody was performed using HRP-conjugated polyclonal rabbit anti-goat antibody (Thermo Scientific, Waltham, MA).

Prediction of transcription factor binding sites

Prediction of transcription factor binding sites was performed with MatInspector (Genomatix Software GmbH, München, Germany) and Match (Biobase, Wolfenbuettel, Germany) bioinformatic tools using the MatBase 9.0 and TRANSFAC 7.0 database versions.

Results

In silico prediction of the main regulatory domains

The sequences of rat *Otc* and human *OTC* promoter reach about 65% similarity. Higher homology is preserved in the proximal part of the promoter up to 450 bp upstream of the TIS. DNase I footprinting and functional characterization of deletion mutants generated from the rat *Otc* 5'-flanking region uncovered four regions protected against DNase I cleavage and two HNF-4α binding motifs located 104 and 187 bp upstream of the TIS (Kimura et al., 1993, Murakami et al., 1990) (Figure 1). Homologous motifs were found in mouse by Inoue et al. (Inoue et al., 2002). Similar binding sites located at c.-105 and c.-187 have been predicted in the human sequence using MatInspector (MI) and Match (M) (Figure 1). While the HNF-4α binding motif located at c.-105 was recognized with both tools (MI score 0.884, M score 0.883), the other located at c.-187 was predicted only with Match (score 0.910). MatInspector recognized at the same position a binding site for another member of nuclear receptor superfamily - the peroxisome proliferator-activated receptor gamma (PPARγ, MI score 0.833). Both HNF-4 and PPARγ recognize similar sequence motifs. Transfection experiments revealed that preferential binding is partially determined by the sequence of the core motif but it is much more dependent on the promoter context (Nakshatri and Bhat-Nakshatri, 1998). Therefore, the sequence motif can be considered a HNF-4α binding site as well.

Two additional potential HNF-4α binding sites located 136 bp (MI/M scores 0.923/0.746) and 588 bp (MI/M scores 0.822/0.835) upstream of the TIS, and one site recognized by C/EBPβ located 221 bp upstream of the TIS (M score 0.970) were predicted within the human *OTC* promoter. None of these binding sites was recognized either in rat or in mouse promoter because of variations in the matrix core motifs.

As mentioned above, human *OTC* lacks a particular site of transcription initiation indicating the absence of a strong core promoter element such as TATA-box. Accordingly, in our previous study (Luksan et al., 2010) we found three alternative TSSs - two major starting at positions c.-95, c.-119 and a minor one starting at c.-169. From these three only the c.-169 transcript variant implies a potential TATA-like element CATAAA located 25-30 bp upstream of the TSS; however, its regulatory role seems to be low because cutting off the corresponding region did not

lead to remarkable decrease in the promoter activity (Figure 2, constructs 173/Enh and 133/Enh). Moreover, this fully conserved motif acts as a known HNF-4α binding site located 187 bp upstream of the TIS in rodents.

Identification of *cis*-acting regulatory elements and delimitation of the core promoter

Transcriptional activity of six 5'-deleted promoter mutants ranging 685, 458, 223, 173, 133 and 81 bp (all with 3´-ends 46 bp upstream the TIS – see Figure 1) was compared with the full 793 bp human promoter in pGL3 constructs containing the human enhancer sequence. The empty plasmid pGL3_0/Enh containing only the enhancer was used as a negative control to eliminate the influence of the enhancer alone. The results have shown that a 5'-deletion of 108 bp has no influence on the reporter activity (both mean and median varied around 116% of the activity of pGL3_793/Enh), while the 5'-deletion of 335 bp leads to a decreased luciferase signal (median and mean reached 57% and 54% respectively) (Figure 2). In contrast, increased reporter activity was observed with the 223 bp promoter variant (median 146%, mean 152%). The response remained constant through the 173 and 133 bp variants (medians 137% and 156%, means 150% and 167% respectively) while the shortest tested variant of 81 bp showed slightly lower transcriptional activity when compared with the full promoter (median 81%, mean 80%) (Figure 2). The decrease was statistically significant when compared with the 133, 173 and 223 bp promoter variants, but not significant when compared with the 458 and 685 bp variants (Figure 2). The reporter gene activity observed with the pGL3 vector carrying the enhancer alone was lower than 20% of the pGL3_793/Enh and significantly lower than all other constructs.

Based on these results we defined the human *OTC* core promoter as the 223 bp region located -46..-268 bp upstream of the TIS containing three transcription initiation sites and two positive *cis*-acting elements: *Cis1+* located within the first 81 base-pairs, reaching 83% of the full promoter activity and *Cis2+* located within the upstream 52 bp sequence.

In the following experiments we evaluated the contribution of *Cis1+* and *Cis2+* to the transcriptional activity of the promoter alone (Figure 2, right plot). The core promoter exhibited twice higher reporter gene activity than the full promoter and the level of luciferase signal remained stable within all tested constructs (81/0, 133/0, 173/0 and 223/0). Finally, the reporter gene signal observed with all promoter variants was significantly higher than with the negative control pGL3_0/0. The influence of *Cis2+* was not observed in this case indicating two possible explanations: First, *Cis2+* regulatory element contributes only to the interaction with the enhancer whereas *Cis1+* plays a role in both interaction with the enhancer and activation of the promoter. Second, *Cis1+* and *Cis2+* operate simultaneously but they contribute to formation of different mRNA variants originating from various TSSs. In such case, absence of one element

responsible for formation of one transcript variant could be compensated by the action of the other element responsible for production of the other transcript originating from different TSS. Such compensatory effect could maintain the overall transcriptional activity of the promoter. However, the compensatory capacity may reach its limit when the transcriptional activity is accelerated by the promoter – enhancer interaction and therefore the promoter variant containing only the *Cis1+* element alone exhibits lower activity when compared with the variant carrying both regulatory units (Figure 2, variants 81 and 133).

Identification of DNA-protein interaction sites by DNase I footprinting

In a series of experiments we identified five DNase I protected regions localized 92-102 (A), 105-114 (B), 135-145 (C), 150-172 (D) and 180-192 (E) bp upstream of the TIS (Figure 3A). The sequence motifs observed in the protected regions B and E are very close to the protected regions in the rat promoter which partially overlap with the HNF-4 & COUP-TF binding sites (Figure 1). Another HNF-4α recognition site predicted with high matrix similarity score is located close to the protected region C.

No conventional transcription factor binding sites were recognized within the protected region A which coincides with the c.-95 bp TSS and overlaps with a sequence motif related to initiator element Inr (MI score 0.967). The same sequence was recognized with Match as a Cap site with matrix similarity score 0.995. A general core promoter motif Inr is known to moderate transcription by direct interaction with the transcription factor II D (TFIID) complex in both TATA-containing and TATA-less promoters. Unfortunately, the role of the initiator element in the *OTC* promoter remains unclear since the strictly unidirectional Inr is located on the template strand in the antisense orientation.

Structural motifs recognized by two different transcription factors with high matrix similarity score were predicted *in silico* within the protected region D: GATA-binding factor 1 (GATA1), located 155 bp upstream of the TIS (MI score 0.962, M score 0.970) and Specificity Protein 1 (Sp1) at position -169..-177 (M score 0.962). Whereas GATA1 is essential for erythroid and megakaryocyte development, its expression pattern is limited to the hematopoietic cell lineage and its involvement in regulation of *OTC* transcription is thus unlikely. The Sp1 is a rather ubiquitous transcription factor capable of regulating basal expression in genes with TATA-less promoters (Black et al., 2001). The Sp1 recognition site overlaps the alternative TSS located 169 bp upstream of the translation origin. Since no other general core promoter has been found in a standard position related to the TSS except for the TATA-like domain discussed earlier, the contribution of Sp1 to the initiation of transcription from the position c.-169 seems likely.

Analysis of the DNA – protein complexes by EMSA

The DNA – protein interactions observed by DNase I footprinting were verified in a set of electromobility shift assays (EMSA). Six double-stranded biotin-labeled probes were designed to cover the regions containing *cis*-acting regulatory elements, DNase I protected regions or *in silico* predicted transcription factor binding sites (see Figure 3C). Interaction with the HepG2 nuclear extract led to an obvious shift: Three shifted bands were observed with the probes covering the regions located at positions -46..-126 and -98..-178, two bands with probes spanning the regions -168..-218 and -204..-268 and one zone was detected using probe corresponding to the sequence localized -98..-126 upstream of the TIS.

The probe [-98..-126] spanning 29 bp was designed to cover the proximal HNF-4α binding site. The motif is conserved in human and rodents and overlaps the region B protected against the DNase I cleavage. The site is also contained in the [-46..-126] probe carrying the *Cis1+* regulatory element and in the [-98..-178] probe. Both probes exhibited a different band shift pattern than the [-98..-126] probe indicating formation of additional DNA-protein complexes (Figure 3B). One of these complexes formed with the probe [-46..-126] may involve the intermediates formed by the factors binding to the Inr element. The pattern obtained with the probe [-98..-178] likely involves complexes with other HNF-4α and Sp1 predicted to bind the sequences of protected regions C and D.

The probe [-168..-218] was designed to overlap the conserved HNF-4α binding site, found also in rat around the position -187 bp upstream of the TIS, and the DNase I protected region E. The probe [-204..-268] covered the distal part of the core promoter region, where a sequence motif recognized by C/EBPβ was predicted.

Specific interaction of HNF-4α with probes [-46..-126] and [-98..-178] was examined in competition assays using a non-labeled probe [-98..-126]. The reaction mixture containing both biotin-labeled and non-labeled probe [-98..-126] was used as a control indicating specific manner of the competition. Repression of nuclear factor binding was observed with all three probes. A gradual decrease of two differentially shifted bands was observed with the probe [-98..-178] carrying two motifs recognized by HNF-4α (Figure 3B).

Confirmation of specific interaction between HNF-4α and the predicted binding sites located 105 bp and 136 bp upstream of the TIS was achieved using probes [-98..-126] and [-132..-156] (Figure 3C). Addition of specific competitor (a non-labeled probe corresponding to the consensus sequence of HNF-4α recognition motif) in 200-fold excess led to significant decrease in the shifted band intensity. Supershifted bands of the DNA-protein-antibody complexes were observed after incubation with the anti-HNF-4α antibody. Presence of the anti-HNF-4α antibody

in these complexes was confirmed using HRP-conjugated secondary antibody: The immunoreactive zones co-localized with the supershifted bands (Figure 3B).

Mutation in *Cis 1+* element abolished HNF-4α binding

To confirm the role of the *Cis 1+* element as a HNF-4α binding site we generated a mutant 793(HNF4mut) in which the matrix sequence CAAAGG was replaced by CAccaG and then compared the transcriptional activity of both variants combined with the enhancer in a set of reporter assays (Figure 4A). The activity of pGL3_793(HNF4mut)/Enh construct was more than two times lower than in the case of normal promoter and the decrease was highly significant in Student's t-test. When compared with the reporter gene activity obtained with other variant constructs, the pGL3_793(HNF4mut)/Enh reached similar level as the core promoter variants without enhancer and it was about two times lower than pGL3_81/Enh (mean and median reached 41% and 34% of pGL3_793/Enh respectively). Interestingly, in spite of the disruption of the HNF-4α binding site, the GL3_793(HNF4mut)/Enh activity was partially preserved indicating the ability of the upstream element(s) compensate function of the disrupted *Cis1+*.

Discussion

We have shown that the minimal promoter responsible for both the promoter activity and activation of the enhancer is located within the region -46..-126 bp upstream of the TIS. As the shortest reported transcription variant starts at the position c.-95, the minimal promoter contains one strong positive *cis*-acting regulatory element *Cis1+* located within 30 bp upstream of the TIS. The motif is highly conserved in vertebrates and the affinity of HNF-4 to this region has been clearly demonstrated in rat (Kimura et al., 1993) and mouse (Inoue et al., 2002). Accordingly, we found that the *Cis1+* spanning region is protected against DNase I cleavage and then experimentally confirmed binding of HNF-4α at this site (Figure 3B).

The *Cis2+* element was identified within the region between -127 and -178 bp upstream of the TIS overlapping with the protected regions C and D. *Cis2+* is involved in the interaction with the distal enhancer, but it does not directly influence the promoter activity alone (Figure 2). Sequence motifs recognized by nuclear factors HNF-4α and Sp1 were predicted *in silico* in this region (Figure 3C). Binding of HNF-4α was confirmed in the band shift competition assay where two bands obtained with the probe [-98..-178] and one band obtained with the probe [-132..-156] decreased proportionally to the increased amount of specific competitor (Figure 3B). Presence of two differentially shifted bands with the probe [-98..-178] can be explained by formation of two different complexes probe-HNF4 and HNF4-probe-HNF4 as the probe carries two HNF-4α

binding sites and the ratio of both complexes remained constant even when repressing their formation by a specific competitor. A direct proof of HNF-4α binding at this site was obtained by detection of supershifted complex DNA-protein-antibody (Figure 3B).

Binding of Sp1 could also explain transcription from alternative *OTC* TSS at the position c.-169 but it cannot explain the role of *Cis2+* in the interaction with the enhancer.

Unlike the two HNF-4α binding sites discussed above, the motif located 187 bp upstream of the translation origin didn't exhibit significant influence on transcriptional activation of *OTC* (Figure 2). Nevertheless, the matrix sequence is highly conserved and it differs from the rodent sequence only in one nucleotide located outside of the matrix core. The region partially correlates with the protected region E and interaction of HNF-4 with that site has been confirmed in rodents (Kimura et al., 1993; Inoue et al., 2002). The motif overlaps with the probe [-168..-218] which competed for HNF-4α with the HNF-4α-specific competitor in EMSA. Therefore it is obvious that the HNF-4α binding site is active, despite that its contribution to the overall promoter function is negligible when compared with the *Cis1+* and *Cis2+* elements. The low contribution is further supported by observations in mouse (Inoue et al., 2002).

Fang et al. (2012) reported recently a genome-wide study evaluating matrix sequence motifs in a large group of HNF4-specific genes using high-throughput ChIP-seq. Motifs recognized by 1371 specific sequences were statistically evaluated based on their binding affinity. Matrices preferred by top 10% (strongest), top 33% (strong) and middle 33% (middle) binders originating from the Fang's study were compared with all four *in silico* predicted HNF-4α binding sites and their rat and mouse homologues (Figure 4B). The matrix sequence is highly conserved and practically all compared regions could be considered as at least middle affinity binders to HNF-4α.

Activation of the enhancer in rat is associated with the interaction of promoter-bound HNF-4 with the enhancer-bound C/EBPβ (Nishiyori et al., 1994). A similar mechanism involving synergistic co-operation of HNF-4 and C/EBPα, both bound to the promoter, has been suggested for human apolipoprotein B gene (Metzger et al., 1993). Indeed, we predicted a binding site for C/EBPβ in the *OTC* core promoter between positions c.-221 and c.-234 and the results of EMSA revealed two shifted bands indicating the formation of DNA-protein complexes within this region. In contrast, cutting off the region located -219 bp to -268 bp upstream of the TIS did not change the level of the reporter signal in luciferase assays.

In silico analysis of the proximal 200 bp part of the enhancer region revealed a set of six repeated sequence motifs corresponding to the C/EBPβ binding site. Accordingly, the DNase I cleavage of the proximal part of the enhancer was decreased and an intensive DNA-protein smear was observed in the gel retardation assay after incubation of the proximal enhancer fragment

(data not shown). Interaction of the liver enriched HNF-4 factors bound to the *OTC* core promoter with the set of periodically occurring liver specific C/EBP-like proteins thus may contribute to the tissue specificity of *OTC* expression in the liver, as was seen in the rat model (Nishiyori et al., 1994).

The distribution of HNF-4α recognition sites within the *OTC* core promoter region exhibits an important feature: All three HNF-4α binding motifs are located within 35 bases upstream of the TSSs (Figure 1). Moreover, the sequence of 223 bp proximal promoter lack any general core promoter elements such as Inr, the downstream promoter element (DPE), the TFIIB recognition element (BRE), the motif ten element (MTE), downstream core element (DCE) and the X core promoter element (XCPE) located at standard positions upstream of the TSSs, thereby suggesting the importance of HNF-4α in the initiation of *OTC* transcription. Initiation of transcription mediated by DNA-dependent RNA-polymerase II is dependent on the assembly of functional pre-initiation complex (PIC) consisting of general transcription initiation factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (reviewed in Hahn, 2004, Myer and Young, 1998). The transcriptional machinery and the PIC formation is well described in promoters equipped with general core promoter elements but it has been poorly studied in promoters lacking these elements. However, few indices suggesting the role of HNF-4α in the transcription from TATA-less and Inr-less promoters can be found in the literature. The HNF-4α was shown to activate transcription *in vitro* by facilitating assembly of the pre-initiation complex through direct physical interaction with TFIIB. The interaction of TFIIB with DNA-bound HNF-4α led to TFIIB-mediated recruitment of TBP. Furthermore, HNF-4α participated in the assembly of downstream basal factors into the PIC (Malik and Karathanasis, 1996). Similar mechanism comprising both the stable complex activator-TFIIB-TBP assembly and involvement in the association of downstream basal transcription factors TFIIE and TFIIF was observed later in estrogen receptor as another member of the nuclear hormone receptor superfamily (Sabbah et al., 1998). In a recent study, the systematic characterization of the HNF4α-TFIID link revealed that the HNF4α DNA-binding domain binds directly to the TBP and, through this interaction, can target TBP or TFIID to promoters containing HNF4α-binding sites (Takahashi et al., 2009).

In conclusion, our data strongly suggest the essential role of HNF-4α in regulation of *OTC* transcription in human. The regulatory mechanism may involve physical interaction with TFIIB, recruitment of TBP and stabilization of TFIID to particular site in the promoter followed by the assembly of the pre-initiation complex.

References

Black, A. R., Black, J. D., Azizkhan-Clifford, J. (2001). Sp1 and kruppel-like factor family of transcription factors in cell growth regulation and cancer. *J Cell Physiol, 188*, 143-160.

Brusilow, S. W., Horwich, A. L. (1995). Urea cycle enzymes. In R. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, eds (Eds.) *The molecular and metabolic bases of inherited disease.* pp. 1187-1232). New York: McGraw-Hill.

Drewes T., Senkel S., Holewa B., Ryffel G. U. (1996). Human hepatocyte nucler factor 4 isoforms are encoded by distinct and differentially expressed genes. *Mol Cell Biol*, 16(3): 925-31.

Fang, B., Mane-Padros, D., Bolotin, E., Jiang, T., Sladek, F. M. (2012). Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. *Nucleic Acids Res, 40*, 5343-5356.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol, 11*, 394-403.

Hamano, Y., Kodama, H., Yanagisawa, M., Haraguchi, Y., Mori, M., Yokota, S. (1988). Immunocytochemical localization of ornithine transcarbamylase in rat intestinal mucosa. Light and electron microscopic study. *J Histochem Cytochem, 36*, 29-35.

Hata, A., Tsuzuki, T., Shimada, K., Takiguchi, M., Mori, M., Matsuda, I. (1988). Structure of the human ornithine transcarbamylase gene. *J Biochem, 103*, 302-308.

Horwich, A. L., Fenton, W. A., Williams, K. R., Kalousek, F., Kraus, J. P., Doolittle, R. F., Konigsberg, W., Rosenberg, L. E. (1984). Structure and expression of a complementary DNA for the nuclear coded precursor of human mitochondrial ornithine transcarbamylase. *Science, 224*, 1068-1074.

Inoue, Y., Hayhurst, G. P., Inoue, J., Mori, M., Gonzalez, F. J. (2002). Defective ureagenesis in mice carrying a liver-specific disruption of hepatocyte nuclear factor 4alpha (HNF4alpha). *J Biol Chem*, 277, 25257-25265.

Jones, S. N., Grompe, M., Munir, M. I., Veres, G., Craigen, W. J., Caskey, C. T. (1990). Ectopic Correction of Ornithine Transcarbamylase Deficiency in Sparse Fur Mice. *J Biol Chem*, 265, 14684-14690.

Kimura, A., Nishiyori, A., Murakami, T., Tsukamoto, T., Hata, S., Osumi, T., Okamura, R., Mori, M., Takiguchi, M. (1993). Chicken Ovalbumin Upstream Promoter-Transcription Factor (Coup-Tf) Represses Transcription from the Promoter of the Gene for Ornithine Transcarbamylase in a Manner Antagonistic to Hepatocyte Nuclear Factor-Iv (Hnf-4). *J Biol Chem*, 268, 11125-11133.

Lindgren, V., de Martinville, B., Horwich, A. L., Rosenberg, L. E., Francke, U. (1984). Human ornithine transcarbamylase locus mapped to band Xp21.1 near the Duchenne muscular dystrophy locus. *Science*, 226, 698-700.

Luksan, O., Jirsa, M., Eberova, J., Minks, J., Treslova, H., Bouckova, M., Storkanova, G., Vlaskova, H., Hrebicek, M., Dvorakova, L. (2010). Disruption of OTC promoter-enhancer interaction in a patient with symptoms of ornithine carbamoyltransferase deficiency. *Hum Mutat*, 31, E1294-1303.

Malik, S., Karathanasis, S. K. (1996). TFIIB-directed transcriptional activation by the orphan nuclear receptor hepatocyte nuclear factor 4. *Mol Cell Biol*, 16, 1824-1831.

Metzger, S., Halaas, J. L., Breslow, J. L., Sladek, F. M. (1993). Orphan Receptor Hnf-4 and Bzip Protein C/Ebp-Alpha Bind to Overlapping Regions of the Apolipoprotein-B Gene Promoter and Synergistically Activate Transcription. *J Biol Chem*, 268, 16831-16838.

Miyajima, N., Kadowaki, Y., Fukushige, S., Shimizu, S., Semba, K., Yamanashi, Y., Matsubara, K., Toyoshima, K., Yamamoto, T. (1988). Identification of two novel members of erbA superfamily by molecular cloning: the gene products of the two are highly related to each other. *Nucleic Acids Res, 16*, 11057-11074.

Murakami, T., Nishiyori, A., Takiguchi, M., Mori, M. (1990). Promoter and 11-Kilobase Upstream Enhancer Elements Responsible for Hepatoma Cell-Specific Expression of the Rat Ornithine Transcarbamylase Gene. *Mol Cell Biol, 10*, 1180-1191.

Myer, V. E., Young, R. A. (1998). RNA polymerase II holoenzymes and subcomplexes. *J Biol Chem, 273*, 27757-27760.

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.

Nishiyori, A., Tashiro, H., Kimura, A., Akagi, K., Yamamura, K., Mori, M., Takiguchi, M. (1994). Determination of tissue specificity of the enhancer by combinatorial operation of tissue-enriched transcription factors. Both HNF-4 and C/EBP beta are required for liver-specific activity of the ornithine transcarbamylase enhancer. *J Biol Chem, 269*, 1323-1331.

Ryall, J., Nguyen, M., Bendayan, M., Shore, G. C. (1985). Expression of nuclear genes encoding the urea cycle enzymes, carbamoyl-phosphate synthetase I and ornithine carbamoyl transferase, in rat liver and intestinal mucosa. *Eur J Biochem, 152*, 287-292.

Sabbah, M., Kang, K. I., Tora, L., Redeuilh, G. (1998). Oestrogen receptor facilitates the formation of preinitiation complex assembly: Involvement of the general transcription factor TFIIB. *Biochem J, 336*, 639-646.

Storkanova, G., Vlaskova, H., Chuzhanova, N., Zeman, J., Stranecky, V., Majer, F., Peskova, K., Luksan, O., Jirsa, M., Hrebicek, M., Dvorakova, L. (2013). Ornithine carbamoyltransferase deficiency: molecular characterization of 29 families. *Clin Genet*, *84*, 552-559.

Takahashi, H., Martin-Brown, S., Washburn, M. P., Florens, L., Conaway, J. W., Conaway, R. C. (2009). Proteomics Reveals a Physical and Functional Link between Hepatocyte Nuclear Factor 4 alpha and Transcription Factor IID. *J Biol Chem, 284*, 32405-32412.

Tuchman, M., Jaleel, N., Morizono, H., Sheehy, L., Lynch, M. G. (2002). Mutations and polymorphisms in the human ornithine transcarbamylase gene. *Hum Mutat, 19*, 93-107.

Wilson, D. O., Johnson, P., McCord, B. R. (2001). Nonradiochemical DNase I footprinting by capillary electrophoresis. *Electrophoresis, 22*, 1979-1986.

Yamaguchi, S., Brailey, L. L., Morizono, H., Bale, A. E., Tuchman, M. (2006). Mutations and polymorphisms in the human ornithine transcarbamylase (OTC) gene. *Hum Mutat, 27*, 626-632.

Zianni, M., Tessanne, K., Merighi, M., Laguna, R., Tabita, F. R. (2006). Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument. *J Biomol Tech, 17*, 103-113.

Figure legends

Figure 1: Sequence alignment of the rat *Otc* (lower case) and human (upper case) *OTC* 5'
flanking region. The regions protected against DNA I cleavage in rat are underlined. The 5´- ends
of the rat promoter fragments studied by (Murakami et al., 1990) and their relative activity in
chloramphenicol acetyltransferase reporter assays are marked with arrows. Transcription factor
binding sites identified as *cis*-acting regulatory elements in rat are boldfaced. Transcription factor
binding sites predicted by Match and MatInspector in human *OTC* promoter are boldfaced and
highlighted in grey. Black arrows indicate human *OTC* promoter variants tested in reporter
assays. Three alternative TSSs identified within the human liver mRNA are marked with black
triangles.

Figure 2: Transcriptional activity of the mutant promoter variants in combination with the
enhancer (left panel) and without the enhancer (right panel). The data are based on at least 6
independent experiments. The box represents the range between the 25th and 75th percentile,
horizontal line within the box shows the median. Whiskers indicate the minimal and maximal
value. Diamonds are used for means. Multiple comparisons of variance using Bonferroni method
are presented in the table below the plots. Statistical significance is represented by asterisks:
(***) indicate the differences with $p \leq 0{,}005$ and (**) is used for statistical significance with $p \leq$
0.01.

Figure 3: Analysis of the DNA – protein interactions within human *OTC* core promoter. **A** –
DNase I footprinting. Peaks represent the fragments obtained after DNase I cleavage (in blue)
and internal standard (in red). Regions protected against the endonuclease cleavage are marked
by red horizontal bars; **B** – Electromobility shift assays (EMSA). Standard EMSA (upper left
panel), competition assays with a specific competitor for HNF-4α (upper right panel) and a
specific confirmation of HNF-4α binding to the predicted sites using competitor and supershift
assays (lower panels). Detection of the anti-HNF-4α antibody was performed by probing the
same membrane with a secondary antibody (right electrophoreograms in lower panels). NX -
incubated with nuclear extract. **C** – Schematic summary of results. Horizontal bars represent
promoter fragments used in the dual liciferase reporter assay (dark blue), protected regions
observed with the DNase I protection assay (red) and EMSA probes (light blue). Upper part of
the figure represents transcription factors and their binding sites (underlined sequence) predicted
with high score using MatInspector and Match programs.

Figure 4: A – Functional analysis of wildtype (793) and mutated (793(HNF4mut)) *OTC* promoter carrying mutation in the proximal binding site for HNF-4α, both subcloned to pGL3 vector containing enhancer. The data are presented as mean + SD. **B** -Sequence comparison of the *in silico* predicted human HNF-4α binding sites and their rat and mouse homologues with the matrix sequence motifs published by Fang et al. (2012). Grey boxes show the core sequence motif. "n" indicates equal incidence of all four nucleotides. Letters below the sequences in the "Matrix" section represent common (uppercase) or probable, but not common (lowercase) alternatives.

Tables and Figures

Table 1: Reporter constructs used in dual luciferase reporter assay

Construct	Promoter variant		Enhancer
	Fragment lenght	Position	
pGL3_793/Enh	793 bp	c.-838 .. c.-46	yes
pGL3_685/Enh	685 bp	c.-730 .. c.-46	yes
pGL3_458/Enh	458 bp	c.-503 .. c.-46	yes
pGL3_223/Enh	223 bp	c.-268 .. c.-46	yes
pGL3_173/Enh	173 bp	c.-218 .. c.-46	yes
pGL3_133/Enh	133 bp	c.-178 .. c.-46	yes
pGL3_81/Enh	81 bp	c.-126 .. c.-46	yes
pGL3_0/Enh	-	-	yes
pGL3-793/0	793 bp	c.-838 .. c.-46	-
pGL3_223/0	223 bp	c.-268 .. c.-46	-
pGL3_173/0	173 bp	c.-218 .. c.-46	-
pGL3_133/0	133 bp	c.-178 .. c.-46	-
pGL3_81/0	81 bp	c.-126 .. c.-46	-
pGL3_793(HNF4m)/Enh	793 bp	c.-838 .. c.-46	yes
pGL3_0/0	-	-	-

Figure 1

```
-838                                                                    CAAGCG
                                                                        |.||
-806                                                                    ctag--

-832 ACTCTCATGCCTCA-GCCTCCCAAATAGCTGGGA--TTACAGGTGTGCA-C-CA--CCACG-TCTAGCTAATTTTTGTATTTTTAGTG-GAGACGGGGAT
     ||||   |||.| |.||..||| ||.||.|| |||.||   .|| | ||  .||.| |.||||||  .||.|  |.|.|| ||.||   ||
-802 --tctc---cctgatgtctgacaa--agatgagaatttaaag----ccagctcatttcaagctgtagctaa---atgaa---tgaatgtgaaac----at

          685
-741 T--CACCATGTTGGCC--AGGCTGGT--CTCGAACT-CCTGGGCTCAAGTGATCCGCCCGCCTCAGCCTCCCAAAGTGCTGGGA---TTACA--GGCGTG
     |  | ||.||.||.|| |||  .||  || ||..| |.||||..|..|.|   |...|..|||  ||   .||.||.|||  |.||| ||..|.
-723 tggc-ccttgatgacctgagg--agttgct-gaggtgcttggggggcggggg----gggggggttca--ct-----ggtacttggaaactcacaccggaatc

                                                        HNF-4
-653 AGCCACCGTGCCCGGCC-AGCAATTATTTCTTTATTGAAGAC---TTATGTGCAAGGCACAAAGGGAG-CTCCAGGACTGAGATATTTTTACTATACCTT
     ||..|  ||||||  .|| | | |||  | ||   |  || ||.||  ||||||||.|||.||.||||||| | |.||.|||||||||.|.|||||||.|||||
-638 agtaa--gtgccc-accta-c--tta---c--ta---aacacctattatgtggaagacataaaggg-gccccctggactgagttctttttattatac---

-558 CTCTATCA--TCTTGCAC---C-CCCA---AAA--TAGCTTCCAGGGCACT-T--CTTTCTATTTGTTTTTGTGGAA-AGACTGGCAATTAG--AGGTAG
     ||||| || ||||||.||  |.|||   |||  .|.|||||.|| .|| | |||||||||||  ||||||| .|| |||.|.|  |  ||||||
-556 ctcta-cacctcttgaacttgctgccatggaaacccaactttgag---tctatcccttctatttg---ttgtggaacgga-tgggagt--gctaggtag
         100%
                                                          458
-475 AAAAGTGAAAT-AA-ATGGAAATAGTACTACTCAGGACTGTCACATCTACATCTGTGTTTTTGCAGTGCCAATTTGCATTTTCTGAGTGAGTTACTTCTA
     .||||||||||| || | |||| ||| |  |||| .|.||.|||||||.|.||.||||||.|.||||||.|.|||||||||||||||||.|||||||.||| |
-457 gaaagtgaaatcaaca-ggaa----tac-a--cagggctgtcatagctccatctgggggttttgtactgtcaatttacattttctgaatgagttatttc-a

-377 CT-CACCTTCACAGCAG--CCGGTACCGCAGTGCCTTGCATATATTATATC-CTCAATGAGTACTTGTCAATTGATTTTGTA--CATGCGTGTGACAGTA
     .|  .||||||.||.|| |.||.|||||||||||||||||||||| |||||||.|  | |||||||||.||||||||| ||  ||.||..|||||||||
-375 gtgtaccttcagagcagcccccagcacggcagtgccttgcatatattaca-cgctcaatgggtacttgtcaactgatttt-tatccaagcaagtgacagca

         223                                       C/EBPβ            110% 173                    HNF-4
-283 TAAATATATTATGAAAAATG-AGGAGGCCAGGCAATAAAAG-AGTCAGGA---TTCTTCCAAAAAAATACACAGCGGTGGAGCTT-GGCATAAAGTTC
     ||||||.||||||||||| ||| |||  ||.|..||||..||||  | |.|||  ||||||  .||||.||.| |||.|   |  .||| ||  |||.||.||||
-277 taaatacattatgaaaaatgaagg-gtctgggctctgaaagca-ttaggactctttctt-taaacaaca-accc--c---agag-ttaggcttaaagttc
                    320%                                                                  430%  HNF-4
            133       ▼                     HNF-4                  81 ▼      HNF-4
-189 AAATGCTCCTACACCCTGCCCTGCAGTATCTCTAACCAGG-GGACTTTGATAAGGAAGCT-GAAGGGTGATATTACCTTTGCTCCCTCACTGCAACTGAA
     ||.|||.|||.|||||||||||||||||||||.|.||||||||||||||||| .|| |.|||||||||..| | |||.|||.||.|||||||||.|||| |.|||||||||||||
-187 aagtgccctgcaccctgccctgctctatctctaaccaggttga-tctgataaggttg-tggaaaggtggtaggacctttgatccc-ctctgcaactgaa
        42%                                          50%         HNF-4          15%
                                                                                           -1
- 91 CA-CATTTCTTAGTTTTTAG---GTGGCCCCCGCTGGCTAACTTG-CTGTGGAGTTTTCAAGGGCATAGAATCGTCCTTTACACAAT-TA--AA-AGAAG
     | || |||||||.|| |    |||||||||.|||.|| .|.|| |||.|||.|.||.|...||||||.||.||.|||| | ||.| || || |||||
- 90 -agca-ttcttagctt---gccagtggcccccactgcct-gcctgcctgcggaactctctagaccatagattcctcct---c-cactctagcaagagaag
```

Figure 2

	81+/+	133+/+	173+/+	223+/+	458+/+	685+/+	WT0/+	81+/0	133+/0	173+/0	223+/0	WT0/0	WT+/0
81+/+		***	**	**	N	N	***	**	***	***	***	***	***
133+/+	***		N	N	***	N	***	***	***	***	***	***	***
173+/+	**	N		N	***	N	***	***	***	***	***	***	***
223+/+	**	N	N		***	N	***	***	***	***	***	***	***
458+/+	N	***	***	***		N	***	N	N	N	N	***	***
685+/+	N	N	N	N	N		***	***	***	***	***	***	***
WT0/+	***	***	***	***	***	***		***	**	N	N	***	N
81+/0	**	***	***	***	N	***	***		N	N	N	***	***
133+/0	***	***	***	***	N	***	**	N		N	N	***	**
173+/0	***	***	***	***	N	***	N	N	N		N	***	N
223+/0	***	***	***	***	N	***	N	N	N	N		***	N
WT0/0	***	***	***	***	***	***	***	***	***	***	***		***
WT+/0	***	***	***	***	***	***	N	***	**	N	N	***	

Figure 3

Figure 4

A
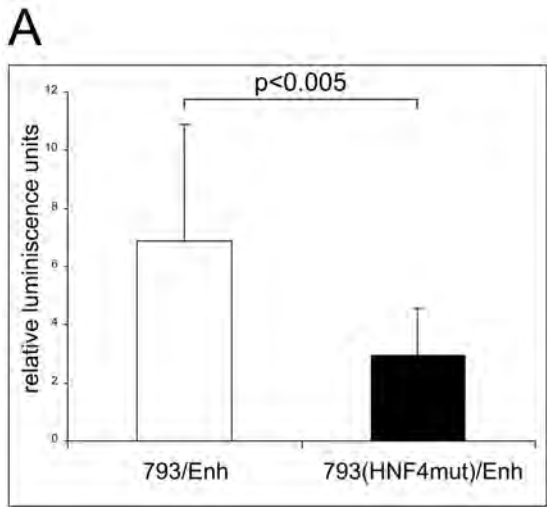
B

HNF-4alpha CORE SEQUENCE MOTIF

Very strong binders(top 10%)*	--GGGTCAAAGTTCA-- A GC G	Matrix
Strong binders (top 33%)*	--nGgnCAAAGGTCA-- ct TC G	
Middle binders (middle 33%)*	--nGnnCAAAGGTCA-- n tcgg TC n	
human -105..-121	AGGGAGCAAAGGTAATA	OTC promoter
human -136..-152	CCTTATCAAAGTCCCCT	
human -187..-203	TTGGCATAAAGTTCAAA	
human -588..-604	AAGGCACAAAGGGAGCT	
human -105..-121 (mut AAG>CCA)	AGGGAGCACCAGTAATA	
rat (homologous to -105..-121)	GGGGATCAAAGGTCCTA	
rat (homologous to -187..-203)	TAGGCTTAAAGTTCAAG	
mouse (homologous to -105..-121)	AGGGAGCAAAGGTCTTA	
mouse (homologous to -187..-203)	TAGGCTTAAAGTTCAAG	

Annex

Acceptance letter of article manuscript

Manuscript Ref.:
- **title:** HNF-4alpha regulates expression of human ornithin transcarbamylase through interaction with two positive cis-acting regulatory elements located in the proximal promoter
- **author:** O. Lukšan
- **number:** FB2013-719M
- **submitted by:** Ondřej Lukšan
- **status:** ACCEPTED
- **editor:** Zdenek Kostrouch

Corresponding author:

Ondrej Luksan, Institute for Clinical and Experimental Medicine (IKEM), Laboratory of Experimental Hepatology, Videnska 1958/9, Praha 4 - 140 21, Czech Republic
E-mail: ondrej.luksan@ikem.cz

February 14, 2014

Dear Dr. Lukšan,

Your revised manuscript has been accepted for publication in Folia Biologica provided that all of the authors fill and sign the Authors Declaration Form which has to be mailed or FAXED to the Folia Biologica Editorial Office.

The manuscript will be further processed by the Folia Biologica editorial staff in order to comply with formal and linguistic requirements of the journal.

You will obtain proofs for a final control of the manuscript. This is restricted only to minor changes. Please contact the Folia Biologica Editorial Office if you have any further questions or difficulties with proof reading.

Thank you for publishing your fine research in Folia Biologica.

Sincerely yours,

Zdenek Kostrouch

4.4. Glucocerebrosidase gene has an alternative upstream promoter, which has features and expression characteristic of housekeeping genes

4.4.1. Commentary

There are five alternative *GBA* transcripts annotated in the databases, four of which apparently originate upstream of the known promoter (P1). These transcripts, differing in the presence of either one or two extra exons –2 and –1, are not predicted to alter the amino-acid sequence of the *GBA* as their variable 5′ ends do not contain initiation codons. Their expression should be apparently driven by an alternative upstream promoter (P2) located 2.6 kb above the translation origin.

In this study, we have shown that the region 1 kb upstream in the 5'-flanking region of exon –2 can function as an alternative promoter of *GBA* and we also analyzed its properties. For that purpose we applied a similar methodical approach as we introduced and optimized within the OTC project.

In a luciferase reporter assay we confirmed the transcriptional activity of both P1 and P2 promoters in Hep-G2 cells. Using the method of 5′ serial deletions we delimited the P2 core promoter and characterized the promoter regions exhibiting *cis*-active regulatory activity. We identified three alternative transcription start sites located 347, 380, and 413 bp upstream of the translation initiation site. Furthermore we studied the expression profile of transcripts originating from P2 and P1 across 20 human tissues and we found that expression is more stable than in the case of housekeeping genes *GAPDH* and *ACTB*. *In-silico* analyses of the P2 promoter revealed the lack of a TATA box (unlike P1) and presence of an unmethylated CpG island and multiple Sp-1 consensus binding sites.

Finally, we analyzed both promoters and the untranslated region in DNA samples from a phenotypically diverse group of twenty Ashkenazi Jewish Gaucher patients homozygous for the common mild mutation p.N370S. Sequence variations explaining phenotypic differences were found neither in P1 and P2 nor exons –2 and –1.
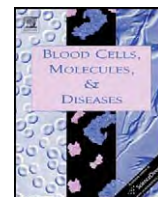
Glucocerebrosidase gene has an alternative upstream promoter, which has features and expression characteristic of housekeeping genes

Eva Svobodová [a], Lenka Mrázová [a], Ondřej Lukšan [b], Deborah Elstein [c], Ari Zimran [c], Larisa Stolnaya [a], Jakub Minks [a,1], Jitka Eberová [b,2], Lenka Dvořáková [a], Milan Jirsa [b], Martin Hřebíček [a,*]

[a] Institute of Inherited Metabolic Disorders, Charles University, 1st Faculty of Medicine and University Hospital, Ke Karlovu 2, Prague 120 08, Czech Republic
[b] Laboratory of Experimental Hepatology, Institute of Clinical and Experimental Medicine, Vídeňská 1958/9, Prague 140 21, Czech Republic
[c] Gaucher Clinic, Shaare Zedek Medical Center, PO Box 3235, Jerusalem 910 31, Israel

ARTICLE INFO

ABSTRACT

Database searches have shown that a part of glucocerebrosidase (GBA) transcripts may originate at an alternative upstream promoter (P2) located 2.6 kb upstream of the known (P1) GBA promoter. The putative alternative transcripts contained one or two extra exons (exon -2 or exons $-2, -1$, respectively), but the first ATG codon and predicted amino-acid sequence are the same as in the transcript from P1. Luciferase assays confirmed promoter activity of both sites in HepG2 cells: the P1 construct exhibited the highest activity of luciferase (17.82 ± 1.10 relative luciferase units), while the P2 construct reached 3.01 ± 0.43 relative luciferase units. Serial 5′ deletions of P2 led to changes in reporter activity, the most prominent decreases were observed in deletion constructs carrying bases -353 to -658, and -353 to -920 (numbered as in NM_001005750.1), respectively. This suggests that the P2 core promoter is contained within the region of -920 bp to -1311 bp.

Three P2 transcription initiation sites were found by 5′ RACE at positions 347, 380, and 413 bp upstream of the $+1$ ATG. The expression stability of transcripts from P2, P1 was studied in 20 human tissues and was higher than that of GAPDH and ACTB, which are commonly used as reference housekeeping genes. The P2 contains an unmethylated CpG island, multiple Sp-1 consensus binding sites and, unlike P1, does not contain a TATA box, features all common to the majority of housekeeping gene promoters.

We have examined DNA samples from a phenotypically diverse group of twenty Ashkenazi Jewish Gaucher patients homozygous for the common mild mutation N370S. Both P1 and P2, as well as exons -2 and -1, did not contain any sequence variations, with the exception of the known polymorphism rs10908459 found on one allele. The phenotypical differences in the patients were thus not explained by nucleotide variations in both promoters.

© 2011 Elsevier Inc. All rights reserved.

Introduction

Gaucher disease is a glycolipid storage disorder caused by an inherited deficiency of activity of glucocerebrosidase, a hydrolase that catalyzes one step in lysosomal degradation of glycosphingolipids [1]. Gaucher patients carry mutations in the gene encoding glucocerebro-

sidase (GBA) located in a gene-rich region at 1q21; the gene spans 7.8 kb and contains 11 exons and 10 introns. A highly homologous transcribed but untranslated pseudogene (psGBA), which is a result of a tandem duplication event involving ancestral GBA and neighbouring metaxin genes, is located 16 kb downstream from GBA [2]. This duplication, which probably occurred before the divergence of great apes and Old World monkeys, has been found only in primates and humans [3].

Although glucocerebrosidase is expressed in all cell types – a feature characteristic of housekeeping genes – differences in its expression levels in cell lines suggested at least some degree of differential regulation of expression. Also, glucocerebrosidase activity to mRNA levels ratios has not been concordant in all examined cell lines. Because of this, it has been suggested that in addition to regulation by mRNA levels, GBA activity must be controlled also by other mechanisms [4,5].

The majority of housekeeping genes characteristically have TATA-less promoters containing CpG islands and are enriched in binding

sites for some transcription factors, including Sp1 [6]. The glucocerebrosidase promoter located immediately upstream of exon 1 (P1) does not have an associated CpG island and contains a TATA box [2]. Other regulatory elements were identified in the promoter, whose ability to drive expression of a reporter gene was shown in several human cell lines. Binding of transcription factors was confirmed by gel-shift assays. It has been elucidated, by deletion analysis, that there are several enhancers and a repressor [7].

Currently, there are five alternative transcripts annotated in the databases, four of which apparently originate at an alternative promoter (P2) located 2.6 kb upstream of the first ATG and contain either one or two extra exons (exon −2 or both exons −2 and −1, respectively) (Fig. 1). These transcripts are not predicted to alter the amino-acid sequence of the glucocerebrosidase as their variable 5′ ends do not contain initiation codons.

In this study, we show that the region 1 kb upstream of exon −2 can function as an alternative promoter of glucocerebrosidase and we also analyze its properties. We found transcripts from P2 promoter in multiple tissues, confirming that the promoter is used in vivo.

We have also examined the sequence of exons −2, −1 and the P1 and P2 promoters in a group of Ashkenazi Jewish Gaucher patients homozygous for the N370S mutation [8] who presented with discordant phenotypes.

Patients and methods

Patients and genotyping

A group of 20 Ashkenazi Jewish type 1 Gaucher patients homozygous for the N370S mutation were included in the study on the basis of written informed consent. Ten of them had mild Gaucher phenotype, which is the usual presentation of the disease in patients with this genotype, while the remaining patients had severe form of the disease. The P1, P2, exons −1 and −2 were amplified from genomic DNA in each patient and sequenced. Primer sequences are in Table 1.

Numbering of sequences

The promoter and cDNA sequences were numbered beginning with the A in the initiation ATG codon according to Reference Sequence NM_000157.2 for the P1 promoter and Reference Sequence NM_001005750.1 for the P2 promoter. No base was assigned to zero position.

Sequencing of RT-PCR products

GBA transcripts originating at P2 were probed by RT-PCR using primers U4 and L9 (Table 1). Upstream primer annealed to exon −2, while downstream primer annealed in exon 9 to the sequence deleted in the pseudogene, making the amplification specific for the active

gene. The products were gel-purified, cloned, and individual clones were sequenced.

Generation of reporter gene constructs

PCR products (921 and 1156 bp long, respectively) containing the sequence of P1 (−948/−27, NM_000157.2; NG_009783.1, g.7639–8559) or P2 (−1509/−353, NM_001005750.1, NG_009783.1, g.3892–5047) promoters were amplified using primers with overhangs containing sequences recognized either by KpnI or XhoI restrictases (forward and reverse primers, respectively, Table 1). Primer sequences were derived from genomic sequence of chromosome 1 (Reference Sequence NG_009783.1). The fragments were gel-purified and directly cloned in the TA cloning vector pCR®4 -TOPO® (Life Technologies (Invitrogen, Carlsbad, USA), and the resulting constructs were verified by sequencing. Positive clones were double digested with KpnI and XhoI (Fermentas), released fragments were cloned in both sense and antisense orientations into pGL4.16 (luc2CP/Hygro) vector (Promega, Madison, USA) upstream of the firefly luciferase reporter gene, generating pGL4 −1509/−353 and pGL4 −948/−27 constructs, respectively.

Generation of serial deletions in P2 constructs

Serial deletions of P2-containing construct pGL4 −1509/−353 were created by site-directed mutagenesis using the QuickChange XL site-directed mutagenesis kit (Stratagene, La Jolla, USA) [9]. Primers listed in Table 1 were used to generate mutant constructs with different 5′end deletions (pGL4 −1311/−353, pGL4 −1164/−353, pGL4 −920/−353 and pGL4 -658/−353) according to the manufacturer's instructions. All deletions were confirmed by DNA sequencing.

Cell culture, transfection and reporter gene analysis

HepG2 (human hepatoblastoma) cells cultured in Opti-MEM® (Invitrogen) were supplemented with 10% (v/v) fetal bovine serum. The cells were grown in 25 cm^2 flasks at 37 °C, 5% CO$_2$. A total of 5×10^4 HepG2 cells per well were seeded into a 24-well culture plate 1 day prior to transfection. 500 ng of each construct was transfected into HepG2 cells using the Tfx™–20 Reagent (Promega). The empty pGL4 vector was transfected simultaneously. The pRL-TK vector (Promega) harboring the Renilla luciferase gene was co-transfected as an internal control to normalize for the transfection efficiency. Each construct was transfected in triplicate, and each transfection experiment was repeated independently at least three times. After 48 h, cells were washed with phosphate-buffered saline (PBS) and lysed with 100 μl of the passive lysis buffer (Promega). The luciferase reporter gene activity was assayed using the Dual-Luciferase® Reporter Assay System (Promega). The intensity of chemiluminescence in the supernatant was measured using a luminometer (Berthold, Bad
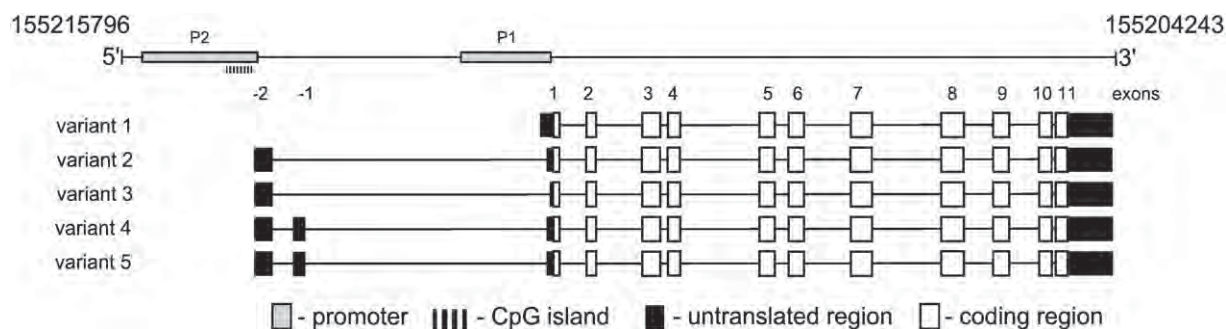

Fig. 1. Alternative transcripts of *GBA*. *GBA* transcripts according to GeneBank reference sequences: variant 1 (Acc.No. NM_000157.2) variant 2 (Acc.No.NM_001005741.1), variant 3 (Acc.No.NM_001005742.1), variant 4 (Acc.No.NM_001005749.1), and variant 5 (Acc.No.NM_001005750.1). The position of P1, P2, and of the P2-associated CpG island are shown.

Table 1
Primers used for amplifications and generation of deletions constructs (see Patients and methods).

Amplified region	Forward primer	5′Position	Reverse primer	5′Position	Annealing temp. (°C)	Length (bp)
P1_sense[a]	5′-GGTACCGTGTGCCACTCCCGCTAAATC-3′	g. 7638 (NG_009783.1)	5′-CTCGAGAGACCACAGGGGTTCCAGAGT-3′	g. 8559 (NG_009783.1)	67	921
P1_antisense[a]	5′-CTCGAGGTGTGCCACTCCCGCTAAATC-3′	g. 7638 (NG_009783.1)	5′-GGTACCAGACCACAGGGGTTCCAGAGT-3′	g. 8559 (NG_009783.1)	67	921
P2_sense[a]	5′-GGTACCTTTCAGTGAGCACCCAATCCC-3′	g. 3892 (NG_009783.1)	5′-CTCGAGGAGAAAAGCAGCCCTGGGGAG-3′	g. 5047 (NG_009783.1)	67	1156
P2_antisense[a]	5′-CTCGAGTTTCAGTGAGCACCCAATCCC-3′	g. 3892 (NG_009783.1)	5′-GGTACCGAGAAAAGCAGCCCTGGGGAG-3′	g. 5047 (NG_009783.1)	67	1156
Exons −1, −2	5′-TGAGGAAGGGCTCTGAGTCC-3′	g. 4958 (NG_009783.1)	5′-TCAAGCGGAGGGTAGGGACC-3′	g. 5828 (NG_009783.1)	65	871
U4, L9	5′-GCTGCTGTTTCTCTTCGCCGAC-3′	g. 5170 (NG_009783.1)	5′-GGTCCTCCTTCGGGGTTCA-3′	c. 1685 (NM_001005750.1)	63	1515
−353/−1311[b]	5′-CCTAACTGGCCGGTACTAGCTGGGATTACAGG-3′	g. 4090 (NG_009783.1)			60	
−353/−1164[b]	5′-CCTAACTGGCCGGTACATTACAGGTGTGACTC-3′	g. 4237 (NG_009783.1)			60	
−353/−920[b]	5′-CTAACTGGCCGGTACAACGTGGTGCCTCC-3′	g. 4481 (NG_009783.1)			60	
−353/−658[b]	5′-CCTAACTGGCCGGTACTGTGGGAATTCAATCG-3′	g. 4743 (NG_009783.1)			60	
U1, L1	5′-CCTGAGGCTCTCTAAAGAGGAAGGTTAGGAA-3′	g. 4683 (NG_009783.1)	5′-GAGTCTCTGAAGGATAGAGGATCCACGTTAA-3′	c. 323 (NM_001005750.1)	65	674
U2, L1	5′-CGACACGCATGCGTAGTTCTC-3′	g. 4795 (NG_009783.1)	5′-GAGTCTCTGAAGGATAGAGGATCCACGTTAA-3′	c. 323 (NM_001005750.1)	65	562
U3, L1	5′-CCAGCGACACTTGTTCGTTCAACTT-3′	g. 4920 (NG_009783.1)	5′-GAGTCTCTGAAGGATAGAGGATCCACGTTAA-3′	c. 323 (NM_001005750.1)	65	437
U4, L1	5′-GCTGCTGTTTCTCTTCGCCGAC-3′	g. 5170 (NG_009783.1)	5′-GAGTCTCTGAAGGATAGAGGATCCACGTTAA-3′	c. 323 (NM_001005750.1)	65	187
U1, L2	5′-CCTGAGGCTCTCTAAAGAGGAAGGTTAGGAA-3′	g. 4683 (NG_009783.1)	5′-AAGACCACAGGGGTTCCAGAG-3′	c. 351 (NM_001005750.1)	65	692
U2, L2	5′-CGACACGCATGCGTAGTTCTC-3′	g. 4795 (NG_009783.1)	5′-AAGACCACAGGGGTTCCAGAG-3′	c. 351 (NM_001005750.1)	65	580
U3, L2	5′-CCAGCGACACTTGTTCGTTCAACTT-3′	g. 4920 (NG_009783.1)	5′-AAGACCACAGGGGTTCCAGAG-3′	c. 351 (NM_001005750.1)	65	455
U4, L2	5′-GCTGCTGTTTCTCTTCGCCGAC-3′	g. 5170 (NG_009783.1)	5′-AAGACCACAGGGGTTCCAGAG-3′	c. 351 (NM_001005750.1)	65	205
Meth1	5′-GGTTGGATTTATTGTGGGAATTTAAT-3′	g. 4731 (NG_009783.1)	5′-TCAAAACCCTTCCTCAAATCTCATT-3′	g. 4973 (NG_009783.1)	56	242
Meth2	5′-AATGAGATTTGAGGAAGGGTTTTGAG-3′	g. 4949 (NG_009783.1)	5′-CCATTACACTCCAACCTAAAC-3′	g. 5340 (NG_009783.1)	56	391
Meth3	5′-GTTTAGGTTGGAGTGTAATGG-3′	g. 5320 (NG_009783.1)	5′-TCATTAATAAAATCTAACCATCCTTACC-3′	g. 5630 (NG_009783.1)	56	310

Primer positions correspond to 5′end of the primers, either in genomic or cDNA reference sequence.

[a] Underlined sequence is a non-hybridizing overhang.

[b] Primer used for generation of serial deletions in P2 insert by site-directed mutagenesis. Underlined sequence anneals to the vector pGL4, while the rest of the primer anneals to P2. The position indicates the 5′end of the P2-hybridizing sequence.

Wildbad, Germany) and expressed as the ratio of Renilla luciferase to firefly luciferase luminescence (RLU, relative luciferase units).

Identification of the transcription initiation site

PCR

RT-PCR was used to approximate the position of the transcription initiation site. Four upper primers and two lower primers were designed (Table 1). PCR products were gel-purified and sequenced. Upper primers were designed to anneal to the presumed 5′ ends of the transcript. Lower primers annealed to the junction of exon −1 and exon 1 and in the exon 1, respectively (Fig. 2).

5′ RACE

Transcription initiation sites of the gene were identified by 5′ RACE technique that amplifies products only from capped mRNAs [10]. 5′
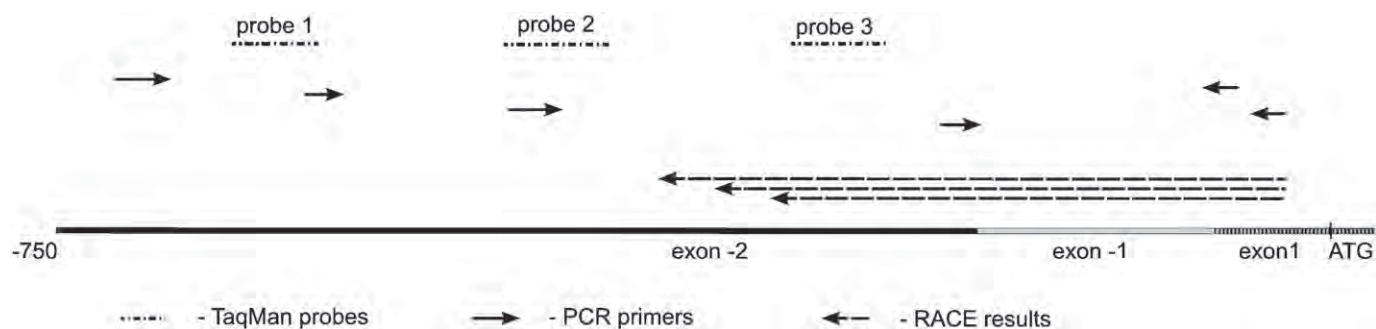

Fig. 2. Location of TaqMan probes and primers relative to transcription initiation sites.

RACE was carried out using the GeneRacer™ Kit (Invitrogen) with RNAs isolated from human leukocytes, liver, and placenta. Alternatively, the FirstChoice® RLM-RACE Kit (Ambion (Life Technologies), Austin, USA) was used with the FirstChoice® Human Placenta Total RNA (Ambion). Nested PCR was performed with 5′ RACE universal primers and gene-specific primers. Primers 5′-aagaccacaggggttcca-gag-3′ or 5′-gagtctctgaaggatagagagatccacgttaa-3′ were used for the first round of amplification, while 5′-caggaattggagaccagcctgac-3′ and 5′-ggcaaaacgaaatcccaccgcag-3′, for the second, respectively. Obtained cDNA fragments were gel purified and sequenced directly and/or after cloning into TA cloning vector pCR®4 -TOPO® (Invitrogen).

Real-time PCR

The utilization of transcription initiation sites was studied by quantitative RT-PCR. Primers and TaqMan probes were designed using Primer Express 3.0 software (Applied Biosystems (Life Technologies), Foster City, USA) to be specific for the sequence upstream (probe 1 and probe 2) and downstream (probe 3) of transcription initiation sites identified by 5′ RACE (Fig. 2, Table 2). Amplification conditions are available upon request. All reactions were performed using StepOne™ Real-time PCR system (Applied Biosystems). cDNA was obtained by reverse transcription of Human Placenta Total RNA (Ambion) using High Capacity RNA-to-cDNA Kit (Applied Biosystems). A standard curve was generated by tenfold serial dilutions of the gDNA for each assay, yielding correlation coefficients ≥ 0.98 in all experiments. Each standard and sample value was determined in duplicate. The expression of targets of all three assays was quantified from the standard curve.

Expression profile of P2 and P1 transcripts

TaqMan probes 3 and 4, respectively (Table 2), were used for the determination of expression stability of promoters P2 and P1. The specificity of probe 4 was verified by absolute quantification using oligonucleotides complementary to GBA gene and pseudogene (Generi Biotech, Prague, Czech Republic, sequence not shown). Expression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β-actin (ACTB) was quantified with probes purchased from Applied Biosystems (Cat. No. 4326317E and 4326315E, respectively). cDNAs were reverse transcribed from FirstChoice® Human Total RNA Survey Panel (Ambion) as described above. Purified RNAs were precisely quantified by the manufacturer and declared concentrations of RNA were used. Negative controls without template were included in each plate. Amplification efficiencies of all probes were tested by construction of standard curves using serial dilutions of cDNA and ranked between 91% and 114%.

The expression stability of GBA and commonly used housekeeping genes GAPDH and ACTB were evaluated using geNorm (version 3.5) and NormFinder software packages, which examine expression stability of genes using different algorithms [11,12]. Ct values were transformed to relative quantities using delta Ct method assuming the amplification efficiency equal to 2 in all cases.

DNA methylation

The methylation status of the CpG island overlapping with P2 was studied using bisulphite sequencing. Bisulphite-modified genomic

DNA [13] of three control individuals served as a template for amplification of the CpG-rich sequence in the vicinity of P2 in three fragments. Primers designed to anneal to the modified DNA are listed in Table 1. Gel-purified products were sequenced.

Bioinformatics—identification of putative transcription factor binding sites

The search for putative transcription factor binding sites was performed using TFSEARCH (http://www.cbrc.jp/research/db/TFSEARCH.html), MatInspector program (http://www.genomatix.de/shop/evaluation.html) [14] and Alibaba2 (http://www.gene-regulation.com/pub/programs.html#alibaba2).

GBA reference transcripts were obtained from NCBI (http://www.ncbi.nlm.nih.gov/refseq/), the sequences were last accessed on 26th August 2010). Expressed sequence tags (ESTs) containing sequence homologous to exons −2 and −1 were identified using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and/or UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

The multiple alignment of −2 exon and P2 region (Supplementary data) was modified from 44-way alignment of assembled genomes available from Galaxy (http://main.g2.bx.psu.edu/) [15].

Results

The constructs containing the sequence of both P1 (in pGL4_P1+) and P2 (in pGL4_P2+), respectively, delivered significantly higher levels of reporter activity than negative controls with the same sequence in antisense orientation as well as the empty vector. The P1 construct exhibited the highest activity of luciferase (17.82 ± 1.10 relative luciferase units), while the P2 construct reached 3.01 ± 0.43 relative luciferase units (Fig. 3).

The pGL4 −1509/−353 containing the whole sequence of P2 and the deletion construct pGL4 −1164/−353 displayed comparable levels of luciferase activity, but the other deletion constructs led to changes in reporter activity. The highest activity was observed in pGL4 −1311/−353, while pGL4 −658/−353 exhibited dramatically lower luciferase activity. The lowest activity was observed in pGL4 −920/−353 (Fig. 4). This suggests that the P2 core promoter is located within the region ranging from −920 bp to −1311 bp. The region −1311 bp to −1509 bp probably contains a negative regulator of transcription activity.

RT-PCR products corresponding to transcript variants containing exon −2 or both exons −1 and −2 were found by RT-PCR in cDNA prepared from placental RNA. 5′ RACE identified three transcription initiation sites at the P2, corresponding to the positions 347, 380 and 413 bp upstream of the +1 ATG (Fig. 2, Supplementary data). Most of the clones had 5′ends at the position −347 and only few clones at the position −413. Two clones (of the total of 40) were at positions −452 and −563, respectively. We were, however, able to amplify weak RT-PCR products up to position −500 bp. RT amplifications with primers annealing upstream of this region did not yield any products.

The levels of transcripts initiating upstream of sites identified by 5′ RACE were studied by quantitative RT-PCR. The signal from probe 3 (downstream of −347) was 300 times higher than that of probe 1 (upstream of −563); the signal from probe 2 (upstream of −413) was only twice as high as the signal from probe 1. This result

Table 2
TaqMan probes and primers.

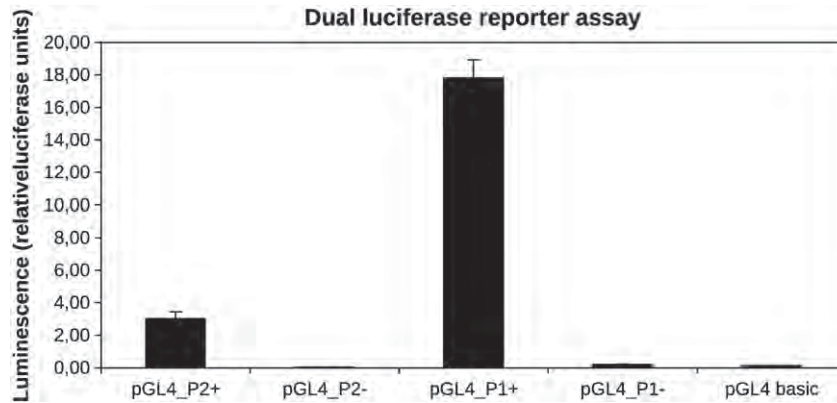| TaqMan assay	Forward primer	T_m (°C)	Reverse primer	T_m (°C)	Probe	T_m (°C)
promIIGBA1	5′-AGAGTCTTACTGCGCGGGG-3′	59	5′-AGCCTGCAAAGGCGCC-3′	60	5′-AGTCTCCAGTCCCGCC-3′	69
promIIGBA2	5′-GCTCGCCAGCGACACTTG-3′	60	5′-GGACTCAGAGCCCTTCCTCAA-3′	59	5′-CGTTCAACTTGACCAATG-3′	69
promIIGBA3	5′-TCAATCGCCCCCATCCA-3′	60	5′-TGCGTGTCGGCGTTTTC-3′	59	5′-CAACAGTGTGCTGGCG-3′	70
promIGBA4	5′-CGGAATTACTTGCAGGGCTTA-3′	58	5′-CAAAAACAAGGATGCAGGTACCT-3′	58	5′-CCTAGTGCCTATAGCTAAG-3′	69

Fig. 3. Reporter activity of P1 and P2 of the GBA. Both the P1 (pGL4_P1+) and P2 (pGL4_P2+) reporter luciferase activities are shown in RLU (relative luciferase unit). The RLU activity of each construct is plotted as fold increases over that of the pGL4 vector. Error bars represent standard deviation. The pGL4 vector, and the constructs with P1 (pGL4_P1-) and P2 (pGL4_P2-) promoters cloned in antisense orientation, were used as a negative control and exhibited negligible luciferase activity.

corresponds well with 5′ RACE, where most clones had 5′end at positions −347, −380, −413 (Fig. 2).

The expression stability of transcripts from P2, P1, and two housekeeping genes (GAPDH and ACTB) was studied in 20 human tissues. The Ct values determined by real-time PCR for the mRNAs from P1, P2, GAPDH, and ACTB are shown in Fig. 5. Both geNorm and NormFinder found the expression stabilities of P1 and P2 transcripts comparable, while both were higher than stabilities of the established housekeeping genes GAPDH and ACTB (Table 3). The expression profile of P1 and P2 followed the same trend among tissues (Fig. 5). The results of bisulphite sequencing showed that the CpG island was unmethylated (data not shown).

The multiple sequence alignment showed that P2 is conserved in primates. The best homology was obtained in chimpanzee, gorilla, and orangutan. Exon −1 homology was found only in gorilla. Analysis of P2 sequence predicted consensus binding sites for a number of transcription factors. No consensus TATA or CAAT boxes were identified, but multiple Sp1 binding sites were found. In addition, conserved binding sites for several other transcription factors, GATA-1, AP-1, YY1, CREB, CRE-BP, E2F, and two CLEAR motifs were also found (Supplementary data).

The results of sequencing study in Ashkenazi Jewish Gaucher patients homozygous for the N370S did not support our hypothesis that sequence variations in P1, P2, exons −2 and −1 may influence Gaucher phenotype. No sequence variations have been found in the

patients with the exception of a known polymorphism rs10908459 (dbSNP build 131, http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp) found in heterozygosity in one of the patients.

Discussion

Four of the five known glucocerebrosidase transcripts contain exon −2 (variants 2, 3, 4, and 5) or both exon −2 and exon −1 (variants 4 and 5). Transcripts 4 and 5 differ in three nucleotides at the junction of exons −2 and −1 and transcripts 2 and 3 differ in nineteen nucleotides at the junction of exon −1 and 1 (Fig. 1). The existence of these variant transcripts is supported by a number of full-length clones and ESTs in databases (not shown). Horowitz and co-workers [2] confirmed that the region immediately upstream of exon 1 (P1 promoter) functions as a promoter of glucocerebrosidase from which variant 1 mRNA is transcribed.

In this paper, we show that the sequence upstream of exon −2 (P2 promoter) drives transcription of mRNA in the in vitro reporter assay and the deletions of parts of the region lead to severe reduction of the reporter activity. The P2 promoter fragment drove the transcription of the reporter gene markedly higher than the control plasmid pGL4; the reached levels of luminescence, however, were about six times lower than that from P1 under the same conditions. This may suggest that the P2 is a weaker promoter than P1. Nonetheless, it is important to note that the fragment may not have contained all functionally
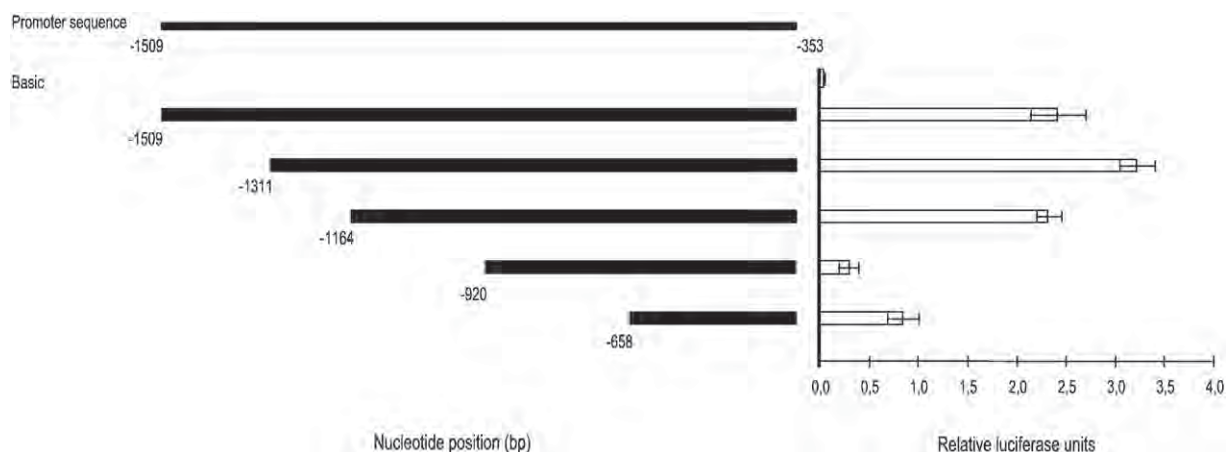

Fig. 4. 5′deletion analysis of the 5′flanking region of the GBA gene. The RLU (relative luciferase activity) of each deletion construct is plotted as fold increases over that of the pGL4 vector. The results are represented as the mean ± S. D. of three separately transfected wells in one experiment. The pGL4 basic vector was used as a negative control and exhibited no luciferase activity.
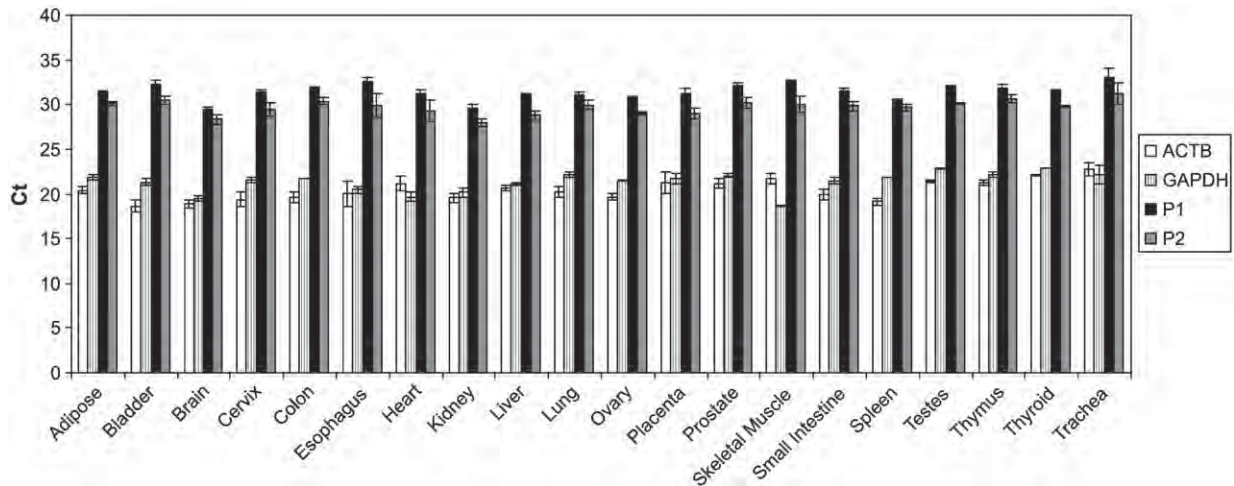
Fig. 5. Cycle threshold values of P1, P2, GAPDH and ACTB in different tissues. The expression levels of P1, P2, GAPDH and ACTB in 20 different tissue samples were evaluated using qRT-PCR. Ct values in P1 are in the range of 29.4 – 32.9 and in P2 27.9 – 30.6. Error bars represent standard deviation ($n=2$).

important elements of the promoter, including distant enhancers as well as all regulatory sequences downstream of the initiation site. We used fragments of similar size, which did not contain a larger part of the downstream exons in both P1 and P2 constructs. For these reasons, reporter gene expression from P1 and P2 constructs may not be proportional to the strength of the native promoters.

Analysis of 5′ RACE products has shown that transcripts originate at three initiation sites (347, 380, and 413 bp upstream of the +1 ATG). We have confirmed by quantitative RT-PCR that the above sites are the major transcription initiation sites of P2. The weak RT-PCR products originating upstream of these sites in the 5′ region may represent promoter upstream transcripts (PROMPTs) [16].

Sardiello and co-workers [17] have recently found that most lysosomal genes exhibit coordinated transcriptional behavior and are regulated by the transcription factor EB (TFEB). They identified a 10-bp motif named CLEAR (Coordinated Lysosomal Expression and Regulation), which is enriched in promoters of lysosomal genes. P1 promoter of *GBA* contains two such elements and overexpression of TFEB led to approximately 3 times higher expression of GBA. We have found two elements compatible with the CLEAR consensus sequence in P2, located −417 and −891 bp from TSS found by 5′ RACE (347 bp upstream of the +1 ATG) (Supplementary data). While CLEAR elements are usually found within 200 bp of the transcription start site (TSS), more distal CLEAR elements were found in promoters upregulated by TFEB.

Multiple Sp-1 binding sites, an unmethylated CpG island, and the absence of a TATA box are features that the P2 has in common with the majority of housekeeping promoters. The P1 does not share these properties, and the presence of the TATA box may suggest that P1 is a regulated promoter [4,5]. However, the results from the quantitative RT-PCR in RNA from normal adult human tissues did not show significant differences in expression stability between the both promoters. Moreover, the expression stability of P1 and P2 appears to be higher than commonly used reference housekeeping genes GAPDH and ACTB. Earlier studies showed that levels of glucocer-

ebrosidase mRNA differ between cell lines of different origin, suggesting that glucocerebrosidase expression may at least partially be regulated. While the mRNA levels differed between cell lines, they did not always correlate with enzyme activity, suggesting that other mechanisms are also important in regulation of glucocerebrosidase activity [4]. These results are not directly comparable with our observations, because we have measured the amount of glucocerebrosidase transcripts in total RNA extracted from normal human tissues which contain a mixture of cells of different origin. Importantly, the lines employed in the above studies included also immortal cancerous cell lines that may not faithfully reflect properties of the underlying cell type.

Transcripts from both promoters were detected in all studied tissues. While we did not measure the absolute levels of transcripts from each promoter, the ratio between P1 and P2 expressions appeared to be similar among tissues.

Although we did not find significant differences in expression from both promoters, steady-state levels of mRNA in normal adult tissues do not exclude that one or both glucocerebrosidase promoters may be differentially regulated under certain conditions including Gaucher disease.

We hypothesized that differences in the manifestation of Gaucher disease in patients with the same genotype may be caused by variable expression of mutant glucocerebrosidase due to variants in the 5′ flanking sequences and the promoters. However, in the group of patients homozygous for the common mild mutation N370S with discordant phenotypes we found only one known polymorphism (rs10908459) in heterozygosity in one case. It remains to be elucidated what role the P2 promoter plays in the regulation of glucocerebrosidase and possibly also in the pathogenesis of Gaucher disease.

Acknowledgments

Appendix A. Supplementary Data

Supplementary data to this article can be found online at doi:10.1016/j.bcmd.2010.12.011.

Table 3
Expression stability coefficients of P1, P2, GAPDH, and ACTB.

Gene	GeNorm	NormFinder
P1	0.923	0.376
P2	0.900	0.397
GAPDH	1.141	0.631
ACTB	1.290	0.740

References

[1] E. Beutler, G.A. Grabowski, Gaucher disease, in: C.R. Scriver, A.L. Beaudet, D. Valle, W.S. Sly (Eds.), The Metabolic and Molecular Bases of Inherited Disease, McGraw-Hill, New York, 2001, pp. 3635–3668.

[2] M. Horowitz, S. Wilder, Z. Horowitz, et al., The human glucocerebrosidase gene and pseudogene: structure and evolution, Genomics 4 (1989) 87–96.

[3] S.L. Winfield, N. Tayebi, B.M. Martin, E.I. Ginns, E. Sidransky, Identification of three additional genes contiguous to the glucocerebrosidase locus on chromosome 1q21: implications for Gaucher disease, Genome Res. 7 (1997) 1020–1026.

[4] R.F. Doll, F.I. Smith, Regulation of expression of the gene encoding human acid beta-glucosidase in different cell types, Gene 127 (1993) 255–260.

[5] R.F. Doll, A. Bruce, F.I. Smith, Regulation of the human acid beta-glucosidase promoter in multiple cell types, Biochim. Biophys. Acta 1261 (1995) 57–67.

[6] W.S. Dynan, S. Sazer, R. Tjian, R.T. Schimke, Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter, Nature 319 (1986) 246–248.

[7] Y.N. Blech-Hermoni, S.G. Ziegler, K.S. Hruska, et al., In silico and functional studies of the regulation of the glucocerebrosidase gene, Mol. Genet. Metab. 99 (2010) 275–282.

[8] S. Tsuji, B.M. Martin, J.A. Barranger, et al., Genetic heterogeneity in type 1 Gaucher disease: multiple genotypes in Ashkenazic and non-Ashkenazic individuals, Proc. Natl Acad. Sci. USA 85 (1988) 2349–2352.

[9] A.R. Shenoy, S.S. Visweswariah, Site-directed mutagenesis using a single mutagenic oligonucleotide and DpnI digestion of template DNA, Anal. Biochem. 319 (2003) 335–336.

[10] M. Fromont-Racine, E. Bertrand, R. Pictet, T. Grange, A highly sensitive method for mapping the 5′ termini of mRNAs, Nucleic Acids Res. 21 (1993) 1683–1684.

[11] RESEARCH0034. J. Vandesompele, K. De Preter, F. Pattyn, et al., Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002).

[12] C.L. Andersen, J.L. Jensen, T.F. Orntoft, Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets, Cancer Res. 64 (2004) 5245–5250.

[13] S.J. Clark, J. Harrison, C.L. Paul, M. Frommer, High sensitivity mapping of methylated cytosines, Nucleic Acids Res. 22 (1994) 2990–2997.

[14] K. Cartharius, K. Frech, K. Grote, et al., MatInspector and beyond: promoter analysis based on transcription factor binding sites, Bioinformatics 21 (2005) 2933–2942.

[15] J. Goecks, A. Nekrutenko, J. Taylor, Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11 R86.

[16] P. Preker, J. Nielsen, S. Kammler, et al., RNA exosome depletion reveals transcription upstream of active human promoters, Science 322 (2008) 1851–1854.

[17] M. Sardiello, M. Palmieri, A. di Ronza, et al., A gene network regulating lysosomal biogenesis and function, Science 325 (2009) 473–477.

5. General discussion

The diagnosis of inherited metabolic disorders is based on clinical suspicion and biochemical and molecular genetic testing. DNA-based methods are commonly considered as final stage giving definitive diagnosis. However, in many cases standard techniques used for genetic testing fail.

In the first article we report 25 different mutations in 37 probands originating from 29 families. Fourteen of identified genomic variants were novel and three of them, classified as gross chromosomal rearrangements, would be overlooked using standard sequencing methods. A novel 3 bp in-frame deletion p.Leu43del, eliminating leucine 43 residue from a highly conserved domain important for correct folding of the enzyme was identified in one patient. None of the three novel missense mutations p.Ala102Pro, p.Pro158Ser and p.Lys210Glu was indicated to affect splicing or to be directly involved either in the enzyme active site or in the intersubunit interactions (Shi D, et al., 2001). Nevertheless, the mutated residues show critical intramolecular interactions with other spatially neighboring and highly conserved residues. Sequencing of cDNA obtained from the liver specimens of four patients did not reveal any abnormally spliced transcripts caused by a deletion c.126_128delTCT or substitutions c.304G>C and c.829C>T, but uncovered an aberrant splicing in the case of silent transition c.867G>A resulting in the usage of an alternative exonic splicing site followed by an in-frame deletion of 12 nucleotides.

Large deletions were identified in two males and one female. In patient 1, a large deletion of 444 kb encompassed 15 out of 19 exons of *RPGR* and the entire *OTC* and *TSPAN7* genes. An insertion of a 63 bp fragment identical with the region g.41,647–41,704 coinciding with the *AluJb* retroelement was identified between the breakpoints. The complex mutation was likely originated through a two-step insertion/deletion process (Chuzhanova NA, et al., 2003). The insertion may have occured through misalignment between two inexact inverted repeats that occur upstream of the 5'-breakpoint and the intermediate breakpoint and consist of *AluJb* repeats in opposite orientation with subsequent strand switching via interchromosomal serial replication slippage (Chen JM, et al., 2005). A recombinogenic motif CCCCACCCC and its complement found in close proximity to the intermediate breakpoint (position g.41,555) and the 3' breakpoint (position g.445,313), respectively, may have been responsible for DSBs and non-homologous end joining (NHEJ) resulting in deletion.

The other two large deletions affected only parts of *OTC*: the 10 kb-deletion encompassing exons 5 and 6 was identified in patient 8, whereas patient 36 was heterozygous for the 24.5 kb-

deletion removing exons 9 and 10. Unlike other large deletions described in the *OTC* gene (Engel K, et al., 2008; Quental R, et al.,2009), no repetitive elements were found in the vicinity of the deletion breakpoints in our case. However, recombinogenic motifs CCTCCCT and CCTCCTT, as same as their complements, were found in close proximity to the mutation breakpoints in both patients respectively. The deletions have most likely been mediated by DSBs and subsequent NHEJ (Lieber MR, et al., 2003). Small insertions, as seen in our patients, are common for this type of recombination. Such a mechanism has not been previously described in association with *OTC* gene deletions.

In addition to clearly pathogenic loss-of function mutations we found in our cohort a set of hypomorphic mutations. In accordance with the published data (Matsuda I, et al., 1996), we observed that such mutations may cause a broad spectrum of phenotypes even among the male members of the same family. We examined the X-inactivation pattern in the peripheral blood cells, urinary cells and saliva cells of a manifesting carrier with a hypomorphic mutation. The patient skewed significant X-inactivation in favor of the mutant allele, while her asymptomatic heterozygous mother showed the opposite pattern favoring the wild type allele. It has been shown that trends in X-inactivation patterns are comparable among different tissues of the same individual and analysis of X-inactivation accessible tissues (e.g. blood) may be useful to predict the status in inaccessible tissues (Bittel, et al., 2008). In line with these findings our data indicate that manifestation of symptoms caused by hypomorphic mutation might depend on highly skewed X-inactivation in the liver, which is reflected by X-inactivation status in peripheral tissues.

The second article follows a case excluded from the cohort mentioned above and published separately. No predictably pathogenic mutation was found in the coding region of *OTC* of female patient with mild form of OTCD. The only detected sequence variation was a heterozygous single nucleotide substitution NG_008471.1:g.4849A>G (c.-366A>G). The variation was inherited from the patient's father who had no obvious complaints except for eczema but clinical testing revealed mild orotic aciduria and mild elevation of plasma glutamine and glutamic acid and abnormal allopurinol test. The substituted nucleotide c.-366A is conserved in mammals including rat and mouse and the variation has not been published in SNP databases. Therefore we considered the variation a suspect mutation.

Because the 5'-flanking region of *OTC* gene was not studied yet, we had to determine the TSS and decide whereas the variation is located in the 5'-UTR or in the upstream, potentially promoter region. We identified three alternative TSSs located 95, 119 and 169 bp upstream of

TIS. Although the human *OTC* TSS was first assigned to the position c.-135 (Horwich AL, et al., 1984), later published data suggested that transcription of human *OTC* does initiate within a region of about 70 bases, mostly at preferred positions c.-95, c.-120, c.-150, c.-161 and c.-166 (Brusilow SW and Horwich AL, 2004). These findings as well as the records from the ESTs library containing spliced clones starting from positions c.-92, c.-120, c.-151 and c.-169 bp upstream of TIS correspond well with our results. Only one TSS was found in mouse and rat, located 136 bp (Veres G, et al., 1986) and 87 bp (Takiguchi M, et al., 1987) upstream of the TSS respectively. These results show that the novel variation c.-366A>G is located upstream of the *OTC* 5'UTR.

A 793 bp fragment encompassing the highly conserved region (72% homology with both, rat and mouse) and ranging from -838 to -46 bp upstream of the TIS was chosen for functional testing of promoter properties (Figure 6). The fragment exhibited significantly increased transcriptional activity in hepatocellular lines Hep-G2, HuH-7 and PCL/PRF/5. Interestingly, similar activity was observed also in HEK293 cells (control renal line). Studies in transgenic mice revealed that introduction of the rat *Otc* controlled by the 1.3 kb rat promoter (Murakami T, et al., 1989) and the human *OTC* controlled by the 0.7 kb mouse promoter (Jones SN, et al., 1990) can conduct small intestine-selective gene expression in mouse at levels comparable with those of the endogenous *Otc*. However, the mRNA levels of the introduced genes were low in the liver. These findings together indicate the presence of some other liver-specific *cis*-acting element(s).

Such an element, a distal enhancer ranging 230 bp and situated 11 kb upstream of the TIS, was indeed identified in rat (Murakami T, et al., 1990). We identified a sequence homologous in 83% with rat enhancer in a conserved region 9 kb upstream of human *OTC* (Figure 6) and tested a fragment of 465 bp for its enhancer activity. Subcloning of the enhancer into the reporter construct led to a significant increase in transcriptional activity in Hep-G2 but not in HEK293, thus indicating the role of promoter-enhancer interaction in liver-specific expression of *OTC*.

When comparing the transcriptional activity of normal promoter and promoter carrying the variation c.-366A>G, no significant difference was observed in constructs with promoter alone, but the situation dramatically changed when enhancer was subcloned. Variant promoter in that case reached only about 50% of the transcriptional activity observed with normal promoter. The results of functional studies thus strongly suggest molecular pathogenicity of the c.-366A>G promoter variation.
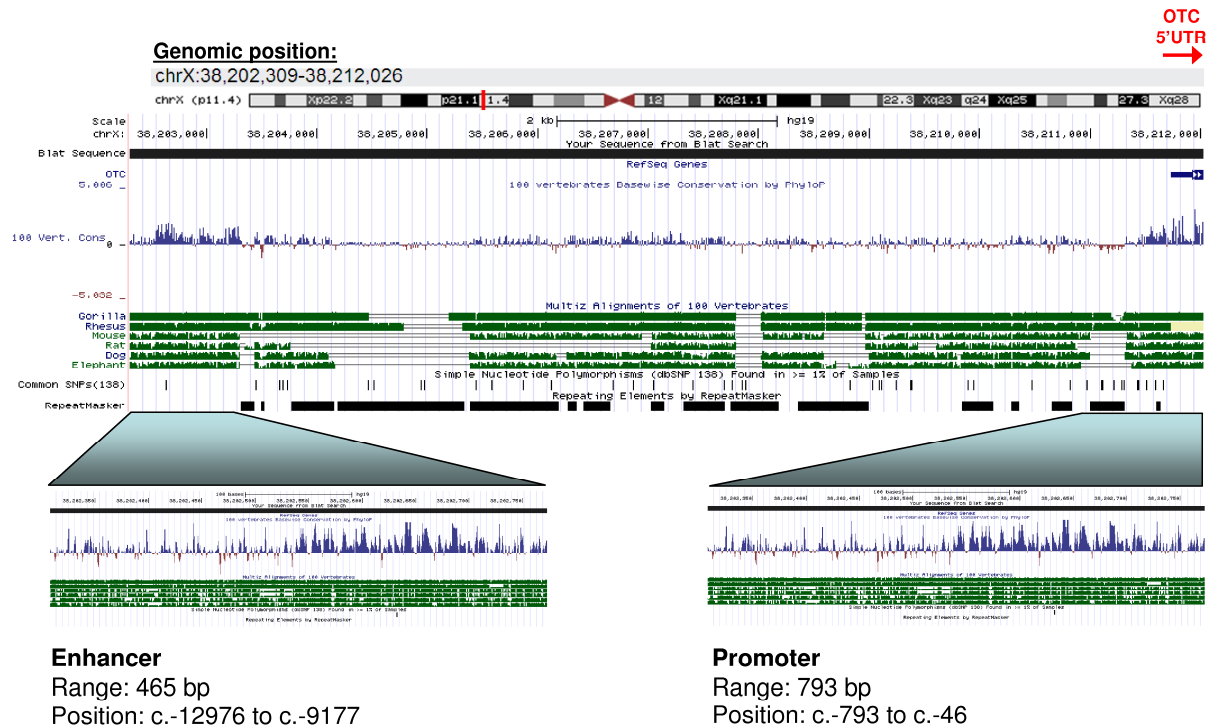
Figure 6: *Prediction of human OTC promoter and enhancer*. *Figure assembled from graphics obtained with the UCSC Genome browser (http://genome.ucsc.edu) represents evolutionary conservation of the OTC 5'flanking region ranging 9717 bp. The blue peaks show evolutionary conservation between 100 vertebrate controls, green peaks indicate levels of homology with selected mammals (gorilla, rhesus, mouse, rat, dog and elephant respectively). Black bars in the middle part correspond to the sequences identified as repeats with the RepeatMasker algorithm, red arrow marks OTC 5'UTR.*

Our goal in the third publication was to characterize *OTC* promoter and enhancer and identify elements responsible for the control of *OTC* transcription. In a set of reporter assays, using a series of 5'-deleted promoter fragments, we defined the 223 bp core promoter located -46..-268 bp upstream of the TIS and containing all three TSS. Within the minimal promoter we identified two positive *cis*-acting elements: Cis1+ located within the first 81 base-pairs, reaching 83% of the full promoter activity and Cis2+ located within the upstream 52 bp sequence. Using a combination of bioinformatic predictive methods, DNase I footprinting and EMSA we identified both, Cis1+ and Cis2+ as HNF-4α binding motifs situated at positions c.-105..c.-121 and c.-136..c.-152 respectively. Another HNF-4α binding site was recognized 187 bp upstream of the TIS. The sequence is highly conserved in vertebrates and the affinity of HNF-4 to this region has been clearly demonstrated in rat (Kimura A, et al., 1993) and mouse

(Inoue Y, et al., 2002). All three HNF-4α binding sites share conserved matrix sequence and could be considered as at least middle affinity binders to HNF-4α (Fang B, et al. 2012). Activation of the enhancer in rat is associated with the interaction of promoter-bound HNF-4 with the enhancer-bound C/EBPβ (Nishiyori A, et al., 1994). A similar mechanism involving HNF-4 and C/EBPα has been suggested for the promoter of human apolipoprotein B gene (Metzger S, et al., 1993). We identified a binding site for C/EBPβ in the *OTC* core promoter between positions c.-221 and c.-234 but deletion of the region did not change the level of reporter signal in luciferase assays.

On the other hand, *in silico* analysis of the proximal 212 bp part of the enhancer region revealed a set of six repeated C/EBPβ binding motifs. Accordingly, the DNase I cleavage of the proximal enhancer region was decreased and DNA-protein interactions were observed even with EMSA. Interaction of the liver enriched HNF-4 factors bound to the *OTC* core promoter with the set of periodically occurring liver specific C/EBP-like proteins thus may contribute to the tissue specificity of *OTC* expression in the liver, as was seen in the rat model (Nishiyori A, et al., 1994).

The distribution of HNF-4α recognition sites within the *OTC* core promoter region exhibits an important feature: All three HNF-4α binding motifs are located within 35 bases upstream of the TSSs (Figure 7). Moreover, the sequence of 223 bp proximal promoter lacks any general core promoter elements such as Inr, DPE, BRE, MTE, downstream core element (DCE) and the X core promoter element (XCPE) located at standard positions upstream of the TSSs (Figure 7), thereby suggesting the importance of HNF-4α in the initiation of *OTC* transcription.

Few indices suggesting the role of HNF-4α in the transcription from TATA- less and Inr-less promoters can be found in the literature. The HNF-4α was shown to activate transcription *in vitro* by facilitating assembly of the pre-initiation complex through direct physical interaction with TFIIB (Malik S and Karathanasis SK, 1996). Similar mechanism was observed later in estrogen receptor, an another member of the nuclear hormone receptor superfamily (Sabbah M, et al., 1998). In a recent study, a systematic proteomic characterization of the HNF4α-TFIID interactions revealed that HNF4α can directly target TBP subunit of TFIID to promoters containing HNF4α-binding sites (Takahashi H, et al., 2009). The findings suggest that regulatory mechanism of *OTC* transcription may involve physical interaction of HNF4α with general transcription factors forming the pre-initiation complex of Pol II. Such a mechanism would explain characteristic features observed within *OTC* minimal promoter: existence of multiple TSS, site-oriented initiation of transcription in the absence of general core

promoter elements and a partial ability to compensate elimination of one *cis*-active element by the function of others.
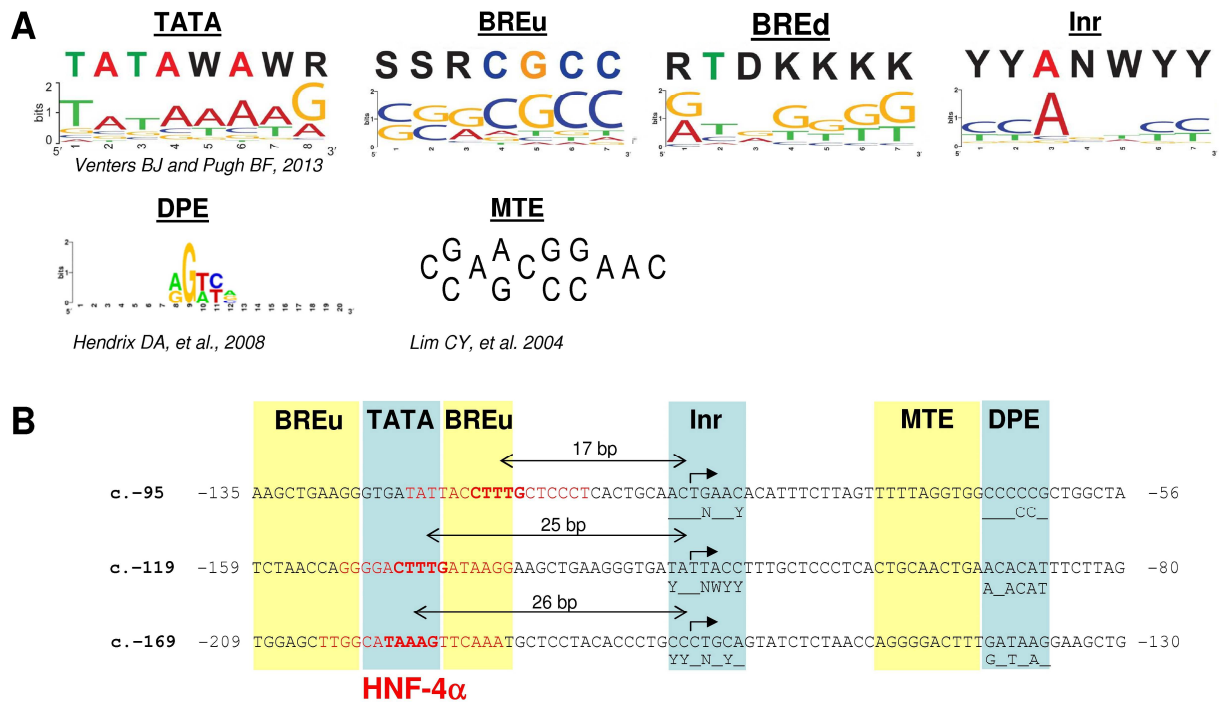
A

TATA — TATAWAWR
Venters BJ and Pugh BF, 2013

BREu — SSRCGCC

BREd — RTDKKKK

Inr — YYANWYY

DPE
Hendrix DA, et al., 2008

MTE — CGAACGGAAC
Lim CY, et al. 2004

B

```
            BREu  TATA  BREu                    Inr              MTE   DPE
                               17 bp
c.-95   -135 AAGCTGAAGGGTGATATTACCTTTGCTCCCTCACTGCAACTGAACACATTTCTTAGTTTTTAGGTGGCCCCCGCTGGCTA -56
                                                  N__Y                              __CC_
                               25 bp
c.-119  -159 TCTAACCAGGGGACTTTGATAAGGAAGCTGAAGGGTGATATTACCTTTGCTCCCTCACTGCAACTGAACACATTTCTTAG -80
                                                 Y__NWYY                           A_ACAT
                               26 bp
c.-169  -209 TGGAGCTTGGCATAAAGTTCAAATGCTCCTACACCCTGCCCTGCAGTATCTCTAACCAGGGGACTTTGATAAGGAAGCTG -130
                                                 YY_N_Y_                           G_T_A_
                 HNF-4α
```

Figure 7: *Characteristic features of human OTC core promoters. Figure represents the distribution of general core promoter elements within the region surrounding three alternative TSSs of human OTC. None of these elements was recognized in standard position according to the TSSs, except for the TATA-like sequence located 20 bp upstream of the c.-169 transcription origin. A – Matrix sequences of common general core promoter elements with marked frequencies of single bases. B – sequences of +/- 40 nucleotides surrounding alternative TSS. Standard positions of general core promoter elements are marked by yellow and blue boxes. The HNF-4α binding sites are in red with the core motif boldfaced. Bent arrow marks the TSS, horizontal bars show the distances between TSS and HNF-4α binding sites.*

The methodological approach introduced within the study of *OTC* regulatory regions was applied on the glucocerebrosidase gene. The *GBA* promoter (P1) was assigned to the region located immediately upstream of the coding-exon 1 (Horowitz M, et al., 1989). Surprisingly, four of the five known glucocerebrosidase transcripts contain alternative non-coding exon −2 or both exons −2 and −1. We identified three alternative TSS located at positions c.-347, c.-380 and c.-413 uptream of the TIS and inspected transcriptional activity of a 1156 bp fragment (P2) ranging from c.-353 to c.-1509 upstream of the non-coding exon -2. The P2 promoter was

found to be active, but the reached levels of reporter gene signal were about six times lower than those of P1. Since we used both fragments of similar size, we might miss some important *cis*-active element and therefore the reporter gene expression from P1 and P2 constructs may not be proportional to the strength of the native promoters.

It was shown that lysosomal genes exhibit coordinated transcriptional behavior and are regulated by the transcription factor EB (TFEB) (Sardiello M, et al., 2009). The coordinated lysosomal expression and regulation motif (CLEAR) recognized by TFEB is widely enriched in promoters of lysosomal genes. While two such motifs were previously recognized and confirmed in the P1, we identified another two CLEAR-like sequences in P2 at positions 764 bp and 1238 bp upstream of the TIS. We found that in contrary to the P1, the P2 lacks a TATA-box but it carries a multiple Sp-1 binding sites and an unmethylated CpG island thus presenting common properties with the majority of housekeeping promoters. However, the results from the quantitative RT-PCR in RNA isolated from normal adult human tissues did not show significant differences in the expression from both promoters. Moreover, the expression stability of both, the P1 and P2, appears to be higher than that of the commonly used reference housekeeping genes GAPDH and ACTB.

In line with these findings we hypothesized that differences in the manifestation of GD in patients presenting the same genotype may be caused by variable expression of different *GBA* transcripts due to variations in the 5'-flanking region. We therefore examined the promoter sequences in a group of 20 Ashkenazi Jewish Gaucher patients homozygous for the common mild mutation p.N370S to inspect if the variable phenotype is not caused by aberrant transcription. Unfortunately, we did not find any sequence variation either in the P1 or in the P2 except for a common polymorphism rs10908459.

6. Conclusions

In the present work we described the molecular basis of OTCD in probands coming from 29 families. We identified and characterized 14 novel mutations affecting the protein-coding region including three large deletions and a pathogenic variation located in the 5'-flanking region. We found and functionally characterized human *OTC* promoter and distal enhancer and suggested mechanism of their involvement in the control of transcription. We found and functionally characterized an upstream *GBA* promoter driving transcription of non-coding exons -2 and -1.

The results, presented in four publications, can be summed up as follows:

1) Fourteen out of twenty five mutations altering the coding sequence of *OTC* gene were found novel in DNA samples from 29 OTCD families. Three of them are missense mutations affecting residues of highly conserved OTC architecture. another three are large deletions. A manifesting heterozygote carrying a hypomorphic mutation with unfavorable X-inactivation was identified and unfavorable mosaic skewing was observed in three peripheral tissues.

2) The sequence spanning 793 bp in the 5'-flanking region of human *OTC* acts as a promoter. The transcriptional activity of the promoter depends on a newly discovered upstream enhancer located 9 kb upstream of the translation initiation site. The promoter – enhancer interaction contributes to tissue specific expression of *OTC* in liver-derived cell lines. Single nucleotide substitutions in the regulatory regions such as the c.-366A>G variation may cause OTC deficiency. Therefore, analysis of the regulatory regions of *OTC* should be considered in patients with OTCD and negative finding in the coding region of the gene.

3) HNF-4α plays an essential role in regulation of human *OTC* expression. Three HNF-4α binding motifs are located within 35 bases upstream each of the transcription start sites in the minimal promoter. Two of these sites were identified as *cis*-active regulatory elements responsible for the full transcriptional activity of the promoter and both were shown to cooperate with distal enhancer in the synergistic manner resulting in significantly increased liver-specific expression of *OTC*. Moreover, DNA-bound HNF-4α seems to act as a direct initiator of *OTC* transcription through recruitment of general transcription factors TFIIB

and TDFIID and stabilization of the pre-initiation complex to the promoter at particular sites lacking TATA-box or initiator.

4) The region 1 kb upstream of the *GBA* non-coding exon −2 acts as an alternative promoter. The transcriptional activity of the alternative promoter P2 seems to be lower than that of the promoter P1 driving the *GBA* expression from exon 1. The promoter P2 lacks TATA-box but contains two CLEAR-like, multiple Sp-1 binding sites and an unmethylated CpG island. Expression of the transcripts originating from both promoters was stable among multiple tissues. The DNA analysis of the P1 and P2 promoters in a group of twenty Ashkenazi Jewish Gaucher patients with pleiotropic phenotype homozygous for the common mild mutation p.N370S did not reveal any sequence variations responsible for the phenotypic differences.

Our data extend the possibilities of genetic testing for *OTC* and *GBA* deficiency.

7. Summary

Molecular analysis is widely used to establish the final diagnosis of inherited metabolic disorders. Nevertheless, standard methods based on the analysis of sequence variations in the protein-coding regions often reach their limits. Particularly complex mutations at the chromosomal level are commonly uncovered by DNA sequencing techniques and require application of more sophisticated methods. In other cases, the disease may be caused by aberrant regulation of gene expression due to mutations in regulatory regions. Special interest should be therefore paid to detailed characterization of promoters, enhancers and other regulatory units, similarly as to the understanding of regulatory mechanisms.

In the present thesis we deal with characterization of complex mutations in a set of OTCD patients followed by a detailed study of *OTC* regulatory regions. Then we move to the *GBA* gene, where we apply similar methodological approach to characterization of an upstream promoter controlling the expression of several alternative transcripts.

First, we have described molecular basis of OTCD in a cohort of patients originating from 29 families. We identified and characterized 25 different mutations, 14 of which were novel. In one female and two male patients we found 3 large deletions: A large 444 kb deletion g.1276_445398delins63 removing the entire *OTC* together with other genes, and two smaller deletions g.42940_52876delinsT and g.57467_82002delinsCCT removing only parts of *OTC*. In all novel mutations we evaluated their clinical significance. Moreover, we described a heterozygote carrying a hypomorphic mutation and manifesting OTCD likely due to the unfavorable X-inactivation which was observed independently in three different peripheral tissues.

In one OTCD familly, which was excluded from the cohort and published separately, we identified likely pathogenic variation c.-366A>G located in the 5'-flanking region. To establish its clinical significance we delimited the *OTC* promoter. We determined 3 alternative transcription start sites located at positions c.-95, c.-119 and c.-169 upstream of the translation origin and defined and functionally characterized the *OTC* promoter as a sequence spanning 793 bp located 46 bp upstream of the initiation codon. Moreover, we found a strong distal enhancer contributing to the tissue-specific expression of *OTC* in the liver. Finally we evaluated the transcriptional activity of mutant promoter containing the c.-366A>G variation in a reporter study and observed about 50% decrease when compared with normal promoter, thus indicating pathogenicity of the variation. To our best knowledge we reported the first disease-causing mutation in a regulatory region of one of urea cycle genes.

In order to understand the mechanism of *OTC* transcription control, necessary for correct interpretation of findings obtained with the mutation analysis of these regions, we inspected the promoter and enhancer regions for functional sites. We delimited the minimal promoter and identified two HNF-4α binding sites acting as strong *cis*-active regulatory elements responsible for the full transcriptional activity of activation of the enhancer. Another, HNF-4α recognition motif was found in the distal part of minimal promoter. All three HNF-4α are located within 35 bp upstream of the transcription start sites in a promoter lacking standard general core promoter elements on standard positions, strongly suggesting direct involvement of these sites in the control of *OTC* transcription. In the proximal part of enhancer we localized a series of C/EBPβ recognition motifs which likely co-operate with promoter bound HNF-4α to increase the rate of transcription. Synergistic operation of these liver-enriched factors seems to contribute to the tissue-specific expression of *OTC* in the liver.

We used a subset of methods introduced within the *OTC* project to characterize an upstream promoter of the β-glucocerebrosidase gene related with another inherited metabolic disorder, GD. We identified three alternative *GBA*. transcription start sites located 347, 380, and 413 bp upstream of the translation initiation site and the alternative *GBA* promoter (P2) controlling the transcription of mRNAs containing untranslated exons -2 and -1. The transcriptional activity of the fragment of the P2 promoter seems to be lower than in the case of the P1 promoter driving the expression from exon 1. The P2 promoter lacks TATA-box but contains two CLEAR-like, multiple Sp-1 binding sites and an unmethylated CpG island. The expression profiles of both, transcripts originating from P2 and P1, were rather stable across 20 human tissues indicating characteristic pattern of housekeeping genes. We hypothesized that differences in the manifestation of Gaucher disease in patients with the same genotype may be caused by variable expression of mutant glucocerebrosidase due to variants in the promoters. However, the hypothesis was not confirmed: except from common polymorphism (rs10908459), no sequence variation was detected in the P1 nor the P2 promoter in a phenotypically diverse group of 20 patients homozygous for the common mild mutation p.N370S.

In conclusion, our findings of novel mutations namely the large deletions in OTC, the discovery of key regulatory sequences essential for liver specific expression of OTC and the discovery of an alternative *GBA* promoter extend the possibilities of molecular genetic testing for OTCD and GD.

8. References

1. Aula P, Renlund M, Raivio KO, Koskela SL. Screening of inherited oligosacchariurias among mentally retarded patients in northern Finland. J Ment Defic Res. 1986;30:365-8.

2. Azevedo L, Stolnaja L, Tietzeova E, Hrebicek M, Hruba E, Vilarinho L, Amorim A, Dvorakova L: New polymorphic sites within ornithine transcarbamylase gene: population genetics studies and implications for diagnosis. Mol Genet Metab. 2003; 78(2):152-7.

3. Bajic VB, Tan SL, Christoffels A, Schönbach C, Lipovich L, Yang L, Hofmann O, Kruger A, Hide W, Kai C, Kawai J, Hume DA, Carninci P, Hayashizaki YBajic VB, Tan SL, Christoffels A, Schonbach C,Lipovich L, Yang L. Mice and men: their promoter properties. PLoS Genet. 2006; 2(4):e54.

4. Bean ES and Atkinson DE: Regulation of the rate of urea synthesis in liver by extracellular pH. A major factor in pH homeostasis in mammals. J Biol Chem. 1984; 259:1552-9

5. Beck M. Variable clinical presentation in lysosomal storage disorders. J Inherit Metab Dis. 2001; 24:47-51

6. Beliveau CG, Cheung CW, Cohen NS, Brusilow S, Raijman L: Regulation of urea and citrulline synthesis under physiological conditions. Biochem J. 1993, 292:241–7

7. Beutler E and Grabowski GA. Gaucher disease. In: Scriver CR, Beaudet AL, Sly WS, Valle D. The Metabolic and Molecular Basis of Inherited Disease. 8th ed. New York, McGraw-Hill, Inc, 2001; p.3635–68.

8. Beutler E, Nguyen NJ, Henneberger MW, Smolec JM, McPherson RA, West C, Gelbart T. Gaucher disease: gene frequencies in the Ashkenazi Jewish population. Am J Hum Genet. 1993; 52(1):85-8

9. Beutler E. Gaucher disease: multiple lessons from a single gene disorder. Acta Paediatr Suppl. 2006; 95:103e9.

10. Beutler E. Gaucher disease: new molecular approaches to diagnosis and treatment. Science. 1992; 256:794–9.

11. Biegstraaten M, van Schaik IN, Aerts JM, Hollak CE: Review 'Non-neuronopathic' Gaucher disease reconsidered. Prevalence of neurological manifestations in a Dutch cohort of type I Gaucher disease patients and a systematic review of the literature. J Inherit Metab Dis. 2008; 31(3):337-49.

12. Bittel DC, Theodoro MF, Kibiryeva N, Fischer W, Talebizadeh Z, Butler MG. Comparison of X-chromosome inactivation patterns in multiple tissues from human females. J Med Genet. 2008; 45:309–13.

13. Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science. 1998; 281(5373):60-3.

14. Blech-Hermoni YN, Ziegler SG, Hruska KS, Stubblefield BK, Lamarca ME, Portnoy ME; NISC Comparative Sequencing Program, Green ED, Sidransky E. In silico and functional studies of the regulation of the glucocerebrosidase gene, Mol Genet Metab. 2010; 99:275–82.

15. Bourrier P, Varache N, Alquier P, Rabier D, Kamoun P, Lorre G, Alhayek G. Cerebral edema with hyperammonemia in valpromide poisoning. Manifestation in an adult, of a partial deficit in type I carbamylphosphate synthetase. Presse Med. 1988; 17:2063-6

16. Bradford N and McGivan J. Evidence for the existence of an ornithine/citrulline antiporter in rat liver mitochondria. FEBS Lett. 1980; 113:294–6

17. Brusilow SW and Horwich AL. Urea Cycle Enzymes. In: Scriver CR, Beaudet AL, Sly WS, Valle D. Metabolic and Molecular Bases of Inherited Diseases - OMMBID. www.ommbid.com. New York, McGraw Hill, Inc, 2004; Chap. 85.

18. Brusilow SW and Maestri NE. Urea cycle disorders: diagnosis, pathophysiology, and therapy. Adv Pediatr. 1996; 43:127–170.

19. Bulger M and Groudine M. Looping versus linking: toward a model for long-distance gene activation. Genes Dev. 1999; 13(19):2465-77.

20. Burke TW and Kadonaga JT. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev. 1997; 11:3020-31.

21. Burke TW, Willy PJ, Kutach AK, Butler JE, Kadonaga JT. The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harb Symp Quant Biol. 1998; 63:75-82.

22. Camacho JA, Obie C, Biery B, Goodman BK, Hu CA, Almashanu S, Steel G, Casey R, Lambert M, Mitchell GA, Valle D. Hyperornithinaemia-hyperammonaemia-homocitrullinuria syndrome is caused by mutations in a gene encoding a mitochondrial ornithine transporter. Nature Genet. 1999; 22:151-8

23. Carcamo J, Maldonado E, Cortes P, Ahn MH, Ha I, Kasai Y, Flint J, Reinberg D. A TATA-like sequence located downstream of the transcription initiation site is required for expression of an RNA polymerase II transcribed gene. Genes Dev. 1990; 4:1611-22.

24. Chalkley GE and Verrijzer CP. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250– TAF(II)150 complex recognizes the initiator. EMBO J. 1999; 18:4835-45.

25. Charrow J, Andersson HC, Kaplan P, Kolodny EH, Mistry P, Pastores G, Rosenbloom BE, Scott CR, Wappner RS, Weinreb NJ, Zimran A. The Gaucher registry: demographics and disease characteristics of 1698 patients with Gaucher disease. Arch Intern Med. 2000; 160(18):2835-43

26. Chen JM, Chuzhanova N, Stenson PD, Krawczak M, Ferec C, Cooper DN. Intrachromosomal serial replication slippage in trans gives rise to diverse genomic rearrangements involving inversions. Hum Mutat. 2005; 26:362–73.

27. Chuzhanova NA, Anassis EJ, Ball EV, Krawczak M, Cooper DN. Meta-analysis of indels causing human genetic disease: mechanisms of mutagenesis and the role of local DNA sequence complexity. Hum Mutat. 2003; 21:28–44.

28. Cohen NS, Cheung CW, Raijman L. Channeling of extramitochondrial ornithine to matrix ornithine transcarbamylase. J Biol Chem. 1987; 262:203-8

29. Deng W and Roberts SG. A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev. 2005; 19:2418-23.

30. Dignam JD, Lebovitz RM, Roeder RG. Accurate Transcription Initiation by Rna Polymerase-Ii in a Soluble Extract from Isolated Mammalian Nuclei. Nucleic Acids Res. 1983; 11(5):1475-89.

31. Dionisi-Vici C, Rizzo C, Burlina AB, Caruso U, Sabetta G, Uziel G, Abeni D. Inborn errors of metabolism in the Italian pediatric population: a national retrospective survey. J Pediatr. 2002; 140:321–7.

32. Dobrovolny R, Dvorakova L, Ledvinova J, Magage S, Bultas J, Lubanda JC, Elleder M, Karetova D, Pavlikova M, Hrebicek M. Relationship between X-inactivation and clinical involvement in Fabry heterozygotes. Eleven novel mutations in the alpha-galactosidase A gene in the Czech and Slovak population. J Mol Med. 2005; 83:647–54.

33. Doll RF and Smith FI. Regulation of expression of the gene encoding human acid beta-glucosidase in different cell types. Gene. 1993; 127:255–60.

34. Doll RF, Bruce A, Smith FI. Regulation of the human acid beta-glucosidase promoter in multiple cell types, Biochim. Biophys Acta. 1995; 1261:57–67.

35. Edelmann L, Dong J, Desnick RJ, Kornreich R. Carrier screening for mucolipidosis type IV in the American Ashkenazi Jewish population. Am J Hum Genet. 2002; 70:1023-27.

36. Engel K, Nuoffer JM, Mühlhausen C, Klaus V, Largiadèr CR, Tsiakas K, Santer R, Wermuth B, Häberle J. Analysis of mRNA transcripts improves the success rate of molecular genetic testing in OTC deficiency. Mol Genet Metab. 2008; 94:292–7.

37. Eyal N, Wilder S, Horowitz M. Prevalent and rare mutations among Gaucher patients. Gene. 1990; 96:277–83.

38. Fang B, Mane-Padros D, Bolotin E, Jiang T, Sladek FM. Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. Nucleic Acids Res. 2012; 40:5343-56.

39. Filocamo M and Morrone A. Lysosomal storage disorders: Molecular basis and laboratory testing. Hum Genomics. 2011; 5(3):156–69.

40. FitzGerald PC, Shlyakhtenko A, Mir AA, Vinson C. Clustering of DNA sequences in human promoters. Genome Res. 2004; 14:1562-74.

41. Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. A code for transcription initiation in mammalian genomes. Genome Res. 2008; 18:1-12.

42. Goker-Alpan O, Hruska KS, Orvisky E, Kishnani PS, Stubblefield BK, Schiffmann R, Sidransky E. Divergent phenotypes in Gaucher disease implicate the role of modifiers. J Med Genet. 2005; 42(6):e37

43. Grabowski GA: Gaucher disease: gene frequencies and genotype/phenotype correlations. Genet Test. 1997; 1:5–12.

44. Gropper SAS, Smith JL, Groff JL. 5th ed. Advanced nutrition and human metabolism. Australia, Wadsworth/Cengage Learning, 2009; p.229

45. Hamano Y, Kodama H, Yanagisawa M, Haraguchi Y, Mori M, Yokota S. Immunocytochemical localization of ornithine transcarbamylase in rat intestinal mucosa. Light and electron microscopic study. J Histochem Cytochem. 1988; 36(1):29-35.

46. Hampsey M. Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev. 1998; 62(2):465-503

47. Hata A, Tsuzuki T, Shimada K, Takiguchi M, Mori M, Matsuda I. Structure of the human ornithine transcarbamylase gene. J Biochem. 1988; 103(2):302-8.

48. Heibel SK, Ah Mew N, Caldovic L, Daikhin Y, Yudkoff M, Tuchman M. N-carbamylglutamate enhancement of ureagenesis leads to discovery of a novel deleterious mutation in a newly defined enhancer of the NAGS gene and to effective therapy. Hum Mutat. 2011; 32:1153–60.

49. Hendrix DA, Hong JW, Zeitlinger J, Rokhsar DS, Levine MS. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc Natl Acad Sci. 2008; 105(22):7762-7.

50. Hopkin R and Grabowski G. Lysosomal storage diseases. In: Kasper DL Braunwald E Hauser S Longo D Jameson LD Fauci AS. Harrison's principles of internal medicine. 16th ed. New York, McGraw-Hill, Inc, 2004; p. 2315-9

51. Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E. The human glucocerebrosidase gene and pseudogene: structure and evolution. Genomics. 1989; 4:87–96.

52. Horwich AL, Fenton WA, Williams KR, Kalousek F, Kraus JP, Doolittle RF, Konigsberg W, Rosenberg LE. Structure and expression of a complementary DNA for the nuclear coded precursor of human mitochondrial ornithine transcarbamylase. Science. 1984; 224(4653):1068-74.

53. Hruska KS, LaMarca ME, Scott CR, Sidransky E. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). Hum Mutat. 2008; 29(5):567-83.

54. Inoue Y, Hayhurst GP, Inoue J, Mori M, Gonzalez FJ. Defective ureagenesis in mice carrying a liver-specific disruption of hepatocyte nuclear factor 4alpha (HNF4alpha). J Biol Chem. 2002; 277(28):25257-65.

55. Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol. 2010; 11(2):113-27.

56. Jones SN, Grompe M, Munir MI, Veres G, Craigen WJ, Caskey CT. Ectopic Correction of Ornithine Transcarbamylase Deficiency in Sparse Fur Mice. J Biol Chem. 1990; 265(24):14684-90.

57. Kalousek F, Francois B, Rosenberg LE. Isolation and characterization of ornithine transcarbamylase from normal human liver. J Biol Chem. 1978; 253:3939-44.

58. Kaufmann J and Smale ST. Direct recognition of initiator elements by a component of the transcription factor IID complex. Genes Dev. 1994; 8:821-9.

59. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA,Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. A high-resolution map of active promoters in the human genome. Nature. 2005; 436:876-80.

60. Kimura A, Nishiyori A, Murakami T, Tsukamoto T, Hata S, Osumi T, Okamura R, Mori M, Takiguchi M. Chicken Ovalbumin Upstream Promoter-Transcription Factor (Coup-Tf) Represses Transcription from the Promoter of the Gene for Ornithine Transcarbamylase in a Manner Antagonistic to Hepatocyte Nuclear Factor-Iv (Hnf-4). J Biol Chem. 1993; 268(15):11125-33.

61. Koprivica V, Stone DL, Park JK, Callahan M, Frisch A, Cohen IJ, Tayebi N, Sidransky E. Analysis and Classification of 304 Mutant Alleles in Patients with Type 1 and Type 3 Gaucher Disease. Am J Hum Genet. 2000; 66:1777–86

62. Krebs HA and Henseleit K. Untersuchungen uber die harnstoffbildung im tierkorper. Hoppe-Seyler's Z Physiol Chem. 1932; 210:325–32.

63. Krebs HA, Hems R, Lund P. Some regulatory mechanisms in the synthesis of urea in the mammalian liver. Adv. Enzyme Regul. 1973; 11:361-77

64. Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev. 1998; 12:34-44.

65. Lanpher BC, Gropman A, Chapman KA, et al. Urea Cycle Disorders Overview. 2003 Apr 29 [Updated 2011 Sep 1] In: Pagon RA, Adam MP, Bird TD, et al. editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993-2014. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1217/

66. Latham TE, Theophilus BD, Grabowski GA, Smith FI. Heterogeneity of mutations in the acid beta-glucosidase gene of Gaucher disease patients. DNA Cell Biol. 1991; 10:15–21.

67. Lee DH, Gershenzon N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol Cell Biol. 2005; 25:9674-86.

68. Leonard JV and Morris AA. Urea cycle disorders. Semin Neonatol. 2002; 7:27–35.

69. Lieber MR, Ma Y, Pannicke U, Schwarz K. Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol. 2003; 4:712–20.

70. Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev. 2004; 18:1606-17.

71. Lindgren V, de Martinville B, Horwich AL, Rosenberg LE, Francke U. Human ornithine transcarbamylase locus mapped to band Xp21.1 near the Duchenne muscular dystrophy locus. Science. 1984; 226(4675):698-700.

72. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Molecular Cell Biology. 5th ed. New York, W. H. Freeman, 2000; p.448.

73. Lopes-Marques M, Pereira-Castro I, Amorim A, Azevedo L. Characterization of the Human Ornithine Transcarbamylase 3′ Untranslated Regulatory Region. DNA Cell Biol. 2012; 31(4):427–33.

74. Malik S, Karathanasis SK. TFIIB-directed transcriptional activation by the orphan nuclear receptor hepatocyte nuclear factor 4. Mol Cell Biol. 1996; 16:1824-31.

75. Marsden D and Levy H. Newborn Screening of Lysosomal Storage Disorders. Clin Chem. 2010, 56(7):1071-9

76. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet. 2006; 7:29-59.

77. Matsuda I, Matsuura T, Nishiyori A, Komaki S, Hoshide R, Matsumoto T, Funakoshi M, Kiwaki K, Endo F, Hata A, Shimadzu M, Yoshino M.Matsuda I, Matsuura T, Nishiyori A. Phenotypic variability in male patients carrying the mutant ornithine transcarbamylase (OTC) allele, Arg40His, ranging from a child with an unfavourable prognosis to an asymptomatic older adult. J Med Genet. 1996; 33:645–8.

78. Meijer AJ and Verhoeven AJ. N-acetylglutamate and urea synthesis. Biochem. J. 1984; 223:559-60

79. Meikle PJ, Hopwood JJ, Clague AE, Carey WF. Prevalence of lysosomal storage disorders. JAMA. 1999; 281:249-54

80. Metzger S, Halaas JL, Breslow JL, Sladek FM. Orphan Receptor Hnf-4 and Bzip Protein C/Ebp-Alpha Bind to Overlapping Regions of the Apolipoprotein-B Gene Promoter and Synergistically Activate Transcription. J Biol Chem. 1993; 268:16831-8.

81. Miyajima N, Kadowaki Y, Fukushige S, Shimizu S, Semba K, Yamanashi Y, Matsubara K, Toyoshima K, Yamamoto T. Identification of two novel members of erbA superfamily by molecular cloning: the gene products of the two are highly related to each other. Nucleic Acids Res. 1988; 16(23):11057-74.

82. Mononen T, Mononen I, Matilainen R, Airaksinen E. High prevalence of aspartylglycosaminuria among school-age children in eastern Finland. Hum Genet. 1991; 87:266-8.

83. Moran D, Galperin E, Horowitz M. Identification of factors regulating the expression of the human glucocerebrosidase gene. Gene. 1997; 194:201–13.

84. Morris SM. Regulation of enzymes of the urea cycle and arginine metabolism. Annu Rev Nutr. 2002; 22:87-105.

85. Moshonov S, Elfakess R, Golan-Mashiach M, Sinvani H, Dikstein R. Links between core promoter and basic gene features influence gene expression. BMC Genomics. 2008; 9:92.

86. Murakami T, Nishiyori A, Takiguchi M, Mori M. Promoter and 11-kilobase upstream enhancer elements responsible for hepatoma cell-specific expression of the rat ornithine transcarbamylase gene. Mol Cell Biol. 1990; 10(3):1180-91.

87. Murakami T, Takiguchi M, Inomoto T, Yamamura K, Mori M. Tissue- and developmental stage-specific expression of the rat ornithine carbamoyltransferase gene in transgenic mice. Dev Genet. 1989; 10(5):393-401.

88. Natowicz MR and Prence EM. Heterozygote screening for Tay-Sachs disease: past successes and future challenges. Curr Opin Pediatr. 1996; 8:625-9.

89. Nishiyori A, Tashiro H, Kimura A, Akagi K, Yamamura K, Mori M, Takiguchi M. Determination of tissue specificity of the enhancer by combinatorial operation of tissue-enriched transcription factors. Both HNF-4 and C/EBP beta are required for liver-specific activity of the ornithine transcarbamylase enhancer. J Biol Chem. 1994; 269(2):1323-31.

90. O'Neill RR, Tokoro T, Kozak CA, Brady RO. Comparison of the chromosomal localization of murine and human glucocerebrosidase genes and of the deduced amino acid sequences. Proc Natl Acad Sci. 1989; 86:5049-53

91. Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. Genes Dev. 1996; 10(21):2657-83.

92. Poorthuis BJ, Wevers RA, Kleijer WJ, Groener JE, de Jong JG, van Weely S, Niezen-Koning KE, van Diggelen OP. The frequency of lysosomal storage diseases in The Netherlands. Hum Genet. 1999; 105:151-6.

93. Quental R, Azevedo L, Rubio V, Diogo L, Amorim A. Molecular mechanisms underlying large genomic deletions in ornithine transcarbamylase (OTC) gene. Clin Genet. 2009; 75:457–64.

94. Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci. 1996; 21(9):327-35.

95. Ryall J, Nguyen M, Bendayan M, Shore GC. Expression of nuclear genes encoding the urea cycle enzymes, carbamoyl-phosphate synthetase I and ornithine carbamoyl transferase, in rat liver and intestinal mucosa. Eur J Biochem. 1985; 152(2): 287-92.

96. Sabbah M, Kang KI, Tora L, Redeuilh G. Oestrogen receptor facilitates the formation of preinitiation complex assembly: Involvement of the general transcription factor TFIIB. Biochem J. 1998; 336:639-46.

97. Sardiello M, Palmieri M, di Ronza A, Medina DL, Valenza M, Gennarino VA, Di Malta C, Donaudy F, Embrione V, Polishchuk RS, Banfi S, Parenti G, Cattaneo E, Ballabio A. A gene network regulating lysosomal biogenesis and function. Science. 2009; 325(5939):473-7

98. Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ Jr. Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol. 2005; 6:33.

99. Shi D, Morizono H, Yu X, Tong L, Allewell NM, Tuchman M. Human ornithine transcarbamylase: crystallographic insights into substrate recognition and conformational changes. Biochem J. 2001; 354:501-9.

100. Shigesada K and Tatibana M. Role of acetylglutamate in ureotelism. I. Occurrence and biosynthesis of acetylglutamate in mouse and rat tissues. J Biol Chem. 1971; 246:5588-95

101. Sidransky E, Bottler A, Stubblefield B, Ginns EI. DNA mutational analysis of type 1 and type 3 Gaucher patients: how well do mutations predict phenotype? Hum Mutat. 1994; 3:25-8.

102. Smale ST and Baltimore D. The "initiator" as a transcription control element. Cell. 1989; 57:103-13.

103. Snodgrass PJ. Ornithine Transcarbamylase: Basic science and Clinical Considerations. Boston, Kluwer Academic Publishers, 2004; Chap. 2.

104. Sonenberg N and Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell. 2009; 136(4):731-45.

105. Sonoda T and Tatibana M. Purification of N-acetyl-l-glutamate synthetase from rat liver mitochondria and substrate and activator specificity of the enzyme. J Biol Chem. 1983; 258:9839-44.

106. Sorge J, Kuhl W, West C, Beutler E. Complete correction of the enzymatic defect of type I Gaucher disease fibroblasts by retroviral-mediated gene transfer. Proc Natl Acad Sci. 1987; 84:906-9

107. Sorge J, West C, Westwood B, Beutler E. Molecular cloning and nucleotide sequence of human glucocerebrosidase cDNA. Proc Natl Acad Sci. 1985; 82:7289-93

108. Sugie K, Yamamoto A, Murayama K, Oh SJ, Takahashi M, Mora M, Riggs JE, Colomer J, Iturriaga C, Meloni A, Lamperti C, Saitoh S, Byrne E, DiMauro S, Nonaka I, Hirano M, Nishino I. Clinicopathological features of genetically confirmed Danon disease. Neurology. 2002; 58:1773-8

109. Summar M. Current strategies for the management of neonatal urea cycle disorders. J Pediatr. 2001; 138: S30–S39.

110. Takahashi H, Martin-Brown S, Washburn MP, Florens L, Conaway JW, Conaway RC. Proteomics Reveals a Physical and Functional Link between Hepatocyte Nuclear Factor 4 alpha and Transcription Factor IID. J Biol Chem. 2009; 284:32405-12.

111. Takiguchi M, Murakami T, Miura S, Mori M. Structure of the rat ornithine carbamoyltransferase gene, a large, X chromosome-linked gene with an atypical promoter. Proc Natl Acad Sci. 1987; 84(17):6136-40.

112. Theisen JW, Lim CY, Kadonaga JT. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. Mol Cell Biol. 2010; 30:3471-9.

113. Tuchman M, Jaleel N, Morizono H, Sheehy L, Lynch MG. Mutations and polymorphisms in the human ornithine transcarbamylase gene. Hum Mutat. 2002; 19(2):93-107.

114. Venters BJ and Pugh BF. Genomic organization of human transcription initiation complexes. Nature. 2013; 502(7469):53-8.

115. Veres G, Craigen WJ, Caskey CT. The 5'-flanking region of the ornithine transcarbamylase gene contains DNA sequences regulating tissue-specific expression. J Biol Chem. 1986; 261(17):7588-91.

116. Voet D and Voet JG. Biochemistry. 3rd ed. New York, J. Wiley & Sons, 1995.

117. Wenger DA, Coppola S, Liu SL. Insights into the diagnosis and treatment of lysosomal storage diseases. Arch Neurol. 2003; 60:322-8.

118. Westbroek W, Gustafson AM, Sidransky E. Exploring the link between glucocerebrosidase mutations and parkinsonism. Trends Mol Med. 2011;17:485–93.

119. Wilcken B. Problems in the management of urea cycle disorders. Mol Genet Metab. 2004; 81(Suppl 1):S86–S91.

120. Wilcox WR. Lysosomal storage disorders: the need for better pediatric recognition and comprehensive care. J Pediatr. 2004; 144:S3-S14

121. Wilson DO, Johnson P, McCord BR. Nonradiochemical DNase I footprinting by capillary electrophoresis. Electrophoresis. 2001; 22(10):1979-86.

122. Woychik NA and Hampsey M. The RNA polymerase II machinery: structure illuminates function. Cell. 2002; 108(4):453-63.

123. Xu YH, Barnes S, Sun Y, Grabowski GA. Multi-system disorders of glycosphingolipid and ganglioside metabolism. J Lipid Res. 2010; 51:1643–75.

124. Yamaguchi S, Brailey LL, Morizono H, Bale AE, Tuchman M. Mutations and polymorphisms in the human ornithine transcarbamylase (OTC) gene. Hum Mutat. 2006; 27(7):626-32.

125. Yang C, Bolotin E, Jiang T, Sladek FM, Martinezd E. Prevalence of the Initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene. 2007; 389(1):52–65

126. Zianni M, Tessanne K, Merighi M, Laguna R, Tabita FR. Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument. J Biomol Tech. 2006; 17(2):103-13.

9. List of publications and presentations

9.1. Publications related to this thesis

Ref I:

Storkanova G, Vlaskova H, Chuzhanova N, Zeman J, Stranecky V, Majer F, Peskova K, **Luksan O**, Jirsa M, Hrebicek M, Dvorakova L. Ornithine carbamoyltransferase deficiency: molecular characterization of 29 families. Clin Genet. 2013; 84(6):552-9. **IF(2012)=4.247**

Ref II:

Lukšan O, Jirsa M, Eberová J, Minks J, Trešlová H, Boučková M, Štorkánová G, Vlášková H, Hřebíček M, Dvořáková L. Disruption of OTC promoter-enhancer interaction in a patient with symptoms of ornithine carbamoyltransferase deficiency. Hum Mutat. 2010; 31(4):E1294-E1303. **IF=5.956**

Ref III:

Lukšan O, Dvořáková L, Jirsa M. HNF-4α regulates expression of human ornithin transcarbamylase through interaction with two positive *cis*-acting regulatory elements located in the proximal promoter. Fol Biol. 2014; *accepted manuscript (Annex)*. **IF(2012)=1.219**

Ref IV:

Svobodová E, Mrázová L, **Lukšan O**, Elstein D, Zimran A, Stolnaya L, Minks J, Eberová J, Dvořáková L, Jirsa M, Hřebíček M. Glucocerebrosidase gene has an alternative upstream promoter, which has features and expression characteristic of housekeeping genes. Blood Cells Mol Dis. 2011; 46(3):239-45. **IF=2.351**

9.2. Publications not related to this thesis

1. Neřoldová M, Fraňková S, Stránecký V, Honsová E, **Lukšan O**, Beneš M, Michalová K, Kmoch S, Jirsa M. Hereditary haemochromatosis caused by homozygous HJV mutation evolved through paternal disomy. Clin Genet. 2014 Feb 12. doi: 10.1111/cge.12346. **IF(2012)=4.247**

2. Subhanova I, Muchova L, Lenicek M, Vreman HJ, **Luksan O**, Kubickova K, Kreidlova M, Zima T, Vitek L, Urbanek P. Expression of Biliverdin Reductase A in peripheral blood leukocytes is associated with treatment response in HCV-infected patients. PLoS One. 2013; 8(3):e57555. **IF(2012)=3.730**

3. Sticova E, Elleder M, Hulkova H, **Luksan O**, Sauer M, Wunschova-Moudra I, Novotny J, Jirsa M. Dubin-Johnson syndrome coinciding with colon cancer and atherosclerosis. World J Gastroenterol. 2013; 19(6):946-50. **IF(2012)=2.547**

4. Fafilek B, Krausova M, Vojtechova M, Pospichalova V, Tumova L, Sloncova E, Huranova M, Stancikova J, Hlavata A, Svec J, Sedlacek R, **Luksan O**, Oliverius M, Voska L, Jirsa M, Paces J, Kolar M, Krivjanska M, Klimesova K, Tlaskalova-Hogenova H, Korinek V. Troy, a Tumor Necrosis Factor Receptor Family Member, Interacts With Lgr5 to Inhibit Wnt Signaling in Intestinal Stem Cells. Gastroenterology. 2012; 144(2):381-91. **IF=12.821**

5. Treepongkaruna S, Gaensan A, Pienvichit P, **Luksan O**, Knisely AS, Sornmayura P, Jirsa M. Novel ABCB11 mutations in a Thai infant with progressive familial intrahepatic cholestasis. World J Gastroenterol. 2009; 15(34):4339-42. **IF=2.092**

9.3. Conference Presentations related to this thesis

1. **Luksan O**, Dvorakova L, Jirsa M (2013): Characterization of the human ornithine carbamoyltransferase promoter. European Human Genetics Conference 2013 (ESHG), Paris, Fr **(Poster)**

2. **Lukšan O**, Dvořáková L, Jirsa M (2013): Charakterizace základních oblastí podílejících se na regulaci exprese genu pro OTC u člověka. XLI. Májové hepatologické dny 2013 – kongres České hepatologické společnosti (ČHS), Karlovy Vary, ČR **(Poster)**

3. **Luksan O**, Dvorakova L, Jirsa M (2012): Characterization of the basic regulatory elements involved in the regulation of OTC gene expression in human. Annual Meeting of The American Society of Human Genetics, San Francisco, USA **(Poster)**

4. **Lukšan O**, Dvořáková L, Jirsa M (2012): Charakterizace promotoru genu pro OTC. XL. Májové hepatologické dny 2012 – kongres České hepatologické společnosti, Karlovy Vary, ČR **(Poster)**

5. **Lukšan O**, Jirsa M Hřebíček M, Dvořáková L (2011): Mutation in the OTC promoter may cause symptoms of ornithine carbamoyltransferase deficienty. European Human Genetics Conference, Amsterdam, NL **(Poster)**

6. **Luksan O**, Jirsa M, Eberova J, Hrebicek M, Dvorakova L (2010): Promoter-enhancer interaction is responsible for tissue specific expression of *OTC,* the disruption may cause the manifestation of ornithine carbamoyltransferase deficiency. 60th Annual Meeting of The American Society of Human Genetics, Washington, USA **(Poster)**

7. Hrebicek M, Svobodova E, Mrazova L, **Luksan O**, Stolnaja L, Jirsa M, Dvorakova L (2010): Glucocerebrosidase alternative promoter has features and expression characteristic of housekeeping genes. 60th Annual Meeting of The American Society of Human Genetics, Washington, USA **(Poster)**

8. **Lukšan O**, Jirsa M, Budišová L, Hřebíček M, Dvořáková L (2010): Mutací vyvolaná porucha interakce promotoru OTC s enhancerem jako příčina kongenitální hyperamonémie. XXXVIII. Májové hepatologické dny 2010 – kongres České hepatologické společnosti, Karlovy Vary, ČR **(Oral presentation)**

9. **Luksan O**, Eberova J, Bouckova M, Jirsa M, Hrebicek M, Dvorakova L (2009): Characterization and functional analysis of OTC gene regulatory regions. 59th Annual Meeting of The American Society of Human Genetics, Honolulu, USA **(Poster)**

10. **Lukšan O**, Hřebíček M, Jirsa M, Dvořáková L (2009): Nové přístupy k molekulárně genetické analýze u pacientů s deficitem ornithin transkarbamylázy. XXXVII. Májové hepatologické dny 2009 – kongres České hepatologické společnosti, Karlovy Vary, ČR **(Poster)**

9.4. Conference Presentations not related to this thesis

1. Neroldova M, Frankova S, Stranecky V, Honsova E, **Luksan O**, Benes M, Michalova K, Kmoch S, Jirsa M (2013): Segmental paternal isodisomy as a cause of juvenile hemochromatosis type 2a. European Human Genetics Conference 2013 (ESHG), Paris, Fr **(Poster)**

2. Koblihová E, **Lukšan O**, Mrázová I a Ryska M (2013): Transplantace hepatocytů v terapii akutního jaterního selhání u potkana. XLI. Májové hepatologické dny 2013 – kongres České hepatologické společnosti (ČHS), Karlovy Vary, ČR **(Poster)**

3. Neřoldová M, Fraňková S, Stránecký V, Honsová E, **<u>Lukšan O</u>**, Beneš M, Michalová K, Kmoch S, Jirsa M (2013): Segmentální paternální isodisomie jako příčina juvenilní hemochromatózy typu 2a. XLI. Májové hepatologické dny 2013 – kongres České hepatologické společnosti (ČHS), Karlovy Vary, ČR **(Oral presentation)**

4. Subhanova I, Vitek L, Muchova L, Lenicek M, **<u>Luksan O</u>**, Zima T, Urbanek P (2012): Biliverdin reductase A expression in peripheral blood mononuclear cells is associated with treatment response to antiviral therapy in chronic hepatitis C patients. 63rd Annual Meeting of the American-Association-for-the-Study-of-Liver-Diseases (AASLD), Boston, USA **(Poster)**

5. Neřoldová M, Fraňková S, Beneš M, Honsová E, **<u>Lukšan O</u>**, Jirsa M (2012): Juvenilní hemochromatóza – první zkušenosti a literární přehled. Pracovní den OKB FN a OS ČLK Olomouc, Olomouc, ČR **(Oral presentation)**

6. Neřoldová M, Fraňková S, Beneš M, Honsová E, **<u>Lukšan O</u>**, Budišová L, Jirsa M (2012): Juvenilní hemochromatóza z deficitu hemojuvelinu – kasuistika. XL. Májové hepatologické dny 2012 – kongres České hepatologické společnosti, Karlovy Vary, ČR **(Oral presentation)**

7. Subhanová I, Muchová L, Leníček M, Vítek L, **<u>Lukšan O</u>**, Zima T, Urbánek P (2011): Biliverdinreduktáza a efekt antivirové terapie u chronické hepatitidy C. XXXIX. Májové hepatologické dny 2011 – kongres České hepatologické společnosti, Karlovy Vary, ČR **(Oral presentation)**

8. Zindr M, **<u>Lukšan O</u>**, Budišová L, Lewindon P, Ee L, Knisely AS, Jirsa M (2011): Rozsáhlá dalece v chromozomu 7 jako příčina progresivní familiární intrahepatální cholestázy 3. typu. XXXIX. Májové hepatologické dny 2011 – kongres České hepatologické společnosti, Karlovy Vary, ČR **(Poster)**

9. Sticová E, **<u>Lukšan O</u>**, Mrázová I, Jirsa M (2010): Úloha matrix metaloproteinázy 9 v progresi jaterního poškození u modelu biliární léze. XXXVIII. Májové hepatologické dny 2010 – kongres České hepatologické společnosti, Karlovy Vary, ČR **(Poster)**

10. Subhanová I, Muchová L, Leníček M, Vítek L, **<u>Lukšan O</u>**, Zima T, Urbánek P (2010): Může biliverdinreduktáza u HCV infikovaných pacientů predikovat odpověď na antivirovou terapii? XXXVIII. Májové hepatologické dny 2010 – kongres České hepatologické společnosti, Karlovy Vary, ČR **(Oral presentation)**

Acknowledgements