UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SALUTE DELLA DONNA E DEL BAMBINO

Scuola di Dottorato di Ricerca in Medicina dello Sviluppo e Scienze della Programmazione

Indirizzo: Malattie Rare, Genetica, Biologia e Biochimica

Ciclo XXV

ANALYSIS OF HUNTER SYNDROME BY RNA-SEQUENCING

DIRETTORE DELLA SCUOLA: Prof. Giuseppe Basso

COORDINATORE D'INDIRIZZO: Prof. Giorgio Perilongo

SUPERVISORE: Dott. Maurizio Scarpa

CO-SUPERVISORE: Dott.ssa Rosella Tomanin

DOTTORANDA: Francesca D'Avanzo

ABSTRACT

Hunter Syndrome (Mucopolysaccharidosis type II, MPS II) is a rare inherited metabolic disease due to an extremely reduced or total absent activity of the lysosomal enzyme iduronate 2-sulfatase (IDS), involved in the degradation of the mucopolysaccharides heparan- and dermatan-sulphate. This causes a progressive pathologic accumulation of the two macromolecules within cell lysosomes and in the extracellular matrix of most tissues and organs, leading to their general malfunctioning and finally to death. In fact, due to the housekeeping nature of IDS, most of the organ systems are involved in the pathology, including the central nervous system in the severe forms of the disease.

MPS II belongs to the group of Mucopolysaccharidoses (MPSs), a cluster of pathologies characterized by accumulation of mucopolysaccahrides (or glycosaminoglycans, GAG). They, in turn, represent a subgroup of the wider class of the Lysosmal Storage Disorders (LSDs), about fifty pathological conditions characterized by the progressive endo- and extra-cellular overstorage of several types of undegraded macromolecules.

LSDs, for long time poorly considered by the medical-scientific community, have received in the past few years an increasing attention due to their elevated overall incidence, up to 1:1500-1:7000 live newborns, dependently on the population analyzed. Although the enzyme or protein defect underlying each of these pathologies is known, almost unknown remains the complexity of the biochemical pathways involved or altered in the lysosomal storage in general, or in specific type of storage. Studies conducted in the last decade have separately highlighted alterations of signalling proteins, intracellular calcium homeostasis, oxidative stress, autophagy, intracellular trafficking, lipid biosynthesis and iron metabolism. However, no systematic and complete studies have been so far conducted for the analysis of the whole pathologic scenario. This would help to acquire a general overview of the lysosomal storage and would also help in defining new, potential therapeutic targets and/or biomarkers useful in the diagnosis, prognosis, progression of LSDs as well as in the evaluation of efficacy of the therapeutic strategies applied. Moreover, since LSDs share several pathological signs and symptoms, it appears evident that a deep analysis of some of them could be of great help in the understanding of others.

For the first time, this project evaluated, by an high throughput technology, the whole transcriptome profile of LSD cells by comparing skin fibroblasts obtained from Hunter patients and healthy controls, thus allowing a deep analysis of MPS II pathogenesis. The

study, conducted by total RNA sequencing, was performed by using the SOLiD technology. Results have shown alterations in: 1) basic cellular processes as cell cycle, apoptosis, intercellular communication; 2) metabolic processes as proteoglycan metabolism, synthesis of lipids, aminoacids and nucleotides; 3) response to stimuli as oxidative stress, insulin, cytokines; 4) alteration of the developmental processes.

From the therapeutic point of view, as for MPS II the major treatment strategy used in the last 5-6 years is represented by the Enzyme Replacement Therapy (ERT), consisting in the weekly systemic administration of the active form of the enzyme, which is missing in the patients. Clinical monitoring of the patients under treatment, organized since 2005 by Shire HGT, has shown, among other results, an important subjectivity in the efficacy of ERT, as expected for a pathology presenting several degrees of severity and a high number of different mutations. Despite this, ERT is administered to all patients following the same therapeutic protocol. Therefore, it becomes necessary to perform a deep clinical and molecular evaluation to identify potential candidate biomarkers of efficacy allowing an adequate follow-up of the patients under ERT; this would allow the set up of a personalized therapeutic protocol.

Starting from these considerations, in this project an *in vitro* evaluation of ERT has been performed in Hunter primary fibroblasts treated with the therapeutic IDS enzyme and collected 24 and 144 h post-treatment. Their transcriptional profile has been studied to characterize the early cellular response to the enzyme supply. Such analysis allowed to highlight 20 candidate biomarkers of therapeutic efficacy. Some of these have been afterwards evaluated by using Real Time PCR, in blood samples obtained from Hunter patients under ERT. Finally, a correlation analysis was performed between clinical parameter obtained by the follow-up of the Hunter population and the gene expression profile of each gene. Such analysis has shown a good correlation for 8 pairs of gene/parameter evaluated. In particular, correlations were found for hearing impairment, seizures, hepatomegaly, splenomegaly, and other clinical parameters, with at least one gene. The analysis of the other candidate genes isolated from transcriptome analysis might indentify other potential biomarkers.

RIASSUNTO

La Sindrome di Hunter (o Mucopolisaccaridosi di tipo II, MPS II) è una malattia metabolica ereditaria rara, causata da un'attività estremamente ridotta o del tutto assente dell'enzima lisosomiale iduronato 2-solfatasi (IDS), deputato alla degradazione dei mucopolisaccaridi eparan- e dermatan-solfato. Tale deficit determina un progressivo accumulo patologico delle due macromolecole sia nei lisosomi cellulari che nella matrice extracellulare di quasi tutti i tessuti ed organi, conducendo progressivamente ad un malfunzionamento generale e, infine, alla morte. Infatti, essendo l'IDS un enzima ubiquitario, quasi tutti i distretti risultano compromessi, compreso il sistema nervoso centrale nelle forme severe della malattia.

La MPS II appartiene al gruppo delle mucopolisaccaridosi (MPSs), un cluster di malattie caratterizzate proprio dall'accumulo di mucopolisaccaridi (o glicosaminoglicani, GAG. Esse rappresentano, a loro volta, un sottogruppo della più ampia classe delle malattie da accumulo lisososmiale (Lysosomal Storage Disorders, LSD), una cinquantina di patologie caratterizzate dall'accumulo endo- ed extra-cellulare di diversi tipi di macromolecole non degradate.

Le LSD, a lungo trascurate dalla comunità medico-scientifica, negli ultimi anni hanno ricevuto una maggiore attenzione a causa della loro elevata incidenza complessiva, fino a 1:1500–1:7000 nati vivi, anche dipendentemente dalla popolazione analizzata. Nonostante il difetto enzimatico o comunque proteico alla base di ciascuna di queste patologie sia ormai noto, per lo più sconosciuta rimane la complessità dei pathways biochimici che risultano coinvolti o alterati nell'accumulo lisosomiale in generale, o in specifici tipi di accumulo. Gli studi condotti nell'ultima decade hanno separatamente evidenziato alterazioni a carico di proteine di segnale, dell'omeostasi del calcio endocellulare, dello stress ossidativo, dell'autofagia, del trafficking intracellulare, della biosintesi dei lipidi e del metabolismo del ferro. Nessuno studio è stato, tuttavia, condotto in modo sistematico e completo per l'analisi dell'intero quadro patologico. Ciò aiuterebbe non solo ad acquisire una visione complessiva del problema dell'accumulo lisosomiale, ma anche alla messa in luce di nuovi, potenziali target terapeutici e/o di biomarcatori utilizzabili nella diagnosi delle patologie, nella definizione della loro prognosi e progressione, nella valutazione di efficacia terapeutica dei trattamenti applicati. Inoltre, poiché le LSD condividono numerosi segni e sintomi patologici, è

evidente come lo studio approfondito di alcune potrebbe risultare di grande aiuto anche per la comprensione delle altre.

Per la prima volta questo progetto ha valutato con tecnologia high throughput l'intero trascrittoma di cellule LSD mediante comparazione di fibroblasti cutanei ottenuti da pazienti Hunter e da controlli sani, consentendo uno studio approfondito della patogenesi della MPS II. Lo studio, condotto mediante sequenziamento di tutto l'mRNA cellulare, è stato effettuato utilizzando la tecnologia SOLiD. I risultati hanno evidenziato alterazioni a livello di: 1) processi cellulari di base, quali il ciclo cellulare, l'apopotosi, la comunicazione intercellulare; 2) processi metabolici quali il metabolismo dei proteoglicani, la sintesi dei lipidi, degli aminoacidi e dei nucleotidi; 3) la risposta agli stimoli quali lo stress ossidativo, l'insulina, le citochine; 4) l'alterazione dei processi dello sviluppo.

Dal punto di vista del trattamento, nel caso della MPS II, valutata in questo progetto di studio, la terapia maggiormente applicata negli ultimi 5-6 anni è rappresentata dalla sostituzione enzimatica (Enzyme Replacement Therapy, ERT), che consiste nella somministrazione sistemica settimanale della forma attiva dell'enzima che è deficitario nei pazienti. Il monitoraggio clinico dei pazienti in trattamento, organizzato a partire dal 2005 dalla ditta che distribuisce il farmaco (Shire HGT) ha evidenziato, tra le altre cose, una importante soggettività nell'efficacia della terapia, come atteso per un trattamento effettuato per una patologia con diverse forme di severità, causata da un elevato numero di mutazioni diverse. Tuttavia l'ERT è di norma somministrato a tutti i pazienti con il medesimo protocollo. Da qui la necessità di effettuare uno studio approfondito sia clinico che molecolare allo scopo di individuare dei potenziali candidati a biomarcatori di efficacia che consentano un follow-up adeguato dei pazienti in ERT; ciò permetterebbe la messa a punto di un protocollo terapeutico personalizzato.

A partire da queste considerazioni, in questo progetto è stata effettuata una valutazione dell'ERT *in vitro*, in fibroblasti primari Hunter trattati con l'enzima IDS terapeutico e raccolti 24 e 144 ore dall'inizio del trattamento. Il loro profilo trascrizionale è stato studiato allo scopo di caratterizzare la risposta cellulare precoce alla somministrazione dell'enzima. Tale analisi ha consentito di evidenziare una ventina di geni candidati a marcatori di efficacia terapeutica. Alcuni di questi sono stati poi valutati, mediante Real Time PCR, in alcuni campioni ematici provenienti da una popolazione di soggetti Hunter in ERT. Infine, è stato effettuato uno studio di correlazione tra l'andamento osservato

dei marcatori molecolari e l'andamento di alcuni parametri clinici, provenienti dal follow-up clinico della popolazione Hunter studiata. Tale analisi ha evidenziato una buona correlazione per 8 appaiamenti gene candidato/parametro clinico valutato; in particolare, correlazioni con almeno un gene sono state trovate per la sordità, le crisi epilettiche, l'epatomegalia, la splenomegalia e altri parametri clinici. E' auspicabile che l'estensione di tale valutazione ai rimanenti geni candidati metta in luce altri potenziali candidati biomarcatori di efficacia terapeutica.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Hunter Syndrome

Hunter Syndrome, or Mucopolysaccharidosis type II (MPS II; OMIM #309900), belongs to a group of lysosomal storage disorders (LSDs) known as Mucopolysaccharidoses (MPS). It is caused by the deficit of activity of iduronate 2-sulfatase (IDS), a key enzyme in the lysosomal catabolism of the mucopolysaccharides heparan- and dermatan-sulfate.

1.1.1 Disease context

Lysosomal storage disorders (LSDs) are rare inherited metabolic diseases mostly due to the deficit or reduced activity of specific lysosomal hydrolases responsible for the degradation of cellular macromolecules. In just a few cases, pathology is due to defective non-enzymatic lysosomal proteins or to non-lysosomal proteins such as membrane proteins, enzyme activators or proteins involved in enzyme targeting and lysosomal biogenesis (Filocamo, Morrone 2011).

The common biochemical hallmark of these diseases is the accumulation, mostly inside lysosomes, of undegraded or partially degraded metabolites. Due to the ubiquitous expression of lysosomal hydrolases, these pathologies are multiorgan and multisystemic, mostly life-threatening, with more than 75% of the patients presenting with degenerating neurological involvement (Sands, Haskins 2008). LSDs present a highly variable phenotypic expression depending on the specifically stored macromolecules, the main anatomic site of production and degradation of the specific metabolites, the residual enzymatic activity and the general genetic background of the patient (Filocamo, Morrone 2011). Clinical conditions generally tend to worsen as more waste substrate accumulates over time, with symptoms varying in severity from mild to severe manifestations, even in the same disease. At present, more than 50 LSDs are known; all LSDs are monogenic disorders inherited as autosomal recessive traits, except for Fabry disease, Hunter Syndrome and Danon disease, which are X-linked conditions. Although individually rare, taken as a whole LSDs present an overall incidence of about 1:1500–1:7000 live births (Staretz-Chacham et al. 2009) also depending on the population examined.

In the past, their low individual incidence has hindered the development of effective therapies because they have been mostly neglected by the pharmaceutical industry. Thus, for many years patients' management mainly consisted of supportive care and treatment of complications. More recently, notable improvements have been made in this field. To date, the most effective strategy proposed seems to be enzyme replacement therapy (ERT) consisting in the weekly/biweekly i.v. infusion of the therapeutic enzyme. ERT has been developed for Gaucher, Fabry and Pompe diseases, for mucopolysaccharidoses type I, II and VI and has generally shown to be effective in treating systemic symptoms and delaying the disease progression (Tomanin et al. 2012). However it cannot treat the neurological deficits, because of the inability of the therapeutic enzyme to cross the blood–brain barrier (BBB) (Beck 2010). Other limitations are the need of repeated infusions, often requiring hospitalization, and the elevated costs sustained by the health care systems.

Haematopoietic stem cell transplantation (HSCT) has led to good results in some cases (mucopolysaccharidosis type I, metachromatic leukodystrophy, Krabbe disease and α-mannosidosis), but only if the procedure is performed at an early stage of disease progression. Also it is limited by the availability of compatible donors and by significant risks of morbidity and mortality and of primary and secondary graft failure (Beck 2010). Substrate reduction therapy, aimed to reduce biosynthesis of new storage material is another strategy presently applied to Gaucher type I and Niemann–Pick type C diseases (Beck 2010).

Among LSDs, Mucopolysaccharidoses (MPSs) are due to a defective mucopolysaccharide metabolism. Each MPS is caused by a deficit in the activity of one of the 11 lysosomal enzymes required for lysosomal sequential degradation of mucopolysaccharides (also called glycosaminoglycans, GAGs). MPSs are usually classified according to their genetic base. The biochemical hallmark of these diseases is the accumulation within lysosomes and extracellular matrix of partially degraded GAG which are also detected at high levels in urine, blood and cerebral spinal fluid (Muenzer 2011).

As LSDs in general, also MPSs are heterogeneous, progressive disorders. Patients typically appear normal at birth, but during early childhood they experience the onset of clinical signs and symptoms including organomegaly, skeletal, joint, airway and cardiac involvement, hearing and vision impairment, and mental retardation in the severe forms of MPS I, MPS II and MPS VII and all subtypes of MPS III (Muenzer 2011).

All MPSs are autosomal recessive disorders, with the exception of MPS II which presents X-linked recessive inheritance. The overall prevalence of MPSs is difficult to assess because of the lack of epidemiological studies; the known incidence values are shown in Table 1.1.

Table 1.1 Mucopolysaccharidoses. Modified from (Muenzer 2011)

Disorder	GAG storage material	Deficient enzyme	Incidence per 100000 live births
MPS I (Hurler, Hurler-Scheie, Scheie syndromes)	dermatan sulphate, heparan sulphate	α-L-iduronidase	0.69 - 1.66
MPS II (Hunter syndrome)	dermatan sulphate, heparan sulphate	iduronate 2-sulphatase	0.30 - 0.71
MPS III (Sanfilippo syndrome) A	heparan sulphate	heparan N-sulphatas	0.29 - 1.89
MPS III (Sanfilippo syndrome) B	heparan sulphate	α-N-acetylglucosaminidase	0.42 - 0.72
MPS III (Sanfilippo syndrome) C	heparan sulphate	acetyl-CoA: α-glucosaminide acetyltransferase	0.07 - 0.21
MPS III (Sanfilippo syndrome) D	heparan sulphate	N-acetylglucosamine 6-sulphatase	0.1
MPS IV (Morquio syndrome) A	keratan sulphate, chondroitin sulphate	galactose 6-sulphatase	0.22 - 1.3
MPS IV (Morquio syndrome) B	keratan sulphate	β-galactosidase	0.02 - 0.14
MPS VI (Maroteaux-Lamy syndrome)	dermatan sulphate, chondroitin sulphate	arylsulphatase B	0.36 - 1.30
MPS VII (Sly syndrome)	dermatan sulphate, heparan sulphate, chondroitin sulphate	β-Glucuronidase	0.05 - 0.29
MPS IX	hyaluronan	hyaluronidase	four cases reported

Concerning treatment, as previously mentioned, therapies presently available are haematopoietic stem cell transplantation for MPS I and VI and enzyme replacement therapy for MPS I, II and VI.

1.1.2 Genetics

1.1.2.1 Molecular genetics

The human gene coding for iduronate 2-sulfatase (IDS) maps on the long arm of X chromosome, at location Xq28, with minus orientation (Figure 1.1). According to Ensembl database, it spans almost 56 kilobases and has 11 transcripts.

Figure 1.1 IDS Gene in genomic location [http://www.genecards.org].

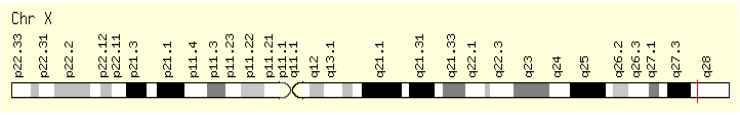


Figure 1.2 and Table 1.2 show respectively the transcript diagram and table of IDS gene from Ensembl website (http://www.ensembl.org/).

IDS-001, the longest transcript, includes 9 exons and codes for the so called long isoform of IDS protein (550 aa). The other transcripts, derived from alternative splicing, are of unknown function, even those which are protein coding. A pseudogene (IDSP1, I2S2) located on the telomeric side of the IDS gene within 20 kb has also been identified (Bondeson et al. 1995); it contains sequences homologous to exons 2 and 3 and to introns 2, 3, and 7 of the longest IDS transcript, in reverse orientation.

Figure 1.2. Transcript diagram of IDS gene from Ensembl genome browser. Individual transcripts are drawn as boxes for exons and connecting lines for introns. Filled boxes show coding sequence, and empty boxes show UTR (Untranslated Region). Red or gold transcripts are protein coding: red transcript comes from either the Ensembl automatic annotation pipeline (number beginning with 2) or manual curation by the VEGA/Havana project (number beginning with 0). A gold, or merged, transcript is identical between Ensembl automated annotation and VEGA/Havana manual curation. Processed transcripts, shown in blue, are noncoding transcripts that do not contain open reading frame (ORF). This type of transcript is annotated by VEGA/Havana. [http://www.ensembl.org/].

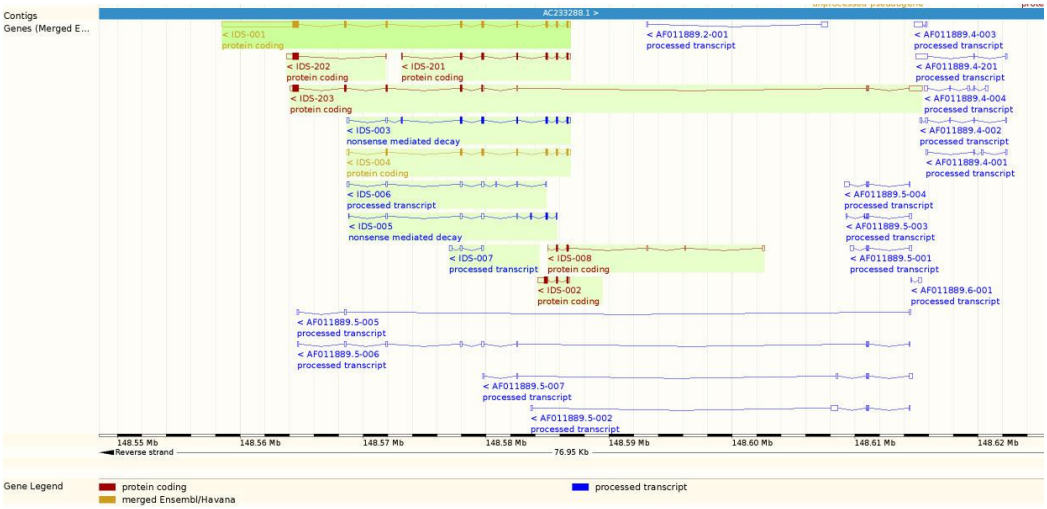


Table 1.2 Table of IDS gene from Ensembl website (http://www.ensembl.org/).

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
IDS-001	ENST00000340855	7619	ENSP00000339801	550	Protein coding	CCDS14685
IDS-203	ENST00000541269	2635	ENSP00000441261	339	Protein coding	-
IDS-003	ENST00000466323	1529	ENSP00000418264	312	Nonsense mediated decay	-
IDS-004	ENST00000370441	1399	ENSP00000359470	343	Protein coding	CCDS14686
IDS-002	ENST00000428056	1213	ENSP00000390241	179	Protein coding	-
IDS-201	ENST00000370443	1156	ENSP00000359472	312	Protein coding	-
IDS-202	ENST00000537071	1032	ENSP00000440324	153	Protein coding	-
IDS-005	ENST00000464251	959	ENSP00000428980	106	Nonsense mediated decay	-
IDS-006	ENST00000490775	945	No protein product	-	Processed transcript	-
IDS-008	ENST00000521702	572	ENSP00000429745	86	Protein coding	-
IDS-007	ENST00000466019	537	No protein product	-	Processed transcript	-

To date, 479 mutations of IDS gene have been reported [www.hgmd.cf.ac.uk/], many of which are private (Burton, Giugliani 2012, Froissart, Da Silva & Maire 2007). They include single nucleotide substitutions, small deletions, insertions and indels, gross deletions, insertions and duplications, complex rearrangements and splicing mutations (Table 1.3).

Table 1.3 Human IDS gene mutations. Data retrieved from Human Gene Mutation Database, accessed 14 January 2012.

Mutation type	Number of mutations	%
Missense/nonsense	244	50,94
Splicing	44	9,19
Regulatory	0	0,00
Small deletions	87	18,16
Small insertions	39	8,14
Small indels	9	1,88
Gross deletions	38	7,93
Gross insertions/duplications	4	0,84
Complex rearrangements	14	2,92
Repeat variations	0	0,00
TOTAL	**479**	

According to a study conducted on 155 unrelated patients in 2007 (Froissart, Da Silva & Maire 2007), most of them (~80%) present with small IDS gene alterations, the others present with large gene alterations.

1.1.2.2 Inheritance

MPS II is inherited as X-linked recessive disorder; nevertheless it has been well documented in a small number of females (Tuschl et al. 2005, Pina-Aguilar et al. 2012) in whom the most common mechanism of expression is skewed X-chromosome inactivation of the non mutant allele. Additional rare causes may include the disruption of the IDS gene through an X chromosome/autosome translocation or homozygosity for an IDS gene mutation (Tuschl et al. 2005). Females carrying a mutation in one IDS allele are usually asymptomatic.

1.1.2.3 Population genetics

Like the other LSDs, Hunter Syndrome is rare, but is one of the most common MPSs with an estimated prevalence of about 1 in 170000 male live births. Table 1.4 shows the values of incidence per 100000 male live births in different populations. Its prevalence among Ashkenazi and Oriental or Sephardic Jews living in Israel is approximately twice than other populations (Martin et al. 2008).

Table 1.4 Incidence of MPS II. Incidence is calculated per 100000 male live births except for data indicated with * which are calculated per 100000 live births and data indicated with † which are calculated per 100000 boys aged 0-14 years.

Population	Incidence	Reference
Germany	1,3	(Baehner et al. 2005)
The Netherlands	1,3	(Poorthuis et al. 1999)
Northern Ireland	1,39	(Nelson 1997)
Taiwan	2,05	(Lin et al. 2009)
Tunisia	0,29	(Ben Turkia et al. 2009)
Estonia	2,16	(Krabbi et al. 2012)
Western Australia	0,6	(Nelson et al. 2003)
Australia	0.62*	(Meikle et al. 1999)
Macedonia	0.36†	(Gucev et al. 2011)
Bulgaria	0.46†	(Gucev et al. 2011)

1.1.2.4 Genotype-phenotype correlation

The high phenotypic and genotypic heterogeneity, the rarity and the presence of almost exclusively private mutations render genotype-phenotype correlation in MPS II difficult to do. However, total or partial gene deletion and gene/pseudogene rearrangement seem to result in the severe phenotype (Froissart et al. 2002). Deletions extending beyond the IDS locus have

been associated with a severe phenotype in combination with symptoms atypical of Hunter syndrome (Timms et al. 1997). Conversely, small gene alterations, such as single nucleotide substitutions, have been reported to be associated with a wide range of phenotypes, spanning the entire spectrum from severe to attenuated. Even the same mutation may be associated with different phenotypes: at least two families have been described with brothers exhibiting severe and attenuated forms (Martin et al. 2008).

Neither the amount of IDS nor its activity correlates with phenotype severity in Hunter patients. Current methods used for measuring IDS enzymatic activity are insufficiently sensitive to differentiate between complete absence and presence of residual activity (Froissart et al. 2002). However, *in vitro* functional studies of IDS mutations reported a residual activity equal to 0.2–2.4% of the wild-type IDS activity for attenuated phenotypes and no detectable activity for those associated with severe phenotypes (Sukegawa-Hayasaka et al. 2006). Lee and colleagues (Lee et al. 2012) observed a low residual IDS activity also in plasma of attenuated patients.

1.1.3 Biochemistry

Iduronate 2-sulfatase (α-L-iduronate sulfate sulfatase, idursulfase) protein (IDS, or I2S, UniProtKB ID: P22304, EC=3.1.6.13) consists of 550 amino acids and is synthesized as a 90 kDa precursor. Thanks to the N-terminal signal peptide of 25 aa, it is co-translationally transported into endoplasmic reticulum (ER) lumen, where it undergoes cleavage of the signal peptide and N-glycosylation. IDS bears 7 potential N-glycosylation sites. From ER, the enzyme is transported to the Golgi apparatus where a mannose 6-phosphate (M6P) group is added as marker for lysosomal sorting. The M6P signal is recognized by M6P receptors bound to the adaptor proteins of clathrin vesicle coat; the clathrin-coated vesicles then bud off from the TGN and transport lysosomal enzyme to early endosomes. The active form of IDS is obtained after proteolytic cleavege leading to the formation of two polypeptide chains of 42 and 14 KDa.

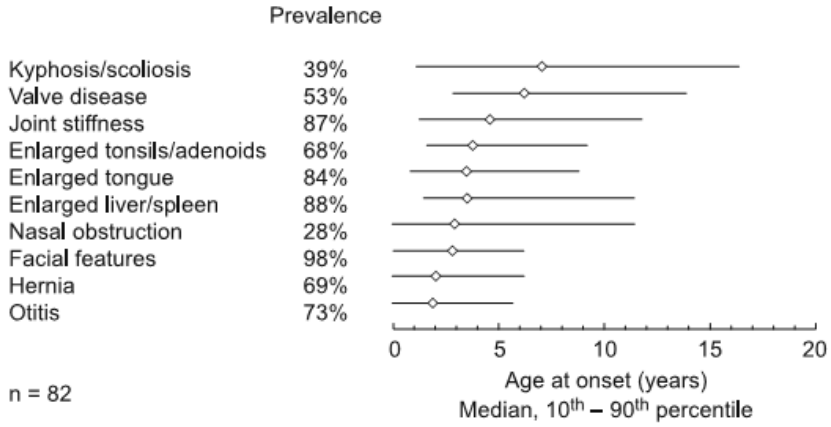
IDS catalyzes a step of glycosaminnoglycan degradation. In particular it removes the sulfate group at position 2 of L-iduronic acid in dermatan- and heparan-sulfate. It presents a high substrate specificity and optimum pH around 4; it is inhibited by many salts especially sulphate and phosphate ones.

IDS is an ubiquitous enzyme; in human it has been purified from liver, kidney, lung, placenta, plasma and urine.

1.1.4 Clinical aspects

MPS II is a multiorgan and multisystem disease with variable age of onset and rate of progression. Figure 1.3 shows the age at which the main signs and symptoms of the disease appear (Wraith et al. 2008).

Figure 1.3 Reported age at onset and prevalence of clinical features in 82 patients with mucopolysaccharidosis type II (Hunter syndrome) enrolled in HOS, the Hunter Outcome Survey (Wraith et al. 2008).



Phenotypic expression spans a wide spectrum of clinical severity. Patients with the more severe form exhibit a chronic and progressive disease involving multiple organs and tissues, with increasing cognitive deterioration. They appear normal at birth, although they tend to be heavy, some have inguinal or umbilical hernias and there is an enhanced incidence of Mongolian blue spots. Clinical features appear between 2 and 4 years of age. Death from a combination of neurological deterioration and cardiorespiratory failure usually occurs in the mid-teenage years (Wraith et al. 2008, Martin et al. 2008). In attenuated patients, clinical signs have a later onset; they are most often diagnosed between the ages of 4 and 8 years. Neurologic dysfunctions are absent or minimal. Survival to adulthood is common, but death often occurs between 20 and 30 years from cardiac or respiratory failure. A few patients are less severely affected. In these, life expectancy can be near normal, and such affected males may have children. However, this traditional classification of patients into ‘mild’ (sometimes

reported as ‘attenuated’) or ‘severe’ forms, on the basis of length of survival and the presence of CNS disease, is a gross simplification. The disorder should rather be regarded as a continuum between two extremes (severe and attenuated) (Wraith et al. 2008).

Patients appear normal at birth; coarsening of facial features, which is the most common presenting trait, generally manifests between 18 months and 4 years of age in the severe forms and about 2 years later for the attenuated ones. The distinctive facies is characterized by broad nose with flared nostrils, prominent supraorbital ridge, large jowls, thick lips and macroglossia; macrocephaly is observed throughout life.

In the first years of life the height of most MPS II children is above the 50th percentile and in some cases it is over the 97th percentile. However, growth velocity decreases with age: starting from eight years height is below the third percentile, and nearly all children exhibit growth retardation before puberty (Schulze-Frenking et al. 2011). The cause of short stature is unknown; it may be related to osseous growth-plate disturbances.

Subjects affected by MPS II have a thickened and inelastic skin sometimes bringing distinctive lesions which are described as ivory-white papules of 2 to 10 mm in diameter, often coalescing to form ridges.

Hunter patients present with joint stiffness and contractures. These skeletal abnormalities are comparable in most patients regardless of the severity of neurological phenotype but are not specific to MPS II. Known as dysostosis multiplex, these radiographic findings are found in all MPS disorders and manifest as a generalized thickening of most long bones. The hands may take on a claw-like appearance and, in combination with carpal tunnel syndrome, loss of hand function can result. These skeletal changes result in profound loss of joint range of motion and restricted mobility and are one of the earliest diagnostic clues. Patients with Hunter Syndrome often walk on their toes and present with hip dysplasia which can become a significant disability.

Progressive airway obstruction is a common finding in Hunter Syndrome and complications are a common cause of death. Narrowed and abnormally shaped trachea and bronchi, enlarged tongue, hypertrophic adenoids and tonsils, large epiglottis, frequent respiratory infections and thick nasal and tracheal secretions are contributing factors. Airway involvement is progressive, first becoming apparent in the upper airways, and gradually involving the lower airways. The progression of airway obstruction usually results in the need of tonsillectomy and adenoidectomy, permanent tracheostomy, continuous positive airway pressure and/or ventilation.

Common oral findings in MPS II patients include macroglossia, hypertrophic adenoids and tonsils, and ankylosis of the temporomandibular joint, which limits opening of the mouth. These changes may be responsible for progressive swallowing impairment. GAG deposition in the larynx typically results in a characteristic hoarse voice. Teeth are often irregularly shaped and gingival tissue is hyperplastic and hypertrophic. Most patients have recurrent ear infections, and nearly all of them experience progressive hearing loss, caused by both conductive and sensorineural deficit.

Cardiac disease is present in almost all patients with Hunter Syndrome and is a major cause of death. Its first signs often present earlier than 5 years of age. The 82% of subjects have cardiovascular signs/symptoms, 62% have a murmur, 57% present valvular disease including, in order of frequency, the mitral, aortic, tricuspid, and pulmonary valves. Valve involvement commonly leads to mitral and aortic regurgitation and/or stenosis. Cardiomyopathy is much less common but may be associated with an increased risk of cardiac arrhythmia. Valve replacement surgery may be necessary.

Because of GAG storage, the liver and spleen of patients with Hunter Syndrome are often enlarged, resulting in abdominal distention. Inguinal hernias are reported in 60% of male patients and also umbilical hernia is commonly observed. Patients with neurologic involvement, often present with chronic diarrhoea but this is not common in the mildly affected patients.

As for the neurological involvement of MPS II, typically the first clue is the delay in global developmental milestones such as the ability to sit, to walk and to speak. In severely affected children, nervous system impairment is profound and progressive; after 48 months of age, cognitive, adaptive, speech, and language functions reach a plateau followed by rapid decline (Holt, Poe & Escolar 2011). Also seizures are reported in more than half of the patients who reach 10 years of age. The decline of cognitive function, combined with progression of severe pulmonary and cardiac disease, generally signals the terminal phase of the disease, with death in the first or second decade of life.

At the opposite end of the phenotype spectrum are patients with minimal-to-no neurological involvement. Although these patients experience all the somatic complications of Hunter Syndrome, they continue to have normal cognitive function and language development is similar to that expected for children with general hearing deficits. Seizures are uncommonly reported in these subjects. They typically survive into adulthood, although death may occur in

the late-teenage years or early adulthood, with airway obstruction and cardiac failure as factors contributing to death.

Nearly all patients older than 72 months have significant reductions in both their gross and fine motor abilities. Other neurological findings may be chronic communicating hydrocephalus, spinal cord compression, caused by narrowing of the spinal canal and instability of the atlantoaxial joint, and carpal tunnel syndrome. Behavioural disturbances, such as hyperactivity, obstinacy, and aggression, occur in both the severe and attenuated forms of the disease (Wraith et al. 2008, Young, Harper 1981) but are more common in the severe form.

MPS II patients also present abnormalities in brain and spine magnetic resonance imaging (MRI) which are summarized in Table 1.5. Among them white matter abnormalities, IIIrd-ventricle dilation and hyperostosis correlate with the severe phenotype (Manara et al. 2011).

Table 1.5 Neuroradiological abnormalities in MPS II patients.

Neuroradiological findings	**%**
Virchow-Robin Perivascular spaces enlargement	89 %
White matter abnormality	97 %
Subarachnoid space enlargement	83 %
IIIrd-ventricle dilatation	100 %
Pituitary sella abnormalities	80 %
Cranial hyperostosis	19 %
Craniosynostosis	19 %
Enlarged cisterna magna	39 %
Dens hypoplasia	66 %
Periodontoid thickening	94 %
Spinal stenosis	46 %
Platyspondylia	84 %
Disc abnormalities	79 %

Also eye is compromised in MPS II but, in contrast to other MPSs, such as MPS I, corneal clouding occurs occasionally. However discrete corneal lesions that do not affect vision may be detected. Also retinal dysfunction, bilateral pigmentary changes and visual field loss have been observed in some patients.

1.1.5 History

The first description of an MPS disorder was made in 1917 by Charles Hunter (Hunter 1917), from whom the disease takes its name, as result of the observation of two brothers with a rare disease. Two years later, Gertrud Hurler described two unrelated girls with features similar to

those described by Hunter (Martin et al. 2008). These two diseases were considered one and were often referred to as Hurler-Hunter syndrome, until the elucidation of their biochemical basis, occurred more than 50 years later (Lorincz 1978). The term mucopolysaccharidosis was used for the first time in 1952 (Brante 1952) after the identification of GAG storages in the liver and, five years later, in urine of patients with Hurler syndrome (Dorfman, Lorincz 1957). Although these storages were initially thought to be caused by a defect of production (excessive synthesis or decreased cellular secretion), in 1968 Fratantoni and colleagues showed that the accumulation of GAGs in skin fibroblasts from Hunter and Hurler patients was a result of reduced degradation (Fratantoni, Hall & Neufeld 1968). They also showed that the metabolic defect in fibroblasts from Hunter patients could be corrected by a factor secreted by fibroblasts from Hurler patients and vice versa. In 1972 the "Hunter-corrective factor" was isolated from normal human urine and shown to be a protein (Cantz et al. 1972), then identified as iduronate 2-sulfatase (Bach et al. 1973). The amino acid sequence of IDS was deduced from its complementary DNA clone in 1990 (Wilson et al. 1990).

1.1.6 Diagnosis

Not all signs and symptoms described are found in each patient and they are not specific for MPS II; therefore long time may elapse between the presentation of first signs and the ultimate diagnosis.

To confirm a clinical suspicion of MPS, analysis of urinary GAG level can be performed. However it is not diagnostic for Hunter Syndrome because an excess of dermatan and heparan sulfate is a common feature of MPS I, II and VII. Thus further tests are required. A negative urine GAG test cannot exclude MPS II diagnosis because of the low sensitivity of test method.

IDS enzyme activity is usually measured in cultured fibroblasts, leukocyte, plasma or serum. Methods based on dried blood spot are preferred for screening purposes and for areas where the transport of fresh samples is not possible. In males, absent or low IDS activity is diagnostic for Hunter Syndrome if the normal activity of another sulfatase is proved ruling out multiple sulfatase deficiency. Conversely enzyme activity cannot be used to identify female carriers, where identification of IDS gene mutation, usually previously identified in an affected male relative, is needed. Neither phenotype severity can be predicted with this method.

Mutation analysis of IDS gene may be used to confirm Hunter Syndrome in males and is the only secure way to identify female carriers. It could also be used for prenatal diagnosis in male foetuses if the family specific mutation is known.

Enzyme analysis and genetic testing are needed to confirm the diagnosis and to discriminate between different MPSs. Other storage disorders present a phenotype similar to MPSs, including mucolipidosis I, II, and III, mannosidosis, fucosidosis, and multiple sulfatase deficiency. Other conditions characterized by macrocephaly and/or organomegaly that are coupled with developmental delay may also be confused with MPSs.

1.1.7 Management and treatment

Until recently, management of Hunter Syndrome has been palliative and focused on the treatment of signs and symptoms. The involvement of specialists for each affected organ system is required to monitor and treat specific problems.

1.1.7.1 Hematopoietic stem cell transplantation

Hematopoietic stem cell transplantation (HSTC) via umbilical cord blood or bone marrow has been proposed as a way of providing sufficient enzyme to delay or stop the progression of the disease (Martin et al. 2008). The potentially beneficial effect of HSCT is thought to be due to the replacement of deficient macrophages with marrow-derived donor macrophages which constitute an ongoing source of enzyme capable of gaining access to various storage sites (Wraith et al. 2008). HSCT has been successful in patients with severe MPS I and MPS VI in early phase of disease progression. In MPS II the use of HSCT has been controversial because of the associated high risk of morbidity and mortality, the lack of sufficient data on long-term follow up and some negative results obtained in few case reports (Vellodi 2005, Araya et al. 2009). However the results of a long-term retrospective study performed in Japan in a cohort of 21 HSCT-treated patients have been recently published (Tanaka et al. 2012). HSCT showed effectiveness when performed before signs of brain atrophy or valvular regurgitation appear. In these patients improvements of brain MRI, cardiac valve regurgitation and urinary GAG levels were observed together with a stabilization of brain atrophy, speech capability and activities of daily living.

1.1.7.2 Enzyme Replacement Therapy

Idursulfase (Elaprase®, Shire Human Genetic Therapies, Inc, Cambridge, MA) is a recombinant human IDS produced in a human cell line that was licensed for treatment of MPS II patients in the USA and the European Union, gaining marketing approval by the Food and Drug Administration (FDA) in 2006 and by the European Medicines Agency (EMA) in 2007. In Italy it was approved by the Italian Medicines Agency (AIFA) in 2006.

The clinical benefits of idursulfase were demonstrated in a randomized, placebo-controlled, double-blind clinical trial (Muenzer et al. 2006), in which 96 patients were randomly assigned to placebo or to a 0.5 mg/kg idursulfase dosage that was infused either weekly or every other week (EOW). After 1 year of treatment, a statistically significant improvement rate of the primary end point, compared with placebo, resulted both in patients in which the idursulfase was infused weekly and in the EOW group, even if in the weekly group it was more significant. A significant reduction of GAG excretion and liver and spleen volumes was observed. On the basis of the larger clinical response in the weekly group compared with the EOW group, idursulfase was approved for the treatment of MPS II in both the United States and European Union at a dose of 0.5 mg/kg weekly.

1.1.7.3 New therapeutic options

Several new approaches to the treatment of MPS disorders are under investigation; some of these could be stand-alone therapies, others would likely be used as adjuvant therapy. Among these, both viral and non-viral approaches have been tested for the treatment of Hunter syndrome. Even if only in preclinical animal experiments, the obtained results have been encouraging. Gene therapy mediated by adeno-associated virus (Cardone et al. 2006, Jung et al. 2010) gave good results as well as cell therapy employing intraperitoneally implanted microcapsules (Friso et al. 2005). Also muscle electro gene transfer has been tested, obtaining an elevated production of therapeutic protein but little curative effects in organs far from the injection site (Friso et al. 2008).

A therapeutic approach of substrate reduction therapy for Hunter syndrome uses genistein, an isoflavone extract from soy beans that was shown to inhibit GAG synthesis in MPS cultured fibroblasts. It could reveal to be useful in combined therapeutic protocols and has shown efficacy in MPS II mice (Friso et al 2010). Being a diet supplement, genistein can be assumed with no prescription, therefore in the last few years numerous MPS patients have regularly

taken it, despite no important clinical trials evaluating both efficacy and safety have been carried on.

1.1.8 Pathogenesis

MPS II has been described and well characterized from the clinical point of view for almost a century. Also, although for less time, its enzymatic, genetic and molecular bases have been determined. Nevertheless little is known about the downstream or secondary biochemical and cellular pathways in the disease pathogenesis.

It must be noted that, despite the distinctive types of storage material, LSDs share many common biochemical, cellular and clinical features. Therefore advances in understanding other specific LSDs could also provide insights into MPS II pathology and vice-versa.

In LSDs alterations occur in many cellular processes, such as signalling pathways, intracellular calcium homoeostasis, oxidative stress response, autophagy, endoplasmic reticulum stress, the unfolded protein response, metabolism, trafficking and inflammation (Ballabio, Gieselmann, Cox, Cachon-Gonzalez 2012).

1.1.8.1 Secondary storages

In many LSDs, there is accumulation of secondary metabolites which cannot be explained by the underlining enzymatic defect and that are thought to be involved in disease pathogenesis.

The most common secondary metabolites include glycosphingolipids (GSLs), mainly the gangliosides GM2 and GM3, cholesterol and phospholipids. In Niemann-Pick Type C (NPC) disease, the accumulation of GM2 and GM3 may be the result of a trafficking defect that is downstream to defects in lysosomal calcium homeostasis (Vitner, Platt & Futerman 2010). Moreover, intraneuronal accumulation of GM2 correlates with ectopic dendritogenesis, a phenomenon unique to lysosomal disease. Cholesterol seems to interfere with normal endosomal/lysosomal processing as well as lead to cholesterol deficits elsewhere in neurons, with consequences for neuronal function (Walkley, Vanier 2009).

In MPS II, GM2 and GM3 gagliosides and the subunit c of ATP synthase have been found as secondary storage materials in brain tissue (Constantopoulos, Iqbal & Dekaban, Elleder, Sokolová & Hrebícek, 1997).

1.1.8.2 Alteration of signalling pathways

Compound accumulated in LSD can affect signal transduction pathways at different levels. They can function as ligands of receptors (Krabbe disease, MPSs), modify receptor response (MPS I, NPC disease), alter subcellular localization of receptors and alter activities of enzymes involved in signal transduction (Ballabio, Gieselmann 2009).

1.1.8.2.1 Non-physiological activation of signal transduction receptors

An example of storage compound which acts as ligand for signal transduction receptors is the lysolipid psychosine in the mouse model of Krabbe disease; the activation of the G protein-coupled receptor TDAG8 by the psychosyne induces the inhibition of cytokinesis providing a possible explanation for the formation of large multinuclear cells observed in Krabbe disease (Kanazawa et al. 2000).

Also in MPSs, storage compound induce a non physiologic activation of a signal transduction receptor, in this case Toll like receptor 4 (TLR4). In fact in MPSs glycosaminoglycan (GAG) storage is not restricted to lysosomes but also occurs in the extracellular matrix (ECM), where they should be physiologically localized as components of ECM proteoglycans, and where GAG fragments with altered sulphatation accumulate. These fragments structurally resemble the bacterial endotoxin lypopolysaccharide (LPS), a ligand and activator of TLR4, whose binding leads to secretion of proinflammatory cytokines and immune response. Several genes involved in TLR signalling, including TLR4 itself, LPS binding protein and MyD88, are upregulated in synovial cells from MPS VI cats and rats and MPS VII dogs (Simonaro et al. 2008). Accordingly, chondrocytes display higher nitric oxide (NO) levels, secrete enhanced amount of proinflammatory cytokines such as IL-1β, TNF-α and TGF-β. Cytokines and NO induce the expression of matrix metalloproteases which through their proteolytic activity may directly contribute to cartilage degeneration (Vitner, Platt & Futerman 2010).

1.1.8.2.2 Modification of signal transduction receptor response

MPSs also provide an example of how a storage compound can alter signal transduction receptor response. Indeed, besides their structural role, GAGs (particularly heparan sulfate) together with the proteoglycans they form at the cell surface and within the extracellular matrix, are essential for cell signalling and distribution of growth factors, cytokines and morphogens (Bishop, Schuksz & Esko 2007). Heparan sulfate (HS) molecules influence ligand-

receptor interaction by stabilizing receptors at the cell surface and by direct binding and presentation of ligands to their respective receptor. Free HS within the extracellular matrix can also sequester factors, thus modulating their biological effects. The growth factor/morphogen binding is determined by the degree and pattern of post-translational modifications of HS, particularly O-sulfation (Campos, Monaga 2012). Fibroblast growth factor 2 (FGF-2) is one of the growth factors which need HS proteoglycan for optimal binding to its receptor (Gallagher 2006). Binding of FGF-2 to this receptor is reduced in cells derived from Hurler syndrome patients and consequently also the proliferative and survival response of these cells to FGF-2. Both the mitogenic and survival-promoting activities of FGF-2 were restored by substitution of Hurler HS by normal HS (Pan et al. 2005). This may explain the increased rate of apoptosis seen in cells of Hurler patients and probably the neurodegeneration, considering its neuroprotective effect on neurons and neuronal precursors cells (Alzheimer, Werner 2002). In the mouse model of MPS IIIB and a dog model of MPS VII (Walton, Wolfe 2007) proliferation of neural progenitor cells, which depends on FGF-2 (Gritti et al. 1999) was reduced too.

Also bone morphogenetic proteins (BMPs), which belong to the TGF-β growth factor superfamily, are regulated by HS proteoglycans. BMPs control proliferation, differentiation and apoptosis in several tissues, including the nervous and skeletal systems and are important for bone and cartilage development (Chen, Zhao & Mundy 2004). It is believed that specific sulfated residues in the GAG side chains bind to BMPs and their antagonists and thereby modulate receptor-mediated signalling, diffusion, and localization of these molecules (Takada et al. 2003, Kuo, Digman & Lander 2010). Comparable to FGF-2, abnormally accumulated GAGs in Hurler cells are shown to impair BMP-4 signalling, which can be restored by enzymatic removal of these GAGs (Khan et al. 2008). Therefore, the alteration in this pathway probably contributes to the common joint and skeletal abnormalities as well as neurological abnormalities seen in patients with MPS I and other MPSs.

Another example of alteration of signal transduction receptor response in LSDs comes from NPC disease in which the stored compounds, cholesterol and sphingolipids, are thought to cause impairment of insulin signalling in murine hepatocytes, by altering the formation of lipid rafts (Vainio et al. 2005). In addition, reduction of cholesterol in the plasma membranes isolated from NPC hepatocytes improved insulin receptor signalling demonstrating that plasma membrane lipid composition has a direct influence on signal transduction (Vainio et al. 2005). So far, there are no indications of insulin resistance in NPC patients or the respective mouse model (Ballabio, Gieselmann 2009).

Alterations of insulin pathway have also been reported in Gaucher disease type I, in which the accumulation of the secondary metabolite GM3 ganglioside, known to modify insulin response (Tagami et al. 2002), is thought to be responsible of the insulin resistance found in Gaucher type I patients (Ballabio, Gieselmann 2009). Experimental evidence, however, for this hypothesis is still missing.

1.1.8.2.3 Modification of activities of the enzymes involved in signal transduction

Some storage molecules directly affect components of signal transduction pathways downstream of receptor activation. For example, in Krabbe disease psychosine accumulation has been associated to the inhibition of protein kinase C and the activation of phospholipase A2, both resulting in apoptosis induction, and to the inhibition of the AMP activated protein kinase with possible influence on cell energy status (Ballabio, Gieselmann 2009).

1.1.8.3 Alterations of intracellular calcium homeostasis

Defective calcium signalling has emerged as a key common pathway in many LSDs. However, the mechanisms leading to this alteration are different in each LSDs, according to the different interaction between stored substrates and specific calcium transporters or other upstream events. Alteration in calcium transporters located in all cellular membranes have been observed in LSDs .

The activity of endoplasmic reticulum (ER) Ca^{2+} channels have been found altered in Gaucher disease, Sandhoff disease, Niemann -Pick A disease and GM1 gangliosidosis (Vitner, Platt & Futerman 2010). In a model of neuronal form of Gaucher disease, an increase of Ca^{2+} release from ER occurs in response to glutamate stimulus, due to overactivation of the ER Ca^{2+} channel, the ryanodine receptor (Korkotian et al. 1999). This is due to the direct regulation of the channel activity by the primary stored metabolite, glucosylceramide (GlcCer) (Lloyd-Evans et al. 2003). This results in an enhanced glutamate toxicity leading to an increased rate of apoptosis which may explain partly the neurodegeneration occurring in the severe type 2 forms of Gaucher disease. This alteration of calcium release from ER was also demonstrated in microsomes isolated from autoptic brain of Gaucher patients with a correlation between the degree of agonist induced release of calcium and disease severity. In Sandhoff disease (a GM2 gangliosidosis) (Pelled et al. 2003) and in Niemann-Pick A disease (Ginzburg, Futerman 2005), a reduction of Ca^{2+} uptake by the sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) causes elevated cytosolic calcium levels; in the case of Sandhoff disease, ganglioside GM2 modulates

SERCA activity *in vitro* (Pelled et al. 2003) by means of an exposed sialic acid residue. ER calcium store depletion also occurs in the GM1 gangliosidosis (Tessitore et al. 2008), where GM1 interacts with the phosphorylated form of the inositol 1,4,5-trisphosphate-gated Ca^{2+} release channel. In Niemann-Pick A disease the reduction of SERCA activity was caused by a severe reduction of SERCA expression rather than activity inhibition (Ginzburg, Futerman 2005). Mitochondrial calcium homeostasis is altered in at least two LSDs. In GM1 gangliosidosis, GM1 accumulation results in elevated mitochondrial membrane permeabilization, opening of the permeability transition pore, and activation of the mitochondrial apoptotic pathway (Sano et al. 2009). In skin fibroblasts from mucolipidosis type IV patients, an oversensitivity of the cells to proapoptotic effects induced by calcium-mobilizing agonists was observed, probably as response to decreased mitochondrial calcium buffering and mitochondrial fragmentation (Jennings et al. 2006); however, calcium-dependent apoptosis appeared to depend on activation of a non-mitochondrial caspase 8 pathway (Vitner, Platt & Futerman 2010).

Lysosomal calcium is altered in NPC1 disease. Sphingosine storage in NPC1 cells induces a reduction in lysosomal calcium stores, resulting in defective endocytic fusion and trafficking, which in turn induces cholesterol, sphingomyelin, and glycosphingolipid storage. Increasing cytosolic calcium, as compensation for altered lysosomal calcium levels, reverses the NPC1 phenotype and prolongs survival of NPC1 mice (Lloyd-Evans et al. 2008). Finally, calcium homeostasis is altered by modulation of a plasma membrane Ca^{2+} channel in a mouse model of the juvenile form of neuronal ceroid lipofuscinosis (NCL). Gnb1 protein, a β1 subunit of the G protein complex which negatively regulates N-type voltage- gated Ca^{2+} channels, is found at elevated levels. Upon inhibition of N-type calcium channels, recovery from depolarization was lower in Cln3-deficient neurons, resulting in a prolonged period of higher intracellular calcium (Luiro et al. 2006).

Calcium plays vital roles in regulating a variety of cellular events, with impaired calcium homeostasis leading to endoplasmic reticulum (ER) stress, oxidative stress and cell death.

1.1.8.4 Oxidative stress

Oxidative stress is thought to be a common mediator of apoptosis in several LSDs including all forms of NCLs, Tay-Sachs disease, Gaucher disease type I and II, NPC1 and 2 and nephropathic cystinosis (Zampieri et al. 2009, Wei et al. 2008).

In Gaucher disease, an up-regulation of apurinic endonuclease 1 (APE1) (a protein that repairs oxidative DNA damage) was observed in fibroblasts (Deganuto et al. 2007) but not in bone marrow mesenchymal stromal cells (Campeau et al. 2009). Inducible nitric oxide synthase and nitrotyrosine were elevated in activated microglia/macrophages of GM1 and GM2 gangliosidoses (Jeyakumar et al. 2003a), and ROS was elevated in Fabry disease models (Shen et al. 2008) and in NPC1 fibroblasts. In the last case also gene microarray analysis confirms oxidative stress hypothesis (Reddy, Ganley & Pfeffer 2006). In MPS IIIB, enhanced oxidative stress resulted in protein, lipid, and DNA oxidation (Villani et al. 2009), and an oxidative imbalance was found in MPS I (Reolon et al. 2009). In fibroblasts and brain extracts from CLN6 sheep, a significant increase in manganese-dependent superoxide dismutase activity was detected (Heine et al. 2003), and the expression of 4-hydroxynonenal was found increased in late infant and juvenile forms of NCL (Hachiya et al. 2006). However, in NCLs, elevated ROS and superoxide dismutase levels were suggested to be downstream to ER stress (Kim et al. 2006). The central role that oxidative stress plays in integrating other cellular pathways and stresses suggests that it is most likely activated in LSDs as a secondary biochemical pathway, rather than as a direct result of accumulation of the primary substrate (Vitner, Platt & Futerman 2010).

1.1.8.5 Endoplasmic reticulum stress and the unfolded protein response

In the ER, folding and post-translational modifications of proteins directed to the secretory pathway take place. If these functions are impaired, misfolded proteins accumulate in the ER lumen and activate the unfolded protein response (UPR), which can initiate apoptosis. Unfolded protein accumulation can occur in response to changes in the ER environment, including glucose starvation, reducing agents, and depletion of ER calcium stores (Vitner, Platt & Futerman 2010). In particular alterations of calcium homeostasis induce URP because some chaperones, which guide proper folding of proteins inside the ER, such as calnexin or calreticulin, critically depend on calcium (Ballabio, Gieselmann 2009). Because calcium homeostasis is altered in LSDs, this pathway could also be potentially involved in LSD pathology.

UPR activation was observed in fibroblasts from many LSDs (including all forms of NCLs, Tay-Sachs disease, Gaucher disease type I and II, NPC2 and nephropathic cystinosis) (Wei et al. 2008) but also in mouse tissues and in neurospheres obtained from GM1 gangliosidosis mice (Tessitore et al. 2004) and in the brain and in cultured cells from infantile NCL (Kim et al. 2006).

However, no UPR activation was observed in spinal cord tissue from a mouse model of sialidosis (Tessitore et al. 2004), nor in various neuronal Gaucher mouse models (Farfel-Becker et al. 2009).

1.1.8.6 Autophagy

The lysosome has a pivotal role in autophagy, an important degradation pathway which mediates the cellular turnover of proteins and organelles. During autophagy a large portion of cytosol, containing the material to be degraded, is sequestered in specific vesicles (called autophagosomes, Avs) and then degraded upon fusion with lysosomes (Mizushima, Ohsumi & Yoshimori 2002). Cells use basal autophagy also to regulate turnover of organelles, such as mitochondria, peroxisomes and endoplasmic reticulum. In addition to this ordinary activity, autophagy can be induced as response to stress event, such as nutrient depletion (when autophagy allows generation of ATP from catabolism of macromolecules) and oxidative stress (when autophagy allows the removal of damaged organelles and proteins from the cytosol), thus acting as pro survival pathway (Ravikumar et al. 2006).

However, the induction of autophagy also leads to abnormal overproduction of autophagosomes in the cells which has been associated to cell death. It remains still to understand whether this represents an attempt to survive, or is itself the mechanism by which the cell dies (autophagic or type 2 cell death). Probably, when autophagic activity overcomes a given threshold, it results in massive degradation of organelles and molecules and in cell death (Maiuri et al. 2007).

Autophagy has been analyzed in a variety of LSDs with different storage molecules, tissues involved and phenotypic severities. Table 1.6 shows the main findings obtained (Lieberman et al. 2012). These studies firmly establish the presence of autophagosome accumulation in lysosomal storage diseases, notwithstanding differences among the diseases and samples analyzed. However, in some cases there seems to be an impairment of autophagic flux, due to defective autophagosomal maturation or autophagosome-lysosome fusion, which cause a secondary accumulation of autophagy substrates such as polyubiquitinated proteins, the autophagic substate p62/SQSTM1 and dysfunctional mitochondria. In other cases, signals of an activation of autophagosome formation, such as high levels of beclin-1 (BECN1) and downregulation of mTOR pathway have been observed. Both processes, the block and the activation of autophagy, seem even to coexist. To solve this discrepancy an hypothesis was proposed according to which lysosomal storage leads to a partial block of autophagy and this

in turn activates a compensatory feedback mechanism through which autophagy is induced (Ballabio, Gieselmann 2009, Lieberman et al. 2012).

Table 1.6. The main results of the study on autophagy in LSDs. AV=autophagic vesicles. Modified from (Lieberman et al. 2012).

Disease	AV accumulation	Defective AV formation	Increased AV formation	Increased polyubiquitinated proteins	Increased dysfunctional mitochondria	Increased p62
GLYCOGENOSES						
Pompe disease	+	+	+	+	NA	+
Danon disease	+	+	NA	NA	NA	NA
MUCOPOLYSACCHARIDOSES						
MSD	+	+	-	+	+	+
MPS III A	+	+	-	+	+	+
MPS VI	+	+	NA	+	+	+
SPHINGOLIPIDOSES						
NPC1, NPC2	+	+	+	+	+	+
Gaucher disease	+	NA	NA	NA	NA	+
Fabry disease	+	+	NA	+	NA	+
GM1 gangliosidosis	+	NA	+	NA	+	NA
MUCOLIPIDOSES						
MLII	+	NA	-	+	+	+
MLIII	+	NA	-	+	+	+
MLIV	+	+	+	+	+	+
CEROID LIPOFUSCINOSES						
CLN10	+	NA	NA	NA	NA	NA
CLN 3	+	NA	+	NA	NA	NA

1.1.8.7 Alteration of trafficking

Since the endosomal and lysosomal pathway are functionally connected it is not surprising that lysosomal storage affects intracellular sorting events. Nonetheless there is only little information about this topic mainly related to lipid and receptor trafficking in lipidoses.

In LSDs, lipid trafficking was studied using a fluorescent analogue of lactosylceramide (BODIPY-LacCer) which, if added from outside, accumulates in the Golgi apparatus in normal fibroblasts and in late endosomes and lysosomes in fibroblasts from some LSD patients. This occurs for GM1 gangliosidosis, GM2 gangliosidosis, prosaposin deficiency, metachromatic leukodystrophy, MPS IV, Fabry disease, Niemann-Pick disease (types A, B, and C) and Krabbe disease while no abnormalities were observed in Pompe disease, Hunter disease, NCL, and

Farber disease (Marks, Pagano 2002, Sillence, Platt 2004). The alteration of lipid is probably related to changes in cholesterol levels because depletion of cholesterol in LSDs restored normal BODIPY-LacCer trafficking. The ras-associated binding protein rab7 and rab9 seem to be involved in this vesicle transport in NPC cells.

Regarding receptor trafficking, reduced endocytosis of the mannose-6-phosphate receptor, but not of the mannose receptor, was observed in cells from Niemann–Pick disease type A patients (Dhami, Schuchman 2004). A similar effect was also observed in metachromatic leukodystrophy and NPC. In the former the internalisation rate of mannose-6-phosphate receptor and the transferrin receptor is enhanced and their recycling rate is reduced ;(Klein et al. 2009) in the second altered localization of the mannose-6-phosphate receptor towards the late endosomal compartment was observed (Kobayashi et al. 1999).

In NPC also TLR4 receptor, responsible for the production of inflammatory cytokines (IFN-β, IL6 and IL8) was found accumulated in endosomes. This probably reduces the lysosomal degradation of TRL4 receptor thus resulting in more intense signalling (Suzuki et al. 2007).

1.1.8.8 Alterations in lipid biosynthesis

Since in many LSDs the primary or secondary stored metabolites are lipids, it is not surprising to find alteration of lipid metabolism in these diseases.

In Sandhoff disease phospholipids were found decreased in murine neurons as was phosphatidycholine in mouse brain, while no alteration was found in liver and spleen. Also activities of two key enzymes in the synthesis of phosphatidylcholine and phosphatidylserine (CTP:phosphocholine cytidylyl-transferase and phosphatidylserine synthase, respectively) are decreased in the brain of Sandhoff mice, compared to wild type. It is thought that regulation of activity of these enzymes occurs post translationally. Since phospholipids are major components of cellular membranes their reduced synthesis may explain the impaired neurite growth of neurons derived from Sandhoff mice (Ballabio, Gieselmann 2009).

In contrast, in Gaucher mouse model the levels of various phospholipids are elevated. This is likely caused by a direct activation, by the accumulated glucosylceramide, of the rate limiting enzyme for the synthesis of phosphatidylcholine (CTP:phosphocholine cytidylyl-transferase).

As a consequence, storing neurons from Gaucher disease mouse model show longer neurites and more axonal branch points than control mice.

Glucosylceramide stimulates phospholipid synthesis not only in storing neurons, but also in macrophages allowing the enlargement of macrophages observed in the disease (Ballabio, Gieselmann 2009).

1.1.8.9 Abnormalities in iron metabolism

In some LSDs alteration of iron metabolism are observed, which could be related to the role of lysosome in the retrieval of storage iron from ferritin and haemosiderin, a proteolytic and polymerized ferritin iron–protein complex.

In GM1 and GM2 gangliosidosis mice a progressive depletion of iron is found in brain tissue. Also key regulators of systemic iron homeostasis, hepatic peptide hormone hepcidin and the inflammatory cytokine IL-6, were increased in the serum when compared with age-matched control animals. Transferrin concentrations were reduced, reflecting a progressive inability of the brain to acquire iron from the circulation, whereas the transferrin receptor and the iron exporter ferroportin were up-regulated in the brain. Administration of iron prolonged survival in the mutant mice by about one-third and also delayed the onset of disease (Cox, Cachon-Gonzalez 2012).

1.1.8.10 Inflammation

Several LSDs with neurologic involvement present activation of inflammatory processes in the brain (Castaneda et al. 2008). In the brain, macrophage lineage cells are represented by microglia, which respond to trauma and disease by activating the inflammatory response. In LSDs with CNS involvement a inflammatory response due to the large and widespread infiltration of microglia is common (Ransohoff, Perry 2009).

However, also in LSDs without CNS pathology, as Gaucher disease type I, a macrophage activation with consequent release of cytokines is observed in several organs (liver, spleen, and bone marrow) (Jmoudiak, Futerman 2005).

The exact mechanisms leading to immune activation are unknown but probably lysosomal storage causes multiple proinflammatory cytokines activation by different molecular mechanisms depending on the biochemical nature of the stored molecules. In fact, it is known that GAGs have proinflammatory properties related to the stabilization and release of cytokines and their oligomerization (Mulloy, Rider 2006).

Despite the fact that inflammation is a secondary event in the LSDs diseases, as GM1 and GM2 gangliosidoses (Jeyakumar et al. 2003b), Sandhoff and NPC1 disease (Jeyakumar et al. 2004, Smith et al. 2009), it may be a target for supportive therapy in these pathologies.

1.2 RNA-Sequencing

The transcriptome is the complete set of transcripts for certain type of cells or tissues in a specific developmental stage or physiological condition. Transcriptome analysis can provide a comprehensive understanding of molecular mechanisms involved in specific biological processes and diseases from the information on gene structure and function. Understanding the transcriptome is essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues, and also for understanding development and disease. Analysis of gene expression alteration in disease condition allows to evaluate the onset and progression of a disease and to identify markers of pathology and possibly of therapeutic efficacy. Initial transcriptomics studies largely relied on hybridization-based microarray technologies and until few years ago, microarrays were the only method capable of providing a high-throughput transcriptomic data. Thanks to the relatively low cost and high efficiency, they are still used worldwide in transcriptomic approaches. However, despite the ability to simultaneously measure the expression levels of hundreds of genes, microarrays have some limits It does not identify unknown transcripts; furthermore it is an indirect method in which the amount of transcript is deduced from the intensity of hybridization, obtaining results with an high background, and this makes the comparison between experiments difficult, requiring complicated standardization systems.

These limitations have led to the development of a new technique called RNA Sequencing (RNA-Seq). RNA-Seq uses next-generation sequencing technologies to allow transcriptome analyses of genomes at a far higher resolution than is available with traditional Sanger sequencing- and microarray-based methods. In the RNA-Seq method, complementary DNAs (cDNAs) generated from the RNA of interest are directly sequenced using next-generation sequencing technologies. The reads obtained from this can then be aligned to a reference genome in order to construct a whole-genome transcriptome map.

RNA-seq studies have been providing a progressively fuller knowledge of both the quantitative and qualitative aspects of transcript biology in both prokaryotes and eukaryotes; it has been used successfully to precisely quantify transcript levels, confirm or revise previously annotated 5' and 3' ends of genes, and map exon/intron boundaries. Moreover this approach allows a more comprehensive understanding of transcription initiation sites, to catalogue sense and antisense transcripts, to improve the detection of alternative splicing events and the detection of gene fusion transcripts.

The next-generation sequencers have permitted to enormously reduce the costs of sequencing, experimental complexity and time of analysis as well as to increase transcriptome coverage.

Among high-throughput next generation sequencing (HT-NGS) technologies, a second generation and a third generation can be distinguished.

In the second generation HT-NGS platforms, the principle was based on an emulsion PCR amplification of DNA fragments, to make the signal strong enough for reliable base detection. In some instances PCR amplification may introduce base sequence errors or favor of certain sequences over others, thus changing the relative frequency and abundance of various DNA fragments that existed before amplification.

To overcome this, the ultimate miniaturization into the nanoscale and the minimal use of biochemicals, would be achievable if the sequence could be determined directly from a single DNA molecule, without the need for PCR amplification and its potential for distortion of abundance levels. This sequencing from a single DNA molecule is now called as the third generation of HT-NGS technology.

The characteristics of the main sequencing systems are shown in Figure 1.4 and Figure 1.5.

Figure 1.4 Advanced technological features of three leading second generation NGS platforms

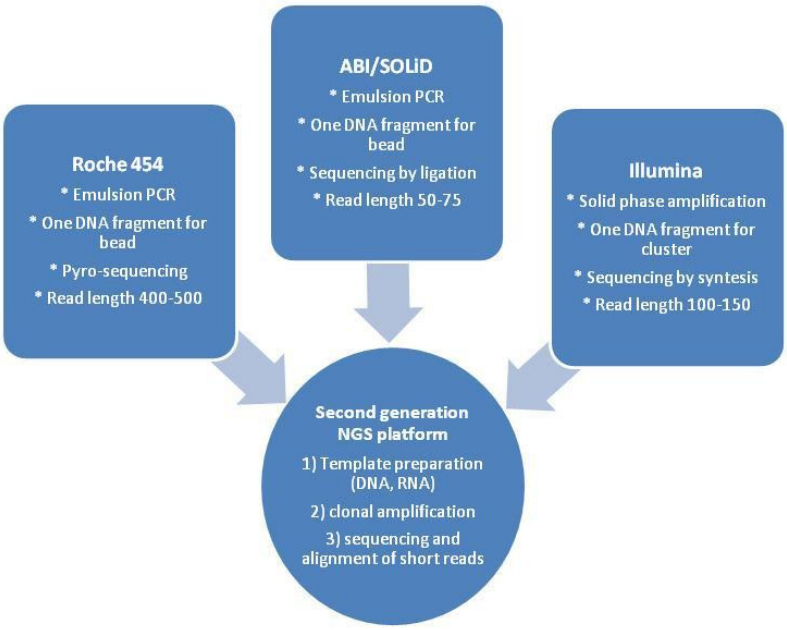
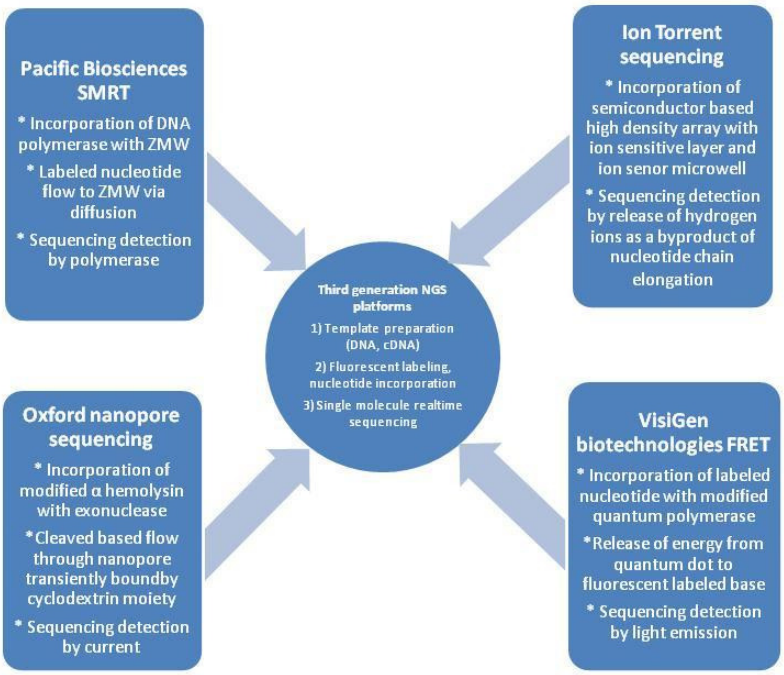


Figure 1.5 Advanced technological features of four leading third generation NGS platforms (SMRT= Single molecule real time sequencing; ZMW= zero-mode waveguide).



2 AIMS

Hunter Syndrome (Mucopolysaccharidosis type II, MPS II) is a rare genetic disease belonging to the group of pathologies called Mucopolysaccharidoses (MPSs), and is due to the lack of activity of the lysosomal enzyme iduronate 2-sulfatase (IDS), involved in the degradation of the mucopolysaccharides heparan- and dermatan-sulphate. As a result such molecules accumulate within cell lysosomes and in the extracellular matrix of most tissues and organs, thus determining a severe pathological phenotype. Due to the housekeeping nature of IDS, enzymatic deficit affects most if not all the peripheral organ systems, also impairing cognitive functions in the severe forms of the disease.

MPSs belong to the wider group of lysosomal storage disorders (LSDs), counting about 50 diverse lysosomal diseases, which, although characterized by different deposits of macromolecules, share several pathological signs and symptoms.

Studies conducted in the past few years to clarify biochemical and molecular pathways involved in lysosomal storage have provided some clues on LSDs and/or more specifically on MPSs or Hunter Syndrome pathogenesis; however, most of the primary molecular alterations causing the disease phenotypes as well as the resulting secondary changes remain unknown. A deep comprehension of these mechanisms would shed light on important cellular functions and would help to identify potential therapeutic targets as well as biomarkers of diagnosis, prognosis and therapeutic efficacy.

At the moment, a part from symptomatic treatments, the most significant therapeutic approach applied to MPS II is represented by the Enzyme Replacement Therapy (ERT), obtained by systemic weekly injection of the purified recombinant enzyme. Although application of ERT has obtained variable and very subjective results in the different patients analyzed, at present same therapeutic dose, time-schedule and modality of administration are adopted for all MPS II patients. Instead, it might be necessary to adjust both doses and timing, based on a personal treatment follow-up, thus favouring the design of a personalized therapeutic approach.

On this basis, the present project has performed for the first time a high throughput analysis of Hunter disease, starting from a total RNA sequencing approach.

Specifically, aims of this project were:

- To study MPS II pathogenesis by comparing the whole transcriptome of primary human cells derived from Hunter and healthy controls.
- To identify molecular biomarkers of therapeutic efficacy in the above cell types undergoing *in vitro* treatment with the recombinant human enzyme.
- To evaluate the candidate biomarkers identified by an *in vivo* follow-up of Hunter patients undergoing ERT.
- To study possible correlations between clinical signs and symptoms and the laboratory biomarkers identified.

This approach will contemporary elucidate molecular pathways involved in the disease pathogenesis/progression and contribute to the search of molecules easily detectable in accessible tissues, helping in assessing ERT efficacy.

3 MATERIALS AND METHODS

3.1 Rationale and experimental workflow

A whole transcriptome analysis on fibroblasts derived from patients affected by Hunter syndrome and treated *in vitro* with therapeutic IDS enzyme was performed with two aims:

1. to study the cellular pathogenesis of MPS II
2. to isolate a list of candidate molecular biomarkers of ERT efficacy.

The transcriptome analysis was performed by RNA-Sequencing (RNA-Seq), a full-length cDNA sequencing application that generates a comprehensive and quantitative view of the entire cell transcriptome.

The list of candidate biomarkers obtained from RNA-Seq data analysis, after a filtering process, was evaluated in a population of Hunter patients undergoing ERT treatment; this evaluation was performed through qPCR in blood RNA samples collected at different time-points from the start of therapy for about 3 years. From the same population, clinical data were collected during the whole observational period and an assessment of therapeutic efficacy was performed considering several clinical parameters. To validate the selected candidate biomarkers, a correlation analysis was performed between clinical data and the blood gene expression profile of each candidate gene.

Therefore for biomarker research on blood samples, the *in vitro* experiment on fibroblasts was used as a preliminary step to obtain, searching among all cellular transcripts by a transcriptomic approach, a list of possible biomarkers of therapeutic efficacy. This step was made necessary also because of the preciousness and uniqueness of blood samples from patients under treatment and because of the influence of different factors, such as sex, age, time of the day, proportion of different cell populations and many physiological and environmental conditions, on blood cells' gene expression (Dumeaux et al. 2010). Study of MPS II pathogenesis and effect of treatment

3.2 Study of MPS II pathogenesis

3.2.1 Cells and treatment

Primary fibroblasts from 5 Hunter patients with 4 different mutations in IDS gene (Table 3.1) and from 4 healthy controls were evaluated. Fibroblasts from Hunter patients were obtained from cutaneous biopsy (Cell line and DNA Biobank from patients affected by Genetic Diseases, G. Gaslini Institute, Genova) while controls fibroblasts derived from circumcision (Department of Histology, Microbiology and Medical Biotechnology - University of Padova). All cells were anonymously obtained and maintained.

Table 3.1 Patient ID of the Human fibroblasts used and related mutations.

PATIENT_ID	MUTATION
20/98	c.1131_1305 with Alu insertion (del ex8)
114/85	c.1403G>A (R468Q)
104/84	c.410_411delTT (F137nt2)
58/86	c.1403G>A (R468Q)
48/03	c.1122_1181del

They were cultured into tissue culture dishes (Ø 10 cm) under 5% CO_2 tension and at 37°C, in RPMI-1640 medium supplemented with 2 mM L-glutamine, 100 U/ml penicillin G, 100 µg/ml streptomycin and 15% (v/v) heat-inactivated fetal bovine serum (all Gibco®, Life Technologies™, Carlsbad, CA, USA). Cells were treated with recombinant IDS enzyme in the same formulation used in ERT (Elaprase, Shire Human Genetic Therapies, Cambridge, MA, USA), as described in Table 3.2; treatment with equal volumes of NaCl 0.9% (Bieffe Medital, Grosotto, SO, Italy) was used as negative control for cells of C and UT samples.

Time points of 24 hours and 144 hours (six days) were chosen with the purpose to register the short- and long-term effect of the therapy on the whole cellular gene expression, considering also the fact that Hunter patients are infused with the recombinant enzyme once a week. 62.5 nM IDS dosage was chosen because, in addition to allowing a good induction of IDS enzyme activity and an important reduction of GAGs, it allowed maintaining treatment at sufficiently low concentrations and similar to those detectable in the plasma of animals in preclinical experiments (data unpublished).

Table 3.2 Pool of fibroblasts analyzed using the SOLiD™ 3 System with relative treatment and treatment duration.

ID_SAMPLE	GENOTYPE	TREATMENT	TREATMENT DURATION (h)	LINES IN POOL
C	Healthy	NaCl 0.9%	24	4
UT	MPS II	NaCl 0.9%	24	5
T1	MPS II	62.5 nM IDS	24	5
T2	MPS II	62.5 nM IDS	144	5

3.2.2 RNA preparation

After the treatment duration indicated in Table 3.2, cells were washed twice with 5 ml of Phosphate Buffer Saline PBS (Gibco®, Life Technologies™) trypsinized with 1 ml of TrypLE™ Express (Gibco®, Life Technologies™) for 5 minutes at 37°C, harvested with 5 ml of complete growth medium and centrifuged 5 minutes at 3000 rcf at 4°C. Cellular pellets were dissolved in 1 ml of TRIzol® Reagent (Sigma-Aldrich®, St. Louis, MO, USA) and kept at -80°C until extraction. Total RNA was extracted following standard TRIzol® Reagent protocol (Sigma-Aldrich®) and RNA quality evaluation was performed by microcapillary electrophoresis chip with 2100 Bioanalyzer and RNA 6000 Pico Kit (Agilent Technologies, Santa Clara, CA, USA). RNA concentration was determined by spectroscopic method through NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, Barnstead, NH, USA). Total RNAs extracted from different cell lines were equally pooled as shown in Table 3.2, using 12.5 µg of RNA from each cell line. From total RNA of each pool mRNA purification was performed trough poliA(+) enrichment by Dynabeads® mRNA Purification Kit (Life Technologies™ Carlsbad, CA, USA). This method relies on A-T base pairing between the polyA tails of mRNA and short sequences of oligo-dT (25 bases), covalently bound to the surface of magnetic beads, and the subsequent magnetic isolation. mRNA enrichment was evaluated by microcapillary electrophoresis chip with 2100 Bioanalyzer and RNA 6000 Pico Kit (Agilent Technologies).

Any step involving manipulation of RNA samples was performed using RNase-free materials and reagents; all surfaces and instruments have been treated with RNaseZAP®(Sigma-Aldrich®) reagent to avoid any contamination by exogenous RNases.

3.2.3 Transcriptome analysis

3.2.3.1 RNA-Sequencing

Sequencing of mRNA samples extracted from fibroblasts was performed in collaboration with CRIBI Biotechnology Centre (University of Padova) using the SOLiD™ 3 (Sequencing by Oligo Ligation and Detection) Sequencing System (Life Technologies™) according to manufacturer's protocol. The whole process can be summarized in the following steps: cDNA library preparation, amplification by emulsion PCR, beads deposition, sequencing by ligation.

3.2.3.1.1 cDNA library preparation

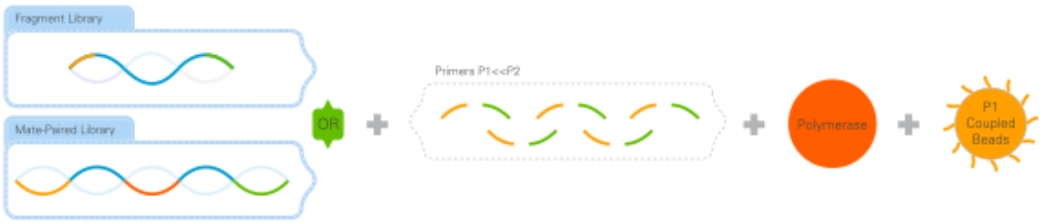
SOLiD™ Whole Transcriptome Analysis Kit (Life Technologies™) was used to convert the full set of mRNA transcripts expressed in the sample into a cDNA library for transcriptome analysis (Figure 3.1). mRNA underwent fragmentation by RNase III, clean-up and assessment of yield and size distribution. The last was performed by Quant-iT™ RNA Assay Kit with the Qubit® Fluorometer (Life Technologies™) and the RNA 6000 Pico Kit with the Agilent® 2100 Bioanalyzer (Agilent Technologies).

Forward and reverse adaptors were hybridized and ligated to the ends of RNA fragments in such a way to provide sequences required for the subsequent sequencing, and to preserve the orientation of the starting RNA molecules. In particular, one library was prepared using the Adaptor Mix A. This Adaptor Mix is set of oligonucleotides with a single-stranded degenerate sequence at one end and a defined sequence (complementary to P1 or P2 primers) required for SOLiD System sequencing at the other end. The Adaptor Mix constrains the orientation of the RNA in the ligation reaction such that hybridization with Adaptor Mix A yields template for SOLiD System sequencing from the 5' end of the sense strand.

Next, the RNA population with ligated adaptors was reverse transcribed to generate single-stranded cDNA copies of the fragmented RNA molecules. After a cleanup step using the MinElute® PCR Purification Kit (Qiagen, Venlo, NL), the samples were subjected to denaturing gel electrophoresis, and gel slices containing cDNA of length between 50 and 150 nt were isolated. The size-selected cDNA was next amplified by PCR reaction that takes place in the gel slices. This step appends terminal sequences, required for subsequent steps, to each molecule and generates sufficient template for SOLiD System sequencing. Following PCR, the amplified cDNA was cleaned up using the PureLink™ PCR Micro Kit (Invitrogen™, Life Technologies™).

The yield and size distribution of each cDNA library was assessed by Agilent 2100 Bioanalyzer with the DNA 1000 Kit (Agilent Technologies) and Qubit® Fluorometer (Life Technologies™).

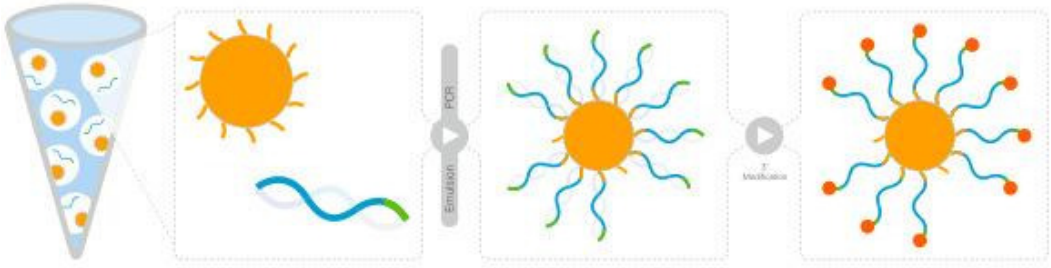
Figure 3.1 cDNA library preparation.



3.2.3.1.2 Amplification by emulsion PCR

An emulsion method was used to clonally amplified cDNA molecules onto 1 µm magnetic beads carrying many copies of P1 Adaptor attached to them (SOLiD™ P1 DNA Beads). Emulsion is made up of an oil phase containing emulsifiers and an aqueous phase, which includes PCR components (cDNA template, primers, DNA polymerase, and SOLiD™ P1 DNA Bead), see Figure 3.2.

Figure 3.2 Emulsion PCR.



Once the emulsion has been created using the ULTRA-TURRAX® Tube Drive (IKA®, Staufen, Germany) droplets of aqueous phase, called micro-reactors, form in which the amplification takes place.

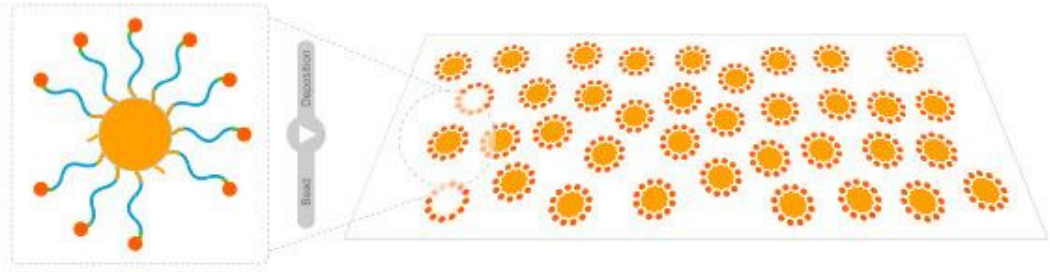
Then amplification was performed at standard PCR condition and about 30,000 or more copies of template were amplified onto each SOLiD™ P1 DNA Bead. After emulsion PCR was complete, the emulsion was broken with 2-butanol, and the beads were washed to clear away the oil and the emulsifiers. In order to isolate templated beads from non-templated or poorly

amplifying beads, an enrichment step was performed in which polystyrene beads with a single-stranded P2 Adaptor attached were used to capture templated beads.

3.2.3.1.3 Beads deposition

The amplified templates underwent a chemical modification in order to prepare the P2-enriched beads for deposition (Figure 3.3). In particular, a dUTP residue was added at the 3'-end of the template using a terminal transferase reaction. Then the 3' modified beads were deposited and covalently attached to the surface of a specially treated glass slide that was placed into a fluidics cassette within the sequencer.
In this study 4 samples (C, UT, T1, T2) was deposited simultaneously on the slide and physically separated in order to be analyzed at the same time.

Figure 3.3 Beads deposition.



3.2.3.1.4 Sequencing by ligation

As Figure 3.4 shows, ligase-mediated sequencing begins by annealing a primer (called universal seq primer n) to the shared adapter sequence on each amplified fragment (P1 adaptor); then DNA ligase is provided along with specific fluorescent-labeled probes. These probes are octamers consisting, from 5' end, of 3 universal nucleotides, 3 degenerated nucleotides, 2 nucleotides which are encoded by the 5'-attached fluorescent dye. The probes are labeled with 4 different colours, to each of which four specific combinations of 2 bases in the 7th and 8th positions, correspond. Therefore 4^5=1024 different probes are used, 256 for each colour. The scheme of two base-encoding is shown in Figure 3.5. When a matching octamer hybridizes to the DNA fragment sequence adjacent to the universal primer 3' end, DNA ligase seals the phosphate backbone. The ligation step is followed by fluorescence detection, a treatment with phosphatase (to cap unextended strands), and a regeneration step in which the three universal

bases at the 5' end of the ligated octamer, including the fluorescent group, are cleaved. Then subsequent cycles of ligation, detection, capping and regeneration are performed allowing extension of synthesizing sequence. In the primer reset step, the synthesized sequence is removed by melting and a second round of sequencing initiates using a primer (called universal seq primer n-1) designed in such a way that it hybridizes on the P1 adapter, one nucleotide upstream with respect to the universal primer used in the first step. Overall five rounds of sequencing were performed with universal seq primer n, n-1, n-2, n-3 and n-4. With this system each template base is interrogated twice in two independent ligation reactions by two different primers, providing increased confidence in each call. The result of sequencing process is a sequence expressed in the so called colour space in which each colour represents 4 potential two base combinations. The conversion into nucleotide base space is possible using the two base-encoding matrix. However the conversion is done after the sequence is aligned to a reference genome transcribed in colour space. The two-base encoding translates into more accurate alignment of short reads, polymorphism discovery and consensus calling.

Each SOLiD (Sequencing by Oligo Ligation and Detection) run required about 5 days and produced 3–4 Gb of sequence data with an average read length of 35 bp.

Figure 3.4 Sequencing by ligation.

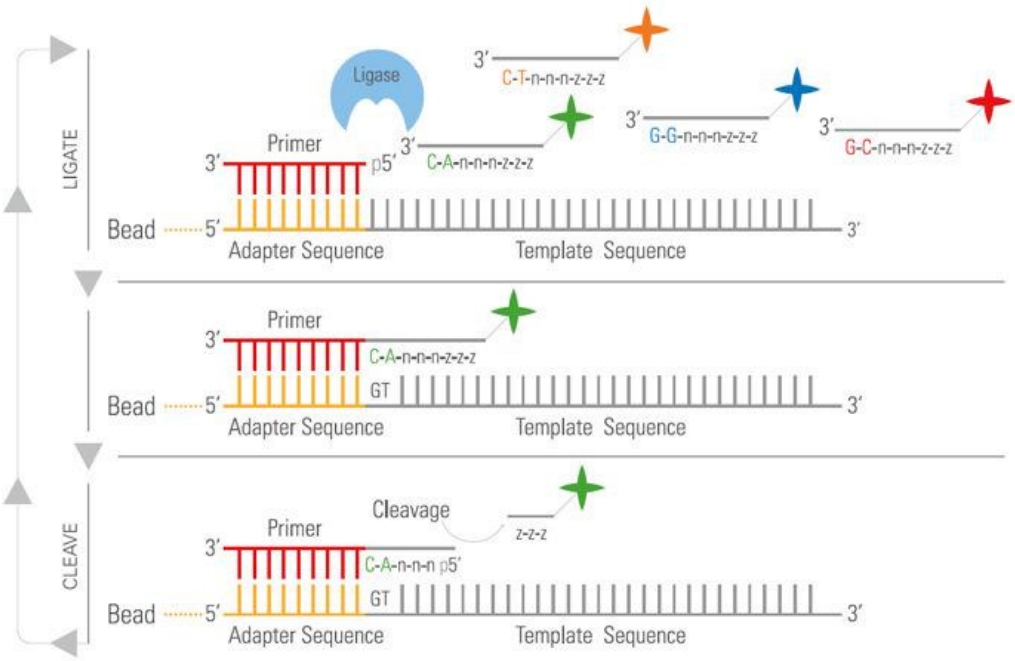
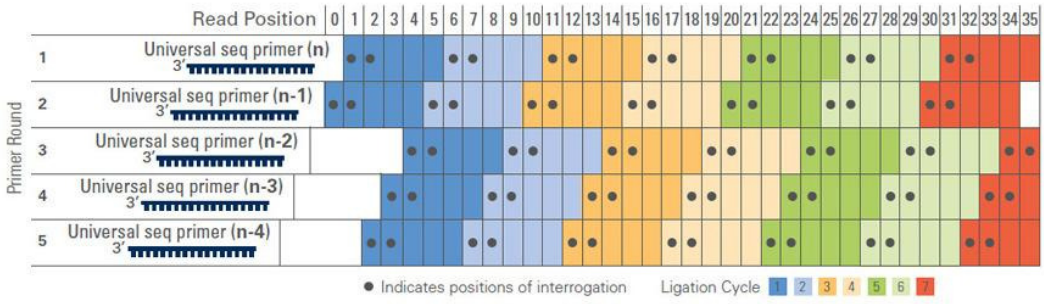


Figure 3.5 2 base colour code.



3.2.3.2 Data analysis

3.2.3.2.1 Alignment

The SOLiD™ 3 System produced sequences of 35 bp (termed reads), which undergo alignment on human genome. The alignment was performed by CRIBI Biotechnology Centre (University of Padova) using the PASS software (Campagna et al. 2009) http://pass.cribi.unipd.it). PASS (Program to Align Short Sequences) performs gapped and ungapped alignment onto a reference sequence and is designed to handle huge amounts of short reads generated by Next Generation Sequence (NGS) technologies, including SOLiD™ System. As a first step, PASS makes an index of seed words (with default length of 12 bases) occurring in the genomic sequence. Then PASS scans every read executing three steps: (1) it identifies seed words matching between read and the genomic index; (2) for every match it checks if it is possible to extend the alignment to the regions flanking the seed words; (3) if step (2) is passed, then it refines the alignment and/or score it. To verify the possibility of extending the seed a precomputed table of all the possible short words aligned against each other is used. The final alignment is performed by exact dynamic programming. The alignment is preceded by a quality control of the reads, defined by three main parameters (w, q and max_trim_leng) which the software calculates on a sample subset of reads randomly chosen from whole library. Bases at the end of each read are trimmed if they did not reach the minimum quality score q into a window of length w; if trimming produces a read shorter than Max_trim_leng parameter, it is discarded.

The February 2009 human genome assembly (GRCh37), publicly available at UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/) was taken as reference sequence and the options best-hit, gap=0 and maximun mismach=3 have been selected. Only unique reads

(aligning on only one gene) and only genes with coverage higher than 50% of gene length were considered.

For each sample, the raw expression value of each gene was calculated as the number of reads uniquely mapping to that gene.

3.2.3.2.2 Identification of differentially expressed genes (DEGs)

Among RNA samples analyzed with RNA-Seq, the following comparisons were performed (see Table 3.2): UT vs C, T1 vs C, T2 vs C, T1 vs UT and T2 vs UT.

The R package DEGseq (Wang et al. 2010) www.bioconductor.org/ packages/2.6/bioc/ html/DEGseq. html] was employed to identify differentially expressed genes. The method used by this software models RNA Sequencing as a random sampling process, in which each read is sampled independently and uniformly from every possible nucleotide in the sample. Under this assumption the number of reads coming from a gene follows a binomial distribution and could be approximated by a Poisson distribution. The method is based on MA-plot, which is a statistical analysis tool having been widely used to detect and visualize intensity-dependent ratio of gene expression data. Let C_1 and C_2 denote the counts of reads mapped to a specific gene obtained from two samples, with $C_i \sim$ binomial (n_i, p_i), i = 1,2, where n_i denotes the total number of mapped reads and p_i the probability of a read coming from that gene. M and A are defined as $M = \log_2 C_1 - \log_2 C_2$, and $A = (\log_2 C_1 + \log_2 C_2)/2$.

It can be proved that under the random sampling assumption the conditional distribution of M given that A=a (a is an observation of A), follows an approximate normal distribution. For each gene on the MA-plot, the hypothesis to be tested are H_0: $p_1=p_2$ versus H_1: $p_1 \neq p_2$. Then a p-value could be assigned based on the conditional normal distribution. Likelihood ratio test was used as statistical hypothesis test.

Genes meeting the subsequent criteria were considered to be differentially expressed:

$$\left|\log_2 \frac{c_1}{c_2}\right| \geq 1; \quad p-value < 0.001$$

Since DEGSeq is no designed to handle null count values, count values equal to zero were converted to 0.5.

3.2.3.2.3 Functional analysis

The lists of genes resulted differentially expressed in each comparison underwent functional analysis.

Enrichment is a measure of the probability that a gene list contains more members belonging to a given biological category than those expected by chance, taking into account the total number of members of the category and the number of genes in the list to be analyzed. Different tools were used based on different annotation systems.

Gene Ontology

Gene Ontology (GO) is a controlled vocabulary of terms which have the aim of annotating genes and gene products and of describing their features in order to standardize their representation across species and databases.

The structure of GO can be described in terms of a graph, where each GO term is a node, and the relationships between the terms are arcs between the nodes. The relationships comprise is *a* (is a subtype of), *part of*, and *regulates*, *negatively regulates* and *positively regulates*. The relationships used in GO are directed and the graph is acyclic, meaning that cycles are not allowed in the graph. The ontologies resemble a hierarchy, as child terms are more specialized and parent terms are less specialized, but unlike a hierarchy, a term may have more than one parent term. The ontology covers three domains: *Biological process*, *Molecular function* and *Cellular component*. A *Biological process* is defined as a series of operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units (cells, tissues, organs, and organisms) in which the gene (or gene product) is involved. *Molecular function* describes the elemental activities of the gene (or gene product), such as catalytic or binding activities, that occur at the molecular level; the *cellular component* ontology refers to the parts of a cell or its extracellular environment, at the levels of subcellular structures and macromolecular complexes, related to the gene (or gene product).

PANTHER (Protein ANalysis THrough Evolutionary Relationships)

The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System [http://www.pantherdb.org/] is designed to classify proteins (and their genes) according to the systems: Family and subfamily, Pathways and Ontologies, the last comprehending the three GO domains plus the PANTHER Protein Class ontology. Statistical analysis performed is based on binomial test. The PANTHER Classifications are the result of human curation as well as bioinformatics algorithms. It contains information about 7729 protein families and 176 Pathways.

ExPlain™

ExPlain™ [http://www.biobase-international.com/] is a data analysis system that combines functional annotation, pathway analysis and promoter analysis tools. It has been used for promoter analysis. Using the databases of eukaryotic transcription factor, TRANSFAC®, ExPlain™ enables also to identify transcription factors affecting gene expression in a list of differentially expressed genes as well as predict how they, in combination, can induce observed gene expression patterns. The system also permits to extend the analysis further and gain insight into the key upstream signaling regulators influencing the activity of these transcription factors.

TRANSFAC® contains published data on eukaryotic transcription factors (TFs), their experimentally-proven binding sites, and regulated genes. It includes 19524 TFs, 37381 TF binding sites, 50219 TF-site links, 277337 promoter sequences (40501 of which are human) and 28425 references.

IPA®

IPA® (Ingenuity® Pathways Analysis) is a web-based software application that enables to analyze and integrate data derived from gene expression, proteomics and metabolomics experiments.

It leverages the Ingenuity® Knowledge Base, a repository of biological interactions and functional annotations created from individually modeled relationships between proteins, RNAs, genes, isoforms, metabolites, complexes, cells, tissues, drugs, and diseases. These modeled relationships, or Findings, are manually reviewed. The database includes 19635 human genes, 1.6 million interactions and functional annotations, 14805 chemicals, 3 million literature referenced Findings. Functional analysis is based on 3 categories: *Molecular and Cellular Functions*, *Physiological System Development and Functions* and *Diseases and Disorders.* These categories are organized in 85 subcategories, in turn classified in categories of lower level.

Canonical Pathway Analysis tool has been used to identify metabolic and signalling pathways associated to gene lists. Functional categories more represented in the gene lists given as input are identified by Fisher's exact test which returns, as measure of the significance of the association, a p-value that represents the probability that the association between the category or pathway and the gene list is due to chance. For pathway analysis, a ratio between

the number of genes of the list associated to that pathway and the total number of genes of that pathway is also returned.

DAVID

DAVID (The Database for Annotation, Visualization and Integrated Discovery) bioinformatics resources consists of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from large gene/protein lists. Functional Annotation Tool was used for enrichment analysis of DEG lists (Huang da, Sherman & Lempicki 2009).

Enrichment analytic algorithm can be applied over 40 annotation categories, including GO terms, protein-protein interactions, protein functional domains, disease associations, bio-pathways, sequence general features, homologies, gene functional summaries, gene tissue expressions, literatures, etc. The tool adopts the Fisher Exact Tests to measure the gene-enrichment in annotation terms and returns the EASE Score, a modified Fisher Exact P-Value.

The annotation systems GOTERM_BP_FAT, GOTERM_MF_FAT, GOTERM_CC_FAT, Swiss-Prot (SP) and Protein Information Resource (PIR) Keywords (SP_PIR_KEYWORD).

BINGO

BiNGO (Biological Networks Gene Ontology tool) is a Java-based tool to determine which GO categories are statistically over- or under-represented in a set of genes or in a subgraph of a biological network (Maere, Heymans & Kuiper 2005). BiNGO is implemented as a plugin for Cytoscape (version used 2.8.0), which is a an open source bioinformatics software platform for visualizing and integrating molecular interaction networks (Smoot et al. 2011). BiNGO maps the predominant functional themes of a given gene set on the GO hierarchy, and outputs this mapping as a Cytoscape graph. BiNGO was launched on lists of DEG and mapped on GO ontology Biological Process selecting the following settings: assessment of Overrepresentation, Hypergeometric test as statistical test, Benjamini & Hochberg False Discovery Rate as multiple testing correction and 0.05 as significance level.

KEGG

KEGG (Kyoto Encyclopedia of Genes and Genomes) [http://www.genome.jp/kegg/] is an integrated database resource consisting of sixteen databases which are broadly categorized into *Systems information*, *Genomic information* and C*hemical information*. The Database *KEGG*

Pathway, belonging to *Systems information* category, is a collection of manually drawn pathway maps grouped in the categories: Global Map, Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems, Human Diseases and Drug Development. Currently KEGG Pathway database contains 435 maps. *KEGG Pathway mapping* tool was used to map RNA-Seq data to the KEGG pathway maps for facilitating visualization and biological interpretation.

3.3 Research of efficacy biomarkers

3.3.1 Selection of candidate biomarkers

RNA-Seq data were used to identify potential candidate biomarkers of therapeutic efficacy considering the following comparisons: UT vs C, T1 vs C, T2 vs C. Genes having, for each of these comparisons, a p-value of the statistical test for differential expression < 0.001 were clustered according to their expression ratios in these comparisons. The software STEM (Short Time-series Expression Miner) v. 1.3.7 (Ernst, Bar-Joseph 2006) , a Java program for clustering, comparing, and visualizing short time series gene expression data, with 8 time-points or fewer, was used. The maximum number of Model Profiles was set at 40.

Among the profiles obtained, those showing a temporal expression pattern which could be potentially ascribable to an efficacy biomarker were selected. The lists of genes associated to these profiles were filtered depending on their expression levels in the biological fluids, more precisely on the expression of transcript in the whole blood and the presence of the codified protein in plasma and urine. Genes meeting at least one of the subsequent criteria were selected:

- gene expression value greater than 3 in the databases BioGPS (http://biogps.org) for the field "whole blood";
- presence of the codified protein in the database HUPO Plasma Proteome Project database (http://www.hupo.org/);
- presence of the codified protein in the database MAPU Proteome database (http://www.mapuproteome.com/).

A final selection was carried out choosing from the selected genes the first 20 presenting the highest percentage gene expression variation in the comparison T1 vs UT.

3.3.2 Clinical data collection and analysis

Identified candidate biomarkers were evaluated in a population of MPS II patients undergoing enzyme replacement therapy. These data are part of a wider study aimed to evaluate ERT efficacy by a 3 years follow up of 27 MPS II patients. Data collection relative to the whole study was performed creating a relational database whose schema is reported below:

PATIENT (Patient ID, Gender, Birth Date, Informed Consent Date, Therapy Beginning Date)
DIAGNOSIS MPS II (Mutation, Phenotype)
PHYSICAL EXAMINATIONS (Date, Weight, Height, Cranial Circumference, Cardiac Frequency, Respiratory Frequency, Systolic Pressure, Diastolic Pressure, Protuberant Liver, Protuberant Spleen, Puberal Development)
HEMATOCHEMICAL EXAMINATIONS (Date, Erythrocytes, Haemoglobin, Hematocrit, MCV, MCH, MCHC, Platelets, Reticulocytes, Leukocytes, Neutrophils, Lymphocytes, Monocytes, Eosinophils, Basophils)
BIOCHEMICAL EXAMINATIONS (Date, BUN, Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase, ALT, AST, LDH, GGT, CPK, Triglycerides, Cholesterol)
URINARY GAGS (Date, Value)
ENT (Date, Tonsillar Hypertrophy, Adenoidal Hypertrophy, Tonsillectomy, Adenoidectomy, Otitis Media, Hypoacusia, Sleep disorders, Audiometry (Air and Bone Conduction Tests), Spirometry)
6MINUTES WALKING TEST (Date, Walking Duration, Distance, With Assistance, Oxigen Saturation*, Cardiac Frequency*, Respiratory frequency*)
EYE EVALUATION (Date, Visual Acuity*, Corneal clouding*, Visual Field*, Intraocular pressure *, Fundus Oculi*, Visual Evoked Potentials*)
NEURORADIOLOGY (Date, Evaluation Performed, Normal, Ventricular Dilation, Cerebral Atrophy, Substantia Lesions, Perivascular Alterations, Other Anomalies)
NEUROPHYSIOLOGY (Date, Somatosensory Evoked Potentials (SSEP)*, ElectroEncefaloGraphy (EEG): Normal)
NEUROLOGICAL EVALUATION (Date, Abnormal Brain Imaging, Cognitive Impairment, Seizure, Neurosurgical Treatment)
ELECTROCARDIOGRAPHY (Date, PR duration, QRS duration, QT duration, QTc duration, Anomalies)

ECOCARDIOGRAPHY (Date, IVSd, IVSs, LVDd, LVDs, PWTd, PWTs, Ejection Fraction, Anomalies, Mitral Regurgitation, Mitral Stenosis, Aortic Regurgitation, Aortic Stenosis, Tricuspid Regurgitation, Tricuspid Stenosis, Pulmonary Regurgitation, Pulmonary Stenosis)
ABDOMEN ULTRASOUND (Date, Liver Dimensions, Spleen Dimensions, Report: Normal, Hepatomegaly, Splenomegaly)

The attribute Paz_Id present in all of the relations satisfies the referential integrity property with the relation PATIENTS. The symbol * denotes composite attributes, whose atomic values are not reported for sake of simplicity.
From these database some parameters were extracted, according to their informative value and data availability, and used for performing correlation analysis with molecular data. The selected parameters are: Hepatomegaly, Splenomegaly, Valvulopathies, Otitis media, Hearing Impairment, Tonsillar Hypertrophy, Adenoid Hypertrophy, Sleep disorders, Brain Imaging abnormalities, Cognitive impairment, Hydrocephalus and Seizure. Data were reported as dichotomous variables representing the presence or absence of the pathological phenotype.

3.3.3 Blood samples

Candidate biomarkers of therapeutic efficacy selected from transcriptome analysis have been evaluated by quantitative PCR in patients under ERT treatment. 7 patients, listed in Table 3.3, are evaluated.

Table 3.3 List of patients whose blood was analyzed and related information about age at start of ERT, age at diagnosis, mutation and phenotype.

Patient ID	Age at start of ERT (YEARS)	Age at DIAGNOSIS (YEARS)	Mutation	PHENOTYPE
A1	1,6	0,9	deletion exons 1-7	severe
A5	2,6	2,4	p.K236K	attenuated
A8	3,4	2,8	p.R468Q	severe
A12	4,7	4,4	p.D198N	severe
B4	8	3,1	IDS-IDS2 recombination	severe
B7	11,4	7,8	p.R271W	attenuated
C4	16,7	6,8	NA	attenuated

In Table 3.4 the list of blood samples analyzed is provided. Samples collected at about 3, 6, 7, 11, 13, 18, 22, 28, 31 and 41 months from the start of therapy and pre-ERT samples, if

available, were analyzed.

Table 3.4 List of the blood samples analyzed for each patient. In the second indicates if there is or not the sample pre-ERT; in the following lines it's shown after how many months and day samples post-ERT for each patient were collected. NA=not applicable; m=mo.

PATIENT ID	A1	A5	A8	A12	B4	B7	C4
PRE-ERT	NA	yes	NA	yes	NA	yes	yes
POST-ERT	2 m, 3 d	2 m, 5 d	3 m, 21 d	4 m, 16 d	2 m, 3 d	3 m, 22 d	5 m, 20 d
	4 m, 6 d	4 m, 6 d	4 m, 9 d	6 m, 19 d	4 m, 13 d	5 m, 9 d	7 m, 23 d
	6 m, 9 d	7 m, 13 d	6 m, 12 d	8 m, 22 d	6 m, 8 d	13 m, 17 d	14 m, 16 d
	13 m, 2 d	10 m, 1 d	14 m, 18 d	10 m, 25 d	13 m, 2 d	18 m, 6 d	18 m, 13 d
	18 m, 23 d	14 m, 7 d	18 m, 18 d	14 m, 24 d	18 m, 22 d	21 m, 27 d	23 m, 3 d
	22 m, 28 d	20 m, 4 d	23 m, 15 d	18 m, 14 d	22 m, 27 d		29 m, 19 d
	29 m, 9 d	22 m, 7 d	27 m, 20 d	23 m, 20 d	29 m, 28 d		
	31 m, 5 d	29 m, 16 d	32 m, 25 d	27 m, 3 d	31 m, 8 d		
	40 m, 8 d	32 m, 11 d	39 m, 9 d	29 m, 6 d	40 m, 10 d		
		41 m, 16 d		39 m, 7 d			

3.3.4 RNA preparation

Blood samples collected from patients were stored in PAXgene Blood RNA Tubes (Qiagen, Venlo, NL) which contain a proprietary reagent composition based on a patented RNA stabilization technology. This reagent composition protects RNA molecules from degradation by RNAses and minimizes ex vivo changes in gene expression. RNA purification was performed by PAXgene Blood RNA Kit (Qiagen). RNA quality was evaluated by microcapillary electrophoresis chip with 2100 Bioanalyzer and RNA 6000 Pico Kit (Agilent Technologies); RNA concentration was determined by spectroscopic method through NanoDrop ND-1000 Spectrophotometer (Thermo Scientific).

3.3.5 Quantitative PCR

Table 3.5 shows the steps of retrotranscription reaction, which were performed on the thermocycler Applied Biosystems GeneAmp® PCR System 9700 (Applied Biosystems, Life Technologies™). The reaction which underwent the annealing step is described in Table 3.6. Then, in the extension step, 10 µl of the mix shown in Table 3.7, were added on ice to each sample.

Table 3.5 Steps of retrotranscription reaction.

Step	T(°C)	time (min)
Anealing step	65	5
Extention step	55	20
Heat inactivation	75	15

Table 3.6 Components of the annealing step.

Component	µl per reaction
RNA template (500 ng)	X
oligo dT primer	1
RNAse/DNAse free water	X
Final volume	**10**

Table 3.7 Mix added to each sample in the extension step.

Component	µl per reaction
nanoScript 10X Buffer	2
dNTP mix 10mM of each	1
DTT 100mM	2
RNAse/DNAse free water	4
nanoScript enzyme	1
Final volume	**10**

The so obtained cDNA was used as template for real-time quantitative PCR exploiting SYBR Green technology. The reaction was performed in 96 well plates PrimerDesign by using 7900HT Real-Time PCR System (Applied Biosystems, Life Technologies™) according to the amplification protocol shown in Table 3.8. Table 3.9 illustrates the reaction mix prepared for each well. Precision™ 2X qPCR Mastermix (PrimerDesign), is a 2x reaction buffer containing a thermo-stable Taq Polymerase (0.025 U/µl), 5 mM MgCl2, dNTP Mix (200µM each dNTP) and the SYBRgreen reagent for DNA stain. As template cDNA obtained for retro-transcription reaction was used after being diluted 1:5. PrimerDesign customer designed primer were used; manufacturer guarantees for priming specificity, tested by post run melt-curve analysis, and for a priming efficiency (>90%). Primers were also designed in such a way that all transcriptional isoforms were amplified, template does not form secondary structures, primers do not contain SNPs and do not form harpin and dimers.

Table 3.8 Amplification protocol used for the real-time qPCR.

	Step	T (°C)	Time
	Enzyme activation - HotStart	95	10'
Cycling x40	Denaturation	95	15''
	Data collection	60	1'

Table 3.9 Reaction mix added to each well for the real-time qPCR.

Component	**µl per reaction**
Precision^{TM} 2X qPCR Mastermix with SYBRgreen	10
cDNA Template	5
Primer mix	1
RNAse/DNAse free water	4
Final volume	**20**

Table 3.10 shows, for each gene analyzed, RefSeq accession number of the transcript variant taken as reference, the sequence and the position of sense and antisense primers and the amplicon length.

To improve the accuracy and reliability of qPCR results, quality control checks were performed at various level: control on the PCR replicate variation, assessment of positive and negative control samples, determination of the reference target expression stability.

As for technical replicates, in each run, each assay was performed in triplicate. The Cq values of all wells of a replicate group were averaged at the very beginning of the calculation process. As replicate variability parameter, the difference in Cq value between the replicate with the highest Cq value and the replicate with the lowest Cq value (ΔCq) was chosen. Replicate variability threshold was set at 0.5; outliers (samples with ΔCq >0.5) were excluded so they do not take part in the calculations.

In each run a negative control, intended as no template control (NTC), was included for each gene (primer pair). An amplification signal in the NTC sample indicates a potential contamination issue or formation of primer dimers. Such problems can be ignored as long as the difference in Cq value between the NTC and the sample with the highest Cq value is sufficiently large. This difference was chosen as negative control parameter with a threshold of 5. A Cq value difference of 5 corresponds to a fold difference of about 32, indicating that approximately 3 % of the signal may be caused by these unwanted signals.

PCR amplification efficiency was established, in a gene specific manner, by means of calibration curves. The calibration curves were constructed from the Cq and quantity values of a dilution series of pooled cDNA samples.

Table 3.10 Primer used for qPCR. RefSeq accession number of the transcript variant taken as reference, the sequence and the position of sense and antisense primers and the amplicon length for each gene are reported.

			Sense Primer		Antisense Primer	
Official gene symbol	**RefSeq ID**	**Amplicon_Length**	**Sequence**	**Position**	**Sequence**	**Position**
CHCHD2	NM_016139	121	CCTGTCATTCCTGGCTTCCT	650	TCATAACATAACAGCTAGTTTAAGGAG	770
COX6C	NM_004374	128	ATGGTTTGGGACTCTGTGTATG	94	CATTTCGTGGTCTGCTTCTTTC	221
DPP7	NM_013379	101	CGCTGTCAGACGAGAAGGA	773	CAGGAAGTCAGTGGGGTAGG	873
DPT	NM_001937	129	GATGATGGGTGGGTGAATTTGA	133	GCAGGCGTAGTTCCATTGTC	261
HIST1H2BK	NM_080593	107	TCCAAAGAGAATCATTTACAAGTTAAT	622	GAGGAGATGGGTAATGTAGACAA	728
ISG20	NM_002201	115	GAGTGAGCGCCTCCTGC	692	GCGGGCTCGGATTCTCTG	806
NMI	NM_004688	97	CATTCACTTTCAACGGGCAAA	1161	AGTCTATTCTTCAAAGTATGCTATGTG	1257
PFN1	NM_005022	122	GTCTGTCCCTTCCCCTTCA	564	CATGTGGTTTTGGCAGCAATA	685
PHPT1	NM_001135861	80	CCAGCACGCCATTTCAACT	982	CTCAGTAGCCGTCGTTAGC	1061
PKN1	NM_213560	92	CTTCGGAGACCCAGGAGAC	1804	CACCGCCAGGAACTTGAAAT	1895
PNPLA6	NM_001166111	110	TCTTTCATCGGAGCGTTGTAC	3289	AGGTCCAACACAGGTTCCA	3398
RAB11B	NM_004218	112	CGACGAGTACGACTACCTATTC	111	GCTCTTGCTCTCCAGGTTGA	222
RGS2	NM_002923	114	GAACAACTCTTATCCTCGTTTCTT	579	ATTTCTGGGCTCCCTTTTACAT	692
RPL41	NM_021104	89	GACCCACCTTGCTCATAAACA	320	GCCACAAATAAACACAGAAATAGTC	408
SCAMP4	NM_079834	103	GAGATGTCAGAAAAGGAGAACAAC	105	GGTGCTCCACTGGGATCTC	207
SPR	NM_003124	122	TGAACAACTACTGGGCACTGA	409	GGCACAGAGGGACGAGATG	530
TBL3	NM_006453	110	CTGTGGCCTCTTCCCAAAG	1444	CTGTTGATGTCCTTATCATGGC	1553
TMSB10	NM_021103	87	AACACCCTGCCGACCAAAG	174	GAGGACGGGGGTAGGAAATC	260
TSPO	NM_000714	84	CAGTGTCCTGTGCTTTCTGC	782	AAAGGAAGTCAAAAACTTTATTCAGC	865
ZFAT	NM_020863	120	CAGTCAGCATCACCAGCAATA	4085	GTGTGGGGTGTGTGTAGGG	4204

The expression values of each gene in each time point were compared to that of an healthy control, obtained as pool of RNA extracted from pediatric blood samples, and to that of a pre-treatment samples, where available. The expression values were calculated through the delta-delta-Ct method modified in order to take into account multiple reference genes, gene-specific amplification efficiencies as well as the errors on all measured parameters along the entire calculation track (Hellemans et al. 2007). As reference genes UBC (ubiquitin C), RPL13A (ribosomal protein L13a) and PPIA (peptidylprolyl isomerase A) were chosen. Each assay was

performed in triplicate.

The experimental set-up was chosen according to the sample maximization method: as many samples as possible are analyzed in the same run in order to minimize the inter-run variation effects.

3.3.6 Statistical analysis

Gene expression data of the 4 selected candidate biomarkers measured in blood samples were correlated with clinical data collected during patients' follow-up. In particular the following clinical parameters were considered: hepatomegaly, splenomegaly, valvulopathies, otitis media, hearing impairment, tonsillar hypertrophy, adenoid hypertrophy, sleep disorders, CNS abnormalities by brain imaging, cognitive disturbances, seizures and hydrocephalus; they were expressed as dichotomous variables i.e. as presence or absence of the symptom at the date the examination was performed. For each gene, qPCR data were expressed as ratio on healthy control.

Where possible, each clinical evaluation was paired with qPCR values obtained from the corresponding blood sample with a maximum gap of 365 and 100 days for data preceding and following the start of therapy respectively; no pairing between pre-ERT and post-ERT samples were performed. The correlation between each clinical parameter and each gene was evaluated through the point biserial correlation coefficient (r_{pb}); r_{bs} significance was tested using Student's t-test with p-value <0.05.

4 RESULTS AND DISCUSSION

The overall aim of this study is to increase our knowledge about Mucopolysaccharidosis type II (MPS II) in terms of pathogenetic mechanisms and response to ERT treatment

The study consists in an *in vitro* experiment of whole transcriptome analysis conducted on fibroblasts derived from MPS II patients. The overall gene expression profile of these pathological cells was compared to that of healthy controls in order to highlight any variation in cellular biology and functions caused by the disease.

In parallel, clinical data were collected from a population of MPS II patients undergoing ERT treatment, taking into account several parameters measured during a 3 year follow-up.

Finally the information obtained by these two levels of analysis were converged into the research of molecular biomarkers able to follow the clinical status of the patients during disease progression and therapy. Candidate biomarkers were first identified starting from gene expression data obtained in the *in vitro* experiment and then evaluated in blood samples collected at different time-points from some patients belonging to the population under clinical study. A correlation analysis between molecular and clinical data is reported in the conclusive part of the study.

4.1 RNA-Seq: sample processing and data extraction

A whole transcriptome analysis was conducted on fibroblasts derived from patients affected by Hunter syndrome and healthy controls. Hunter cells were also treated with the enzyme IDS, in the same formulation used in ERT, and collected at two different time-points, at 24 and 144 hours from the beginning of the treatment. Four and five different lines were treated separately for Hunter cells and control fibroblasts respectively and pooled as described in Table 4.1, illustrating the experimental design, just before RNA sequencing.

Table 4.1 Experimental design of RNA-Seq experiment.

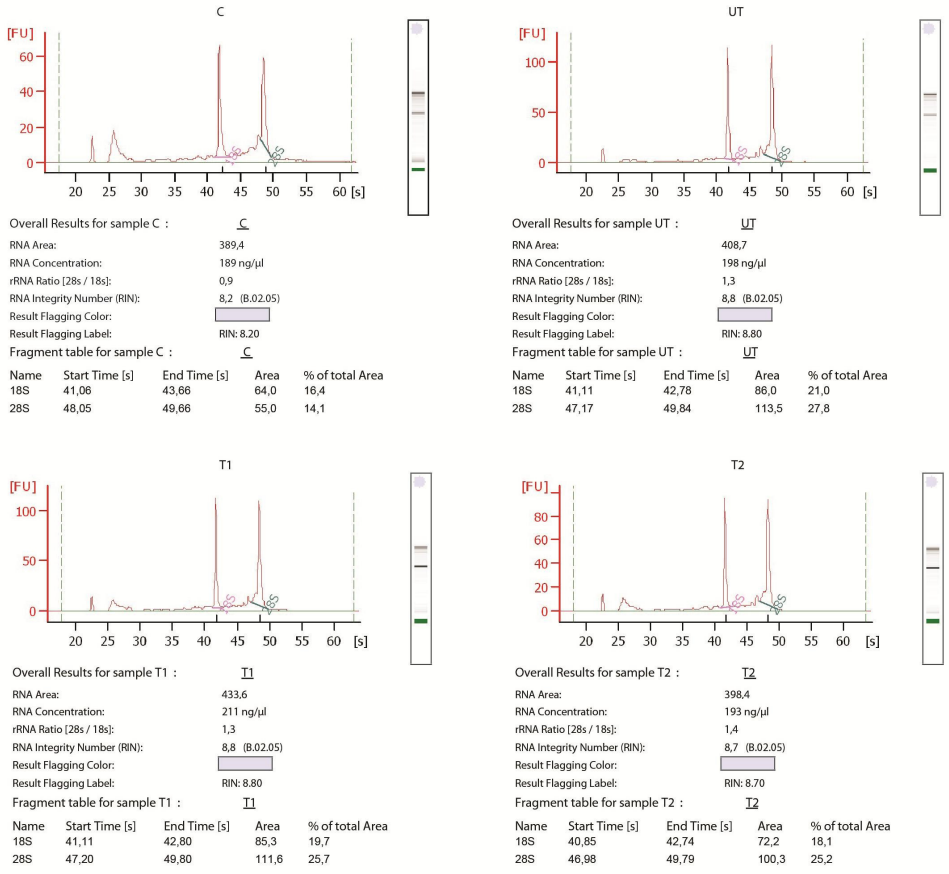
ID SAMPLE	GENOTYPE	TREATMENT	TREATMENT DURATION (h)	LINES IN POOL
C	Healthy	NaCl 0.9%	24	4
UT	MPS II	NaCl 0.9%	24	5
T1	MPS II	62.5 nM IDS	24	5
T2	MPS II	62.5 nM IDS	144	5

As described in Chapter 3, transcriptome analysis was preceded by some steps of RNA preparation, including extraction of total RNA from each cell line, pooling into 4-5 samples, evaluation of RNA quality and mRNA purification.

4.1.1 RNA preparation

Total RNA extracted from fibroblasts were evaluated separately for each cell line before pooling the four or five samples. As shown by electropherograms reported in Figure 4.1, RNA is of good quality with a mean R.I.N. value of 8.6.

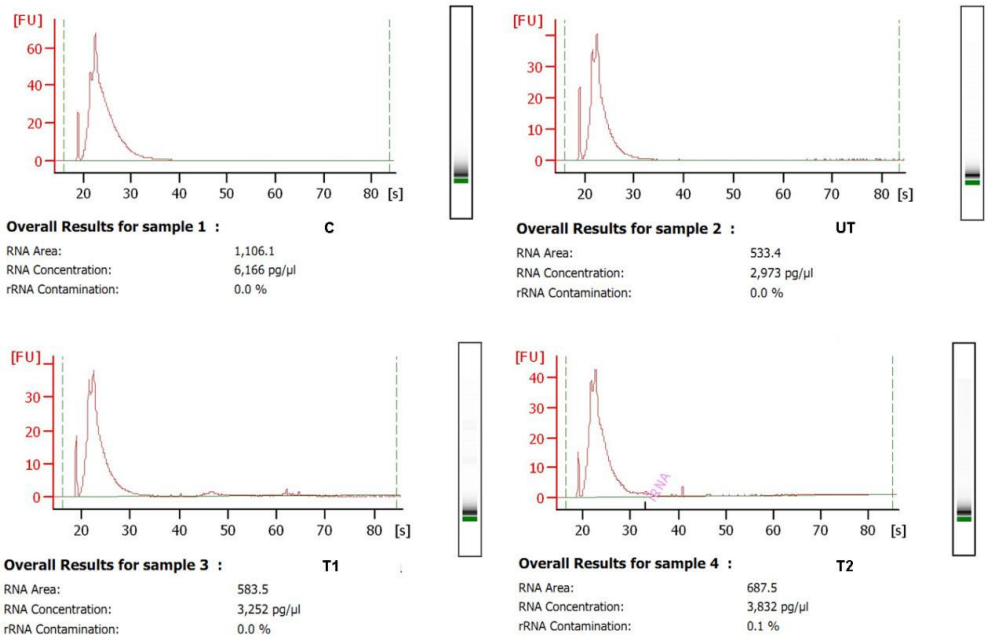
Figure 4.1. Examples of Electropherograms of total RNA extracted from each cell line. Data obtained through 2100 Bioanalyzer, RIN=RNA Integrity Number.



Following RNA extraction, total RNAs from different cell lines were pooled, as shown in Figure 4.1, into the four samples: C (healthy fibroblasts), UT (MPS II fibroblasts, untreated), T1 (MPS II fibroblasts, treated for 24h) and T2 (MPS II fibroblasts, treated for 144h). Hereafter in the text samples will be referred using this abbreviations.

mRNA, which represents only 1-5 % of total RNA in a typical mammalian cell, was purified trough poliA(+) enrichment, obtaining on average 0.96 µg of mRNA from 50 µg of total RNA. Purification efficacy was evaluated by 2100 Bioanalyzer, whose results are shown in Figure 4.2 rRNA contamination resulted almost null.

Figure 4.2. Electropherograms of RNA sample after mRNA purification.



4.1.2 RNA sequencing and data extraction

4.1.2.1 Sequencing

Sequencing of mRNA samples obtained from fibroblasts was performed in collaboration with CRIBI Biotechnology Centre (University of Padova) using the SOLiD™ 3 (Sequencing by Oligo Ligation and Detection) System.

Thanks to this technology the whole set of mRNA transcripts expressed in the sample can be converted into a cDNA library which is amplified by emulsion PCR and sequenced through the next generation sequencing methodology described in details in Chapter 3. In this way a high number of short sequences (of about 35 bp), called reads, were produced, with an overall throughput, for the 4 samples, of about 3GB.

4.1.2.2 Alignment

In order to extract information on gene expression values of all known genes in the 4 samples, these reads need to be aligned to a human genome.

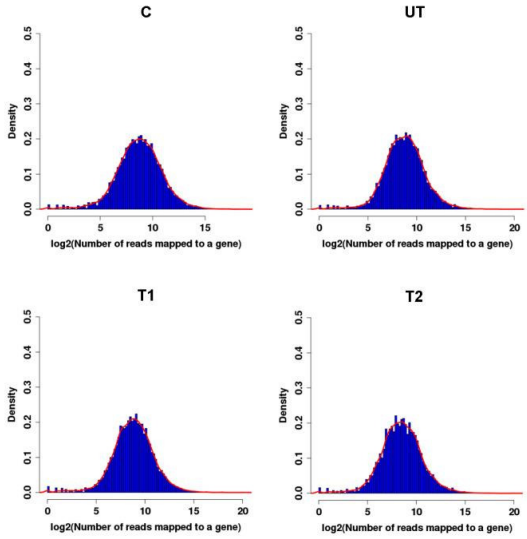
Table 4.2 shows, for each sample, the number of reads obtained from sequencing, the number of reads passing the quality control and the number of reads aligned to the genome. On average, 50.5% of reads passed quality control, 72.2% of which align to human genome.

Table 4.2. Sequencing and alignment result in terms of reads. The alignment was performed through PASS software; http://pass.cribi.unipd.it].

		READS PASSING QUALITY CONTROL (QC)		ALIGNED READS	
SAMPLE ID	**TOTAL NUMBER OF READS**	**NUMBER**	**% OF TOTAL READS**	**NUMBER**	**% OF READS PASSING QC**
C	54,059,655	26,437,774	48.90	19,176,330	72.53
UT	57,062,801	27,408,797	48.03	20,404,290	74.44
T1	49,420,131	28,292,871	57.25	20,323,138	71.83
T2	53,384,390	25,640,885	48.03	17,913,467	69.86

In the subsequent analysis only unique reads (aligned to only one gene) and only genes showing a coverage greater than 50% of gene length were considered. The resulting total number of genes is 10545. The raw expression value of a gene in a sample, also called as row count, is considered as the number of unique reads which align to that gene obtained from that sample. In Figure 4.3 the distribution of raw gene expression values is shown. As it can be observed, values show a normal distribution it follows a normal distribution.

Figure 4.3. Distribution of raw gene expression values for each sample. The graphs were produced by the DEGSeq software.



4.1.2.3 Identification of differentially expressed genes (DEGs)

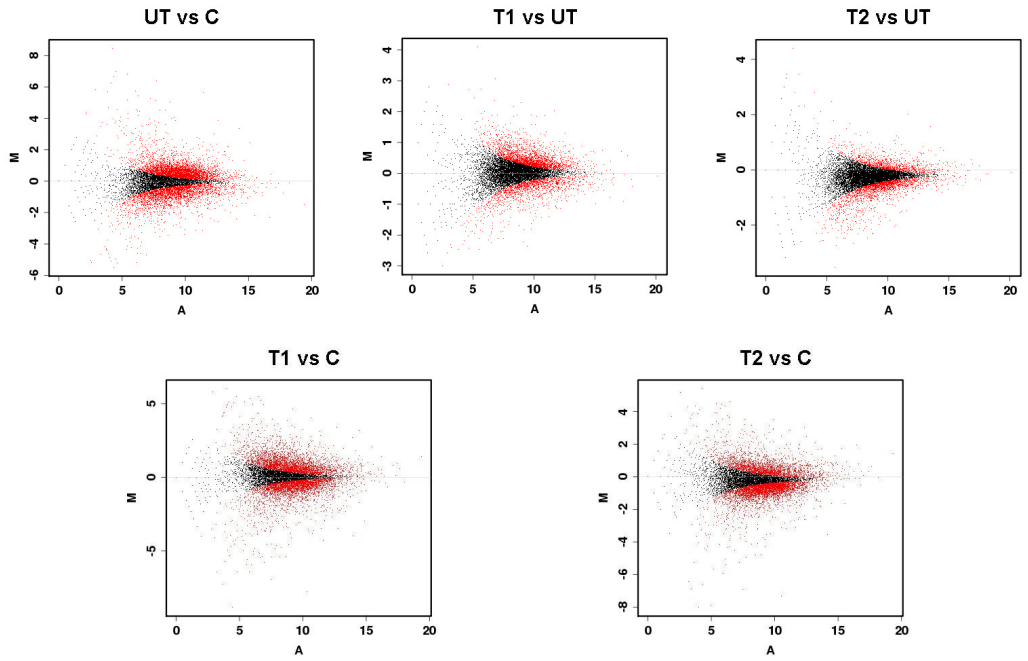
Starting from raw expression data, different comparisons between the 4 samples were considered. UT vs C comparison was used to analyse, at cellular level, alterations induced by the pathological condition compared to healthy one, T1 vs UT and T2 vs UT were analyzed for studying variations induced in Hunter cells by therapeutic enzyme supply. The results of these 3 comparisons will be reported in this first section of the chapter. The comparisons T1 vs C and T2 vs C were analyzed for the research of biomarkers of ERT response and will be taken into account in the third section of the chapter.

For each comparison the software DEGSeq was used to identify differentially expressed genes (DEGs), setting as criteria of differential expression $|\log_2(\text{ratio})| \geq 1$ and p-value$\leq$0.001, were ratio is intended between the row counts of the two samples to be compared and p-value is relative to the statistical analysis performed by the software. In other words, in each comparison were considered as DEGs the genes showing at least a 2-fold change between the two samples with a statistical significance.

In Figure 4.4 the MA-plots relative to each comparison are shown. MA-plots are used to study dependences of gene expression ratios between two samples from raw gene expression values. For each gene, MA-plot reports in Y-axis the log2 ratio between the raw counts of the two samples and in the X-axis the average log2 row count of the two samples. DEGs are shown

in red. As expected for an unbiased distribution, the graphs appear as a point cloud scattered about the M= 0 axis, excluding the need to further normalization in addition to that performed by DEGSeq software.

Figure 4.4. MA plots of each comparison. Each point represents a gene with DEGs shown in red. $M=\log_2 E_1 - \log_2 E_2$, $A=(\log_2 E_1+\log_2 E_2)$ for the comparison E_1 vs E_2. The graphs are produced by DEGSeq software.



In Table 4.3 the number of DEGs obtained for each comparison is listed, together with the subdivision in up and down regulated genes. Noteworthy, according to DEG numbers, gene expression variation due to the pathologic condition seem to be about 5-fold greater than that due to treatment.

Table 4.3. Differentially expressed genes (DEGs) obtained for each comparison by using DEGSeq software.

COMPARISON	DEGs	UP REGULATED GENES	DOWN REGULATED GENES
UT vs C	1656	871	785
T1 vs UT	339	173	166
T2 vs UT	301	60	241
T1 vs C	1955	1070	885
T2 vs C	1427	591	836

4.2 Study of cellular pathogenesis

In order to highlight cellular processes and functions altered in the MPS II pathology, the comparison UT vs C (MPS II cells vs Healthy cells) has been analyzed. The lists of DEGs or, in some cases gene expression data of all aligned genes, underwent to functional, pathway and promoter analysis.

4.2.1 Functional analysis

Functional analysis was performed by using different tools which, although based on different annotation systems and different statistics, all performs an enrichment analysis. Enrichment is a measure of the probability that the gene list under examination contains more members belonging to a given biological category than those expected by chance, taking into account the total number of members of the category and the number of genes in the list. The result of each analysis will be reported as a table, or its graphical visualization, of the categories most represented in the list with the indication of number of genes of the list associated to that category and/or the p-value resulting from the statistical test.
These results will be presented as divided according to the annotation system the tools used are based on. Result obtained using the most popular annotation system, the Gene Ontology, will be reported as first, followed by those obtained with other annotation systems.

4.2.1.1 Gene Ontology (GO)

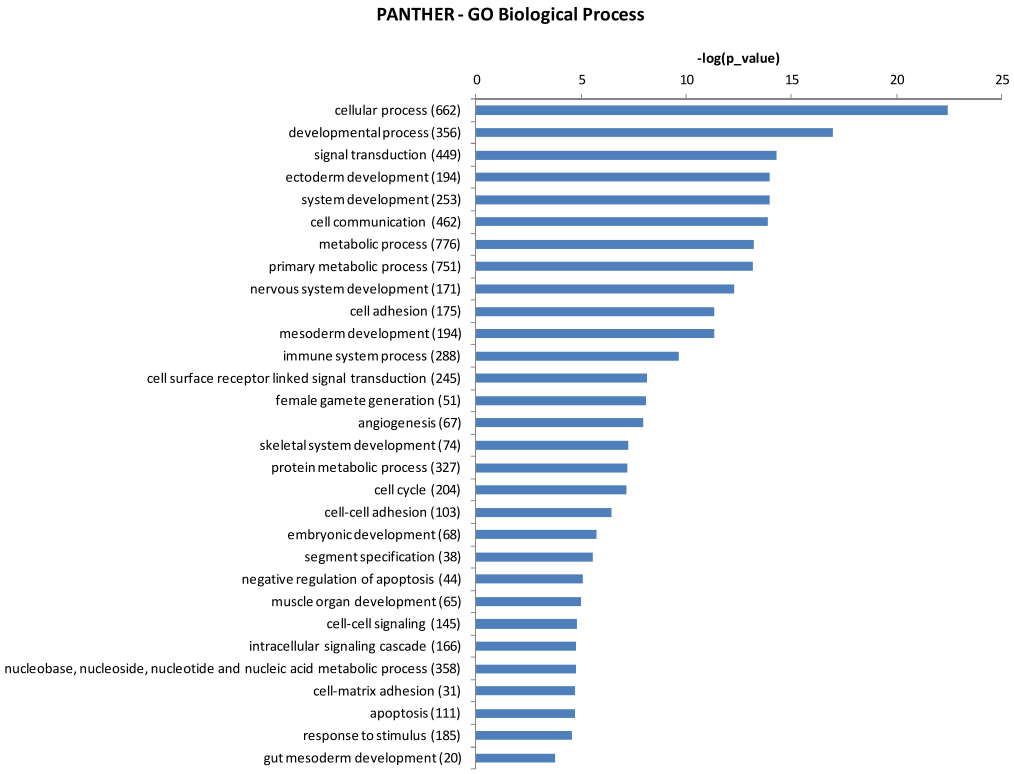
For GO analysis, the tools DAVID (Database for Annotation, Visualization and Integrated Discovery), Panther (Protein ANalysis THrough Evolutionary Relationships) and BiNGO (Biological Networks Gene Ontology tool) were used.
GO annotation system allows to classify genes and their product according to a hierarchy, based on 3 domains *Biological process*, *Molecular function* and *Cellular component.*

4.2.1.1.1 Biological process domain

The analysis of the *Biological process* domain conducted by means of the PANTHER software has identified 62 terms significantly enriched (p-value<0.05). The first 30 are shown in Figure 4.5. They include many basic cellular processes such as intra- and inter- cellular signalling

(signal transduction, intracellular signalling cascade, cell communication, cell-cell signalling), cell and cell-matrix adhesion, apoptosis, cell cycle and metabolism (protein and nucleic acid metabolism), as well as developmental processes (development of ectoderm, mesoderm, embryo, nervous system, muscle organ, skeletal system and segment specification and angiogenesis). Also response to stimulus and immune system process have been evidenced.

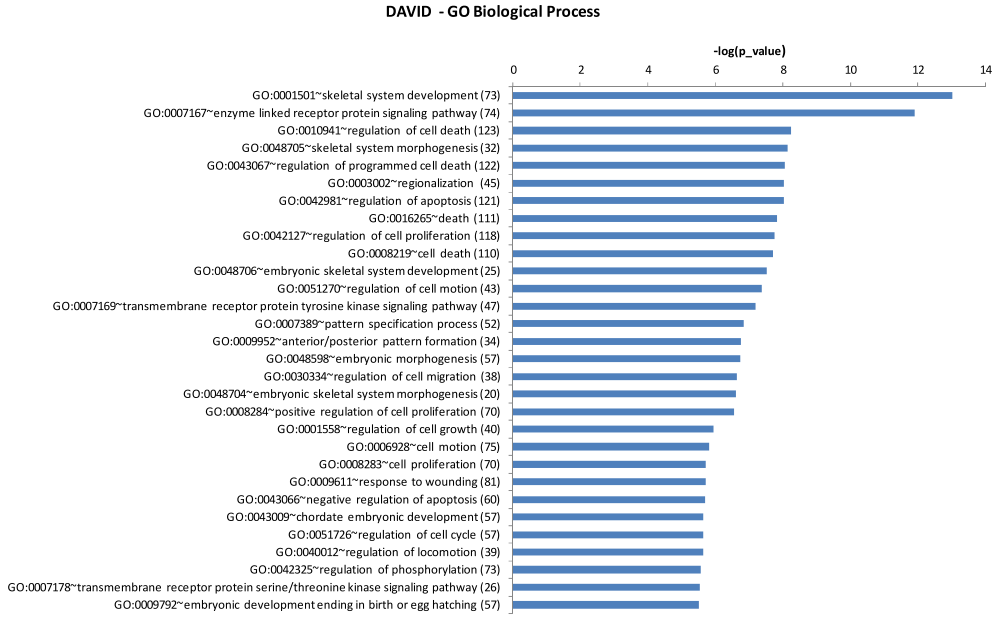
Figure 4.5. Enrichment analysis of DEGs in UT vs C comparison for the GO domain *Biological Process* obtained by PANTHER software. The first 30 significantly enriched categories are shown (p-value<0.05). The number of genes associated to each category is reported in brackets.



Regards to the DAVID tool, using the annotation GOTERM_BP_FAT, 368 categories resulted as statistically more enriched (p-value<0.05). In Figure 4.6 the first 30 categories are listed. They include terms associated to developmental processes ("skeletal system development", "skeletal system morphogenesis", "anterior/posterior pattern formation", "chordate embryonic development", "regionalization"), cell cycle and proliferation ("regulation of cell cycle", "regulation of cell proliferation", "regulation of cell growth", "regulation of cell death"), cell motility ("cell motion", "regulation of cell migration") and cell signalling ("enzyme linked receptor protein signalling pathway", "transmembrane receptor protein tyrosine kinase

signalling pathway" "transmembrane receptor protein serine/threonine kinase signalling pathway"). Also the term response to wounding has been found.

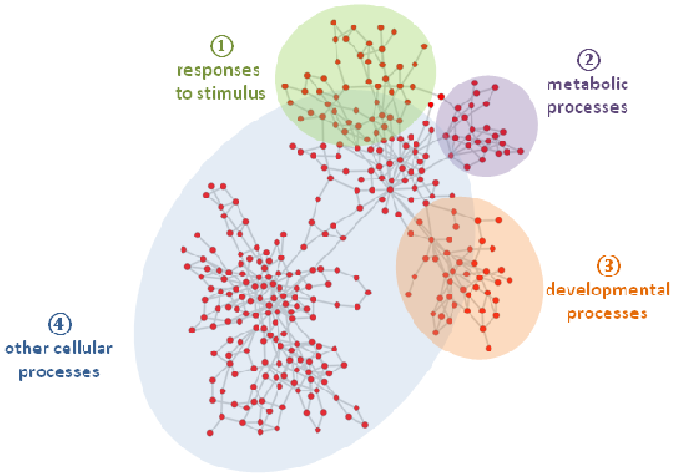
Figure 4.6. Enrichment analysis of DEGs in UT vs C comparison for the GO domain *Biological Process* obtained by DAVID tool. The first 30 significantly enriched categories are shown (p-value<0.05). The number of genes associated to each category is reported in brackets.



GO annotation was explored also by using the software BiNGO, performing a slightly different analysis. This software is implemented as a plug-in of Cytoscape, a platform for visualizing and integrating molecular interaction networks. It performs an enrichment analysis, and provides as output a Cytoscape graph, helping the visualization and interpretation of the result. In each graph nodes represent the GO categories that were found significantly over-represented in the GO hierarchy. The size (intended as area) of the nodes is proportional to the number of genes in the test set which are annotated to that node. The colour of the node represents the (corrected) p-value, with darker colour tone indicating lower, and then more significant, values. White nodes are not significantly over-represented, the other ones are, with a colour scale ranging from the lightest colour tone (p-value = significance level, in this case 0.05) to the darkest one (p-value = 5 orders of magnitude smaller than significance level, in this case $5*10^{-7}$). The list of DEGs obtained from the UT vs C comparison was submitted to BiNGO and analyzed for the GO domain *Biological Process;* the resulting graph is shown in Figure 4.7 as a red network, at low magnification. The graph layout was manually adjusted in such a way that

4 sub-networks, corresponding to more restricted branches of GO hierarchy, could be identified and extracted.

Figure 4.7. Graph of enrichment analysis of DEGs in UT vs C comparison for the GO domain *Biological Process* obtained by BiNGO.



A name was assigned to each sub-network: (1) responses to stimulus, (2) metabolic processes, (3) developmental processes, (4) other cellular processes. Then they were separately analyzed. Figure 4.8 shows the sub-network "responses to stimulus"; the corresponding significantly over-represented GO terms are listed in Table 4.4. As seen in the graph, the overrepresented terms include cellular response to stress, comprising oxidative stress and wound healing, response to organic substances (lipid, cytokine and insulin) and immune response, in particular neutrophil-mediated immunity.

Figure 4.8 Sub-networks responses to stimulus.

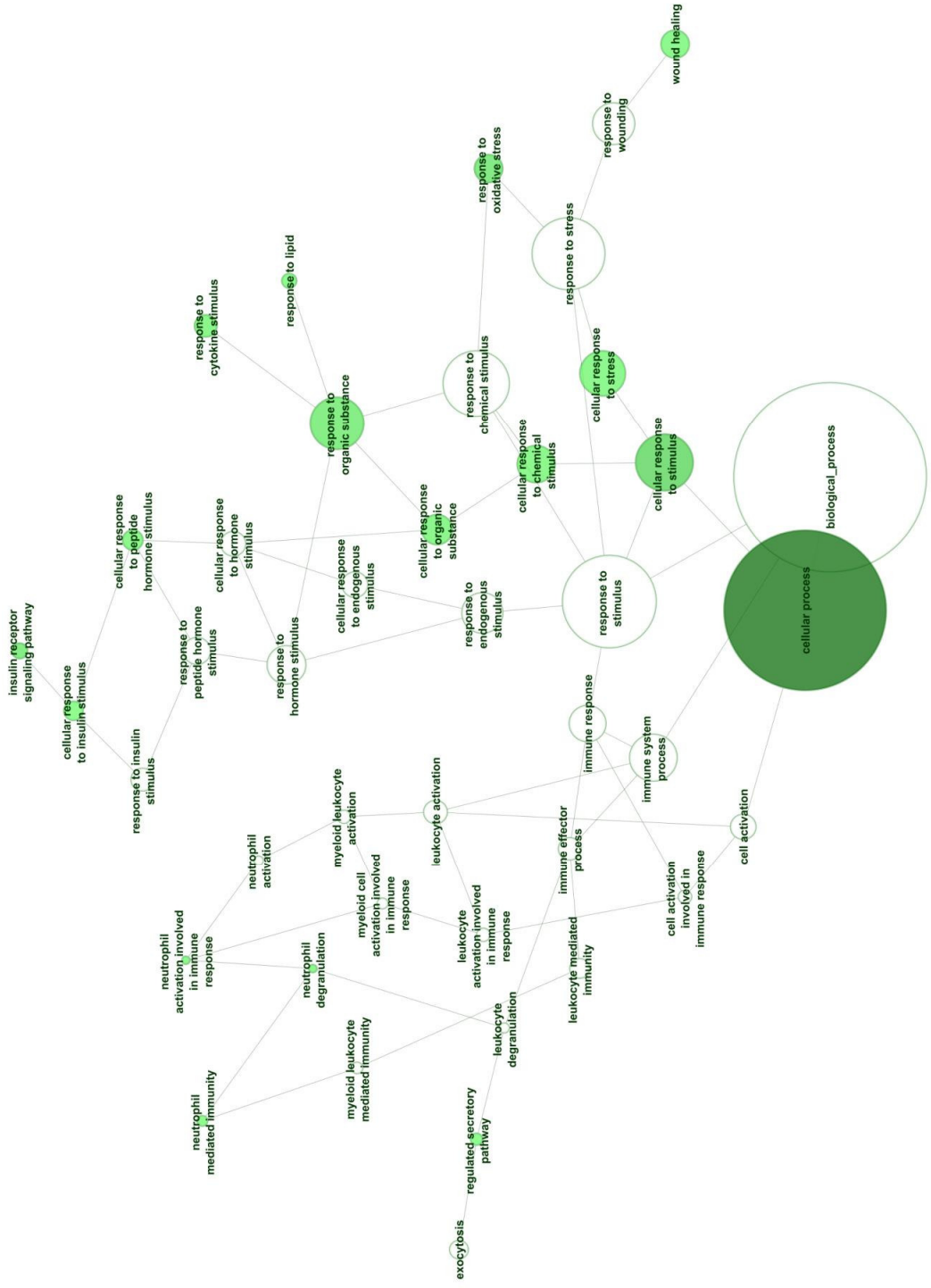


Table 4.4. Enrichment analysis of DEGs in UT vs C comparison for the GO domain *Biological Process* on sub-network "response to stimulus". Only significantly over-represented GO terms are shown (p-value<0.05).

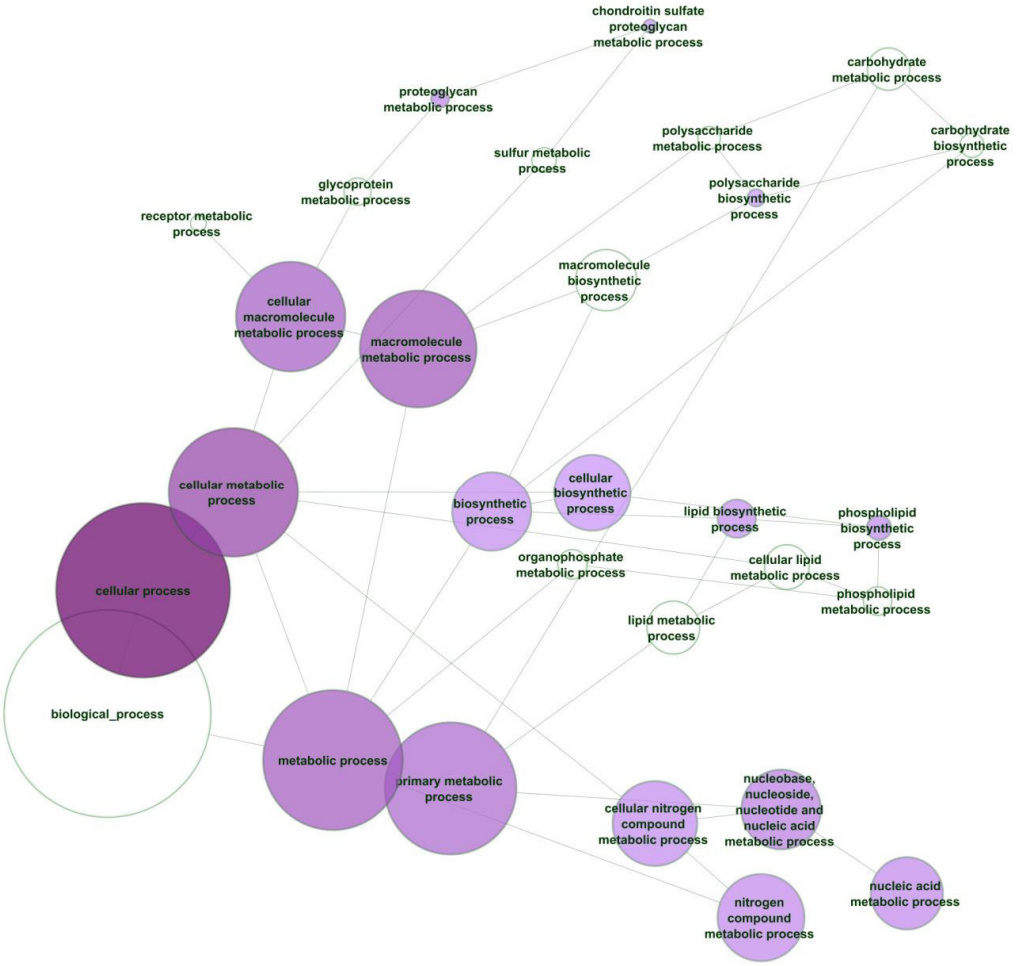
Description	adjusted p-value
cellular response to stimulus	1.42E-03
response to oxidative stress	5.74E-03
insulin receptor signaling pathway	7.66E-03
response to organic substance	8.14E-03
response to cytokine stimulus	1.93E-02
cellular response to peptide hormone stimulus	2.29E-02
cellular response to organic substance	2.30E-02
cellular response to chemical stimulus	2.66E-02
cellular response to stress	2.66E-02
wound healing	2.66E-02
neutrophil activation involved in immune response	2.81E-02
neutrophil degranulation	2.81E-02
response to lipid	4.05E-02
regulated secretory pathway	4.31E-02
cellular response to insulin stimulus	4.37E-02
neutrophil mediated immunity	4.47E-02

Figure 4.9 shows the sub-network "metabolic processes" with the corresponding significantly over-represented GO terms listed in Table 4.5. They include proteoglycan metabolism, in particular chondroitin sulfate metabolism, polysaccharide biosynthesis, lipid (mainly phospholipid biosynthesis) and nucleic acid metabolism.

Table 4.5. Enrichment analysis of DEGs in UT vs C comparison for the GO domain Biological Process on sub-network "metabolic processes". Only significantly over-represented GO terms are shown (p-value<0.05).

Description	adjusted p-value
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	5.59E-03
proteoglycan metabolic process	1.02E-02
nucleic acid metabolic process	1.23E-02
nitrogen compound metabolic process	1.25E-02
phospholipid biosynthetic process	1.41E-02
cellular nitrogen compound metabolic process	1.50E-02
biosynthetic process	1.65E-02
lipid biosynthetic process	1.93E-02
chondroitin sulfate proteoglycan metabolic process	2.74E-02
cellular biosynthetic process	2.81E-02
polysaccharide biosynthetic process	4.47E-02

Figure 4.9 Sub-network metabolic processes.



The sub-network "developmental process" is reported as graphic representation in Figure 4.10 and as list of enriched terms in Table 4.6. It can be observed that terms associated with the development of many organs and systems are present: skeletal, nervous and urinary systems, heart and vasculature with the most significant representation of terms associated with skeletal system development. Among the terms linked to anatomical structure formation, anterior/posterior pattern formation and appendage morphogenesis are the most represented.

Figure 4.10 Sub-network developmental processes

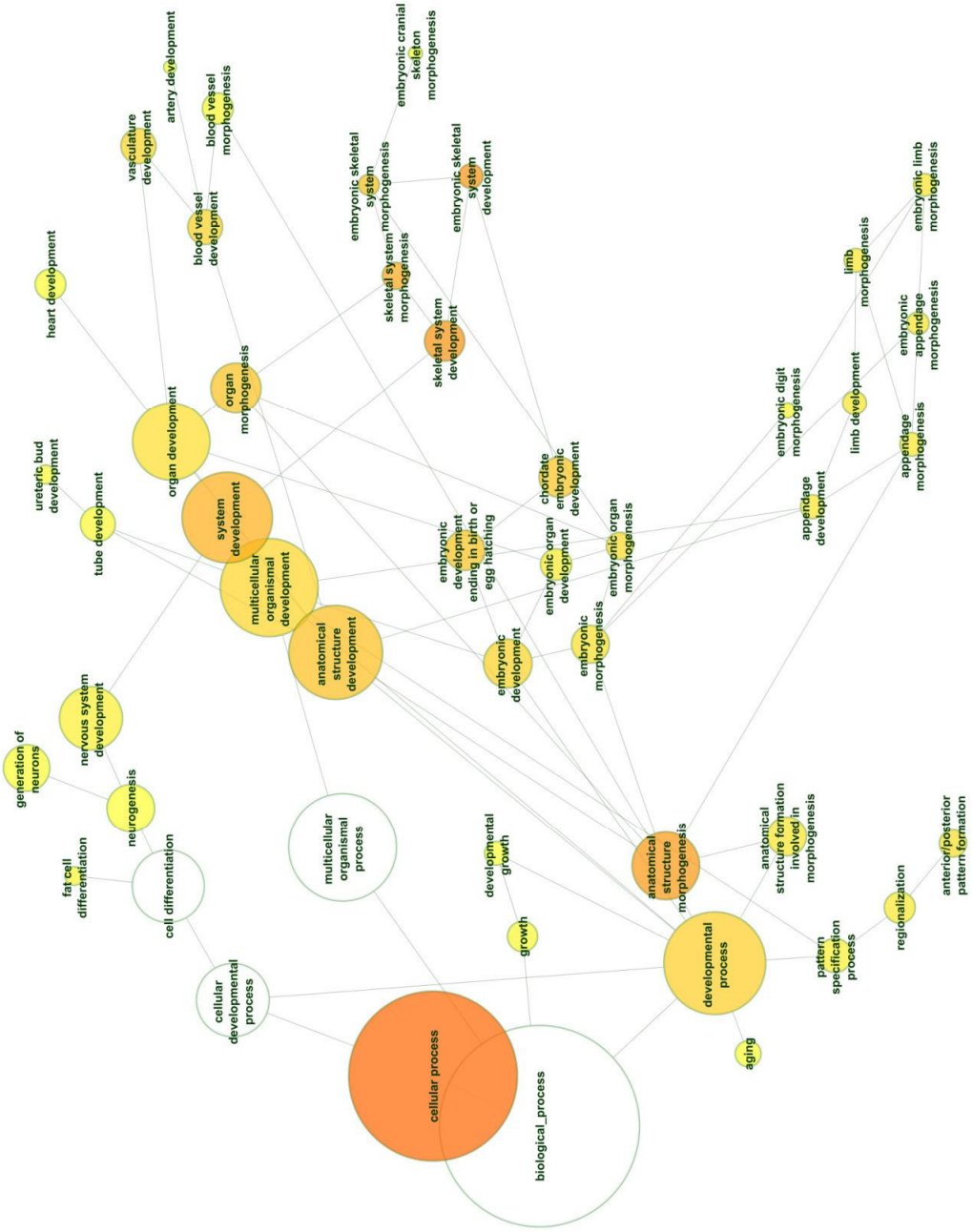


Table 4.6. Result of enrichment analysis of DEGs in UT vs C comparison for the GO domain Biological Process on sub-network "developmental processes". Only significantly over-represented GO terms are shown.

Description	adjusted p-value
skeletal system development	1.26E-05
embryonic skeletal system development	1.36E-05
anatomical structure morphogenesis	2.21E-05
skeletal system morphogenesis	6.64E-05
system development	1.13E-04
anatomical structure development	3.68E-04
embryonic skeletal system morphogenesis	4.03E-04
organ morphogenesis	4.38E-04
embryonic development ending in birth or egg hatching	5.47E-04
chordate embryonic development	5.68E-04
developmental process	1.34E-03
blood vessel development	1.42E-03
embryonic development	1.58E-03
multicellular organismal development	1.92E-03
vasculature development	1.97E-03
embryonic organ morphogenesis	2.89E-03
embryonic morphogenesis	4.23E-03
anterior/posterior pattern formation	4.27E-03
organ development	4.62E-03
embryonic limb morphogenesis	5.74E-03
appendage development	6.53E-03
limb development	6.53E-03
appendage morphogenesis	6.54E-03
limb morphogenesis	6.54E-03
anatomical structure formation involved in morphogenesis	6.54E-03
embryonic organ development	8.85E-03
regionalization	9.35E-03
pattern specification process	1.46E-02
nervous system development	1.46E-02
embryonic digit morphogenesis	1.84E-02
blood vessel morphogenesis	2.26E-02
tube development	2.61E-02
embryonic cranial skeleton morphogenesis	2.65E-02
growth	2.66E-02
fat cell differentiation	2.68E-02
ureteric bud development	2.71E-02
heart development	3.60E-02
developmental growth	4.19E-02
artery development	4.31E-02
aging	4.41E-02
generation of neurons	4.46E-02

For the last sub-network of BiNGO graph including "other cellular processes", Table 4.7 lists the over-represented terms. They include many basic cellular processes such as: cell death, cell cycle, cell proliferation, signal transduction, cell communication, cell differentiation, cellular component movement and organization, cell migration, cell matrix adhesion, protein metabolic processes.

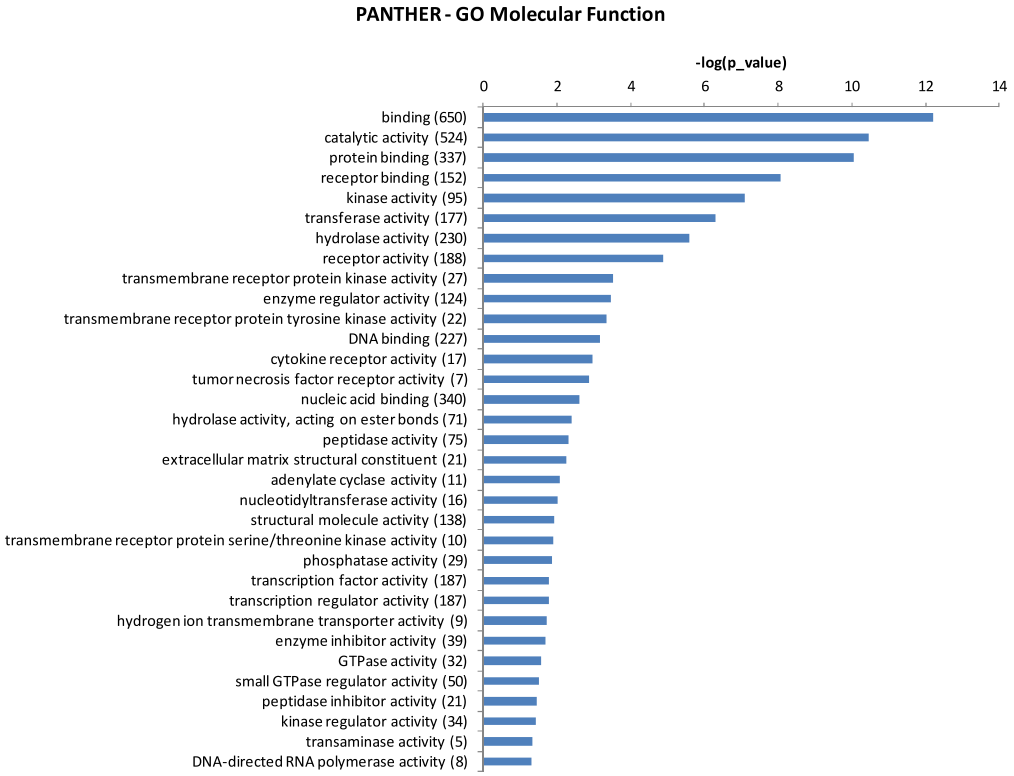
Table 4.7. Result of enrichment analysis of DEGs in UT vs C comparison for the GO domain Biological Process on sub-network "other processes". Only significantly over-represented GO terms are shown.

Description	adjusted p-value
cell death	9.21E-07
death	9.21E-07
regulation of signaling pathway	8.35E-06
regulation of programmed cell death	1.82E-05
regulation of cell death	2.07E-05
regulation of apoptosis	2.74E-05
regulation of cellular protein metabolic process	3.15E-05
regulation of protein metabolic process	3.33E-05
programmed cell death	4.24E-05
apoptosis	6.42E-05
regulation of developmental process	1.27E-04
negative regulation of apoptosis	1.27E-04
negative regulation of programmed cell death	1.30E-04
regulation of cell cycle	1.39E-04
enzyme linked receptor protein signaling pathway	1.39E-04
cell cycle	1.43E-04
biological regulation	2.67E-04
regulation of signal transduction	2.75E-04
negative regulation of cell death	2.90E-04
regulation of cell proliferation	2.90E-04
regulation of cell communication	3.13E-04
regulation of signaling process	3.68E-04
regulation of cell differentiation	4.29E-04
regulation of intracellular protein kinase cascade	5.47E-04
regulation of cellular component movement	5.68E-04
regulation of protein modification process	6.60E-04
positive regulation of signaling pathway	6.61E-04
positive regulation of cell communication	6.90E-04
negative regulation of metabolic process	7.33E-04
regulation of phosphorylation	7.61E-04
negative regulation of protein metabolic process	8.85E-04
regulation of phosphorus metabolic process	1.15E-03
regulation of phosphate metabolic process	1.15E-03
positive regulation of signal transduction	1.15E-03
negative regulation of macromolecule metabolic process	1.18E-03
positive regulation of signaling process	1.19E-03
cellular component organization	1.22E-03
developmental process	1.34E-03
negative regulation of phosphorylation	1.35E-03
cellular component movement	1.42E-03
regulation of cell migration	1.42E-03
negative regulation of cellular metabolic process	1.42E-03
regulation of protein amino acid phosphorylation	1.54E-03
intracellular signal transduction	1.58E-03
signal transmission via phosphorylation event	1.82E-03
regulation of epithelial cell proliferation	1.82E-03
intracellular protein kinase cascade	1.82E-03
cell-matrix adhesion	1.82E-03
multicellular organismal development	1.92E-03

4.2.1.1.2 Molecular Function domain

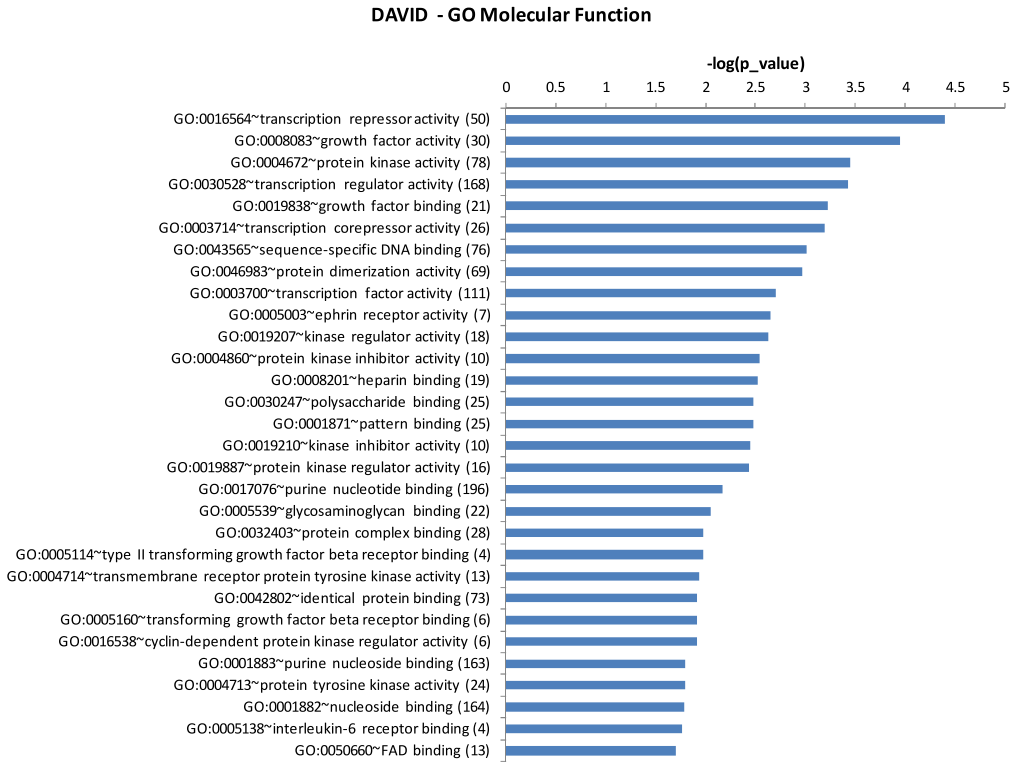
The analysis of the *Molecular Function* GO domain conducted by means of PANTHER software highlights 33 significantly enriched terms (p-value<0.05), showed in Figure 4.11. Many of them refers to receptor activity ("receptor binding", "receptor activity", "transmembrane receptor protein kinase activity", "transmembrane receptor protein tyrosine kinase activity", "cytokine receptor activity", "tumor necrosis factor receptor activity", "transmembrane receptor protein serine/threonine kinase activity", "hydrogen ion transmembrane transporter activity") and other key molecular functions in signal transduction ("kinase activity", "phosphatase activity", "enzyme inhibitor activity", "small GTPase regulator activity", "transcription regulator activity"). Also catabolic activities ("hydrolase activity, acting on ester bonds", "peptidase activity", "hydrolase activity") and structural functions ("structural molecule activity", "extracellular matrix structural constituent") are among the most enriched terms.

Figure 4.11. Enrichment analysis of DEGs in UT vs C comparison for the GO domain *Molecular Function* obtained by PANTHER software. Only the significantly enriched categories are shown (p-value<0.05). The number of genes associated to each category is reported in brackets.



The analogous analysis performed by DAVID tool, using the annotation GOTERM_MF_FAT, detected 55 significantly overrepresented terms (p-value<0.05), the first 30 of which are displayed in Figure 4.12. They comprise, as the previous evaluation, many functions associated to signal transduction; in particular binding to transforming growth factor beta and interelukin-6 receptors are here evidenced. Interestingly, among the most represented categories there are growth factor activity and binding, and glycosaminoglycan and heparin binding.

Figure 4.12. Enrichment analysis of DEGs in UT vs C comparison for the GO domain *Molecular Function* obtained by DAVID tool. The first 30 significantly enriched categories are shown (p-value<0.05). The number of genes associated to each category is reported in brackets.



4.2.1.1.3 Cellular Component domain

The analysis of the *Cellular Component* GO domain performed using PANTHER software evidenced 6 significantly enriched terms (p-value<0.05) (Figure 4.13). Although not much informative, this result underlines the important involvement of extracellular region and matrix in MPS II pathology. Moreover, also the term proton-transporting ATP synthase complex has been detected.

The extracellular space and matrix were highlighted among the GO *Cellular Component* terms also by DAVID tool, together with the other terms indicated in Figure 4.14. Also plasma membrane is present as external side, membrane proteins and receptor complexes such as those of integrin, interleukin-6 and insulin-like growth factor binding protein. However all the endomembrane system resulted involved as vesicles, both as luminal and cytoplasmic part, organelle membrane and Golgi apparatus.

Figure 4.13. Enrichment analysis of DEGs in UT vs C comparison for the GO domain *Cellular Component* obtained by PANTHER software. Only the significantly enriched categories are shown (p-value<0.05). The number of genes associated to each category is reported in brackets.

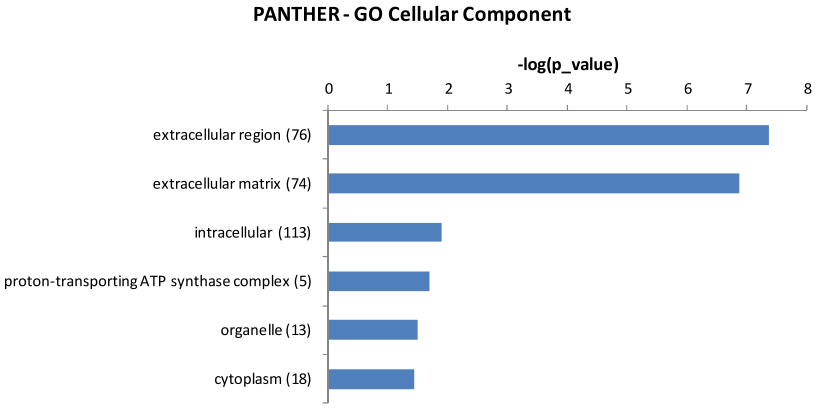
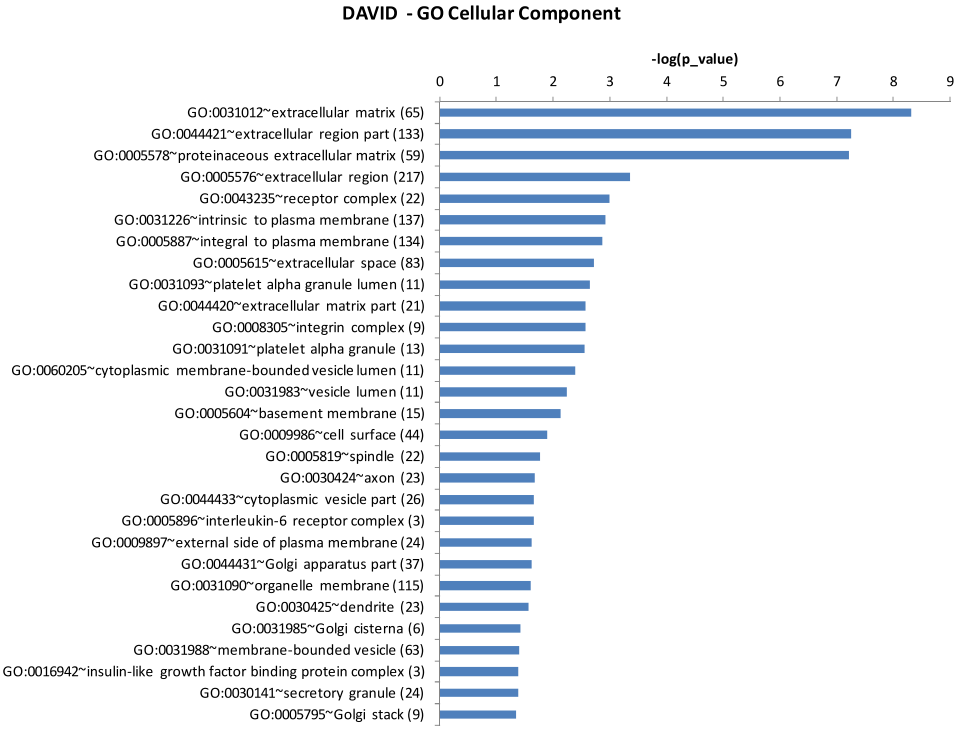


Figure 4.14. Enrichment analysis of DEGs in UT vs C comparison for the GO domain Cellular Component obtained by PANTHER software. Only the significantly enriched categories are shown (p-value<0.05). The number of genes associated to each category is reported in brackets.

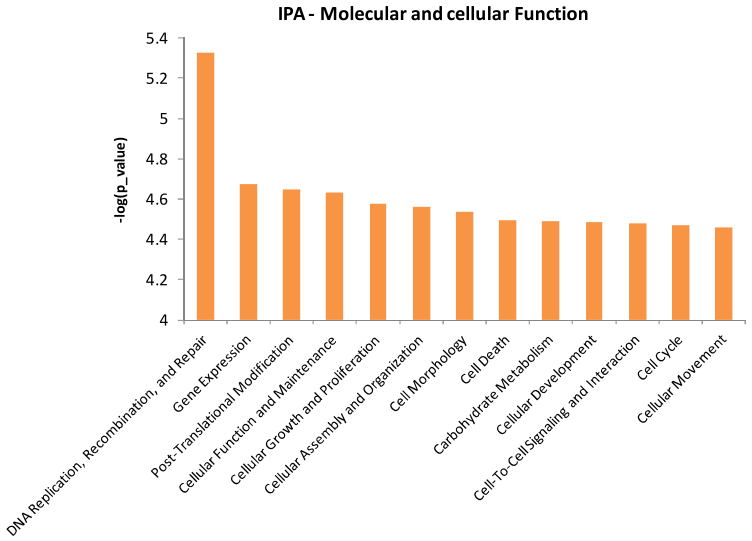


4.2.1.2 Other annotation systems

In addition to Gene Ontology, other annotation systems were used to study gene expression variations in the comparison between MPS II vs healthy cells.

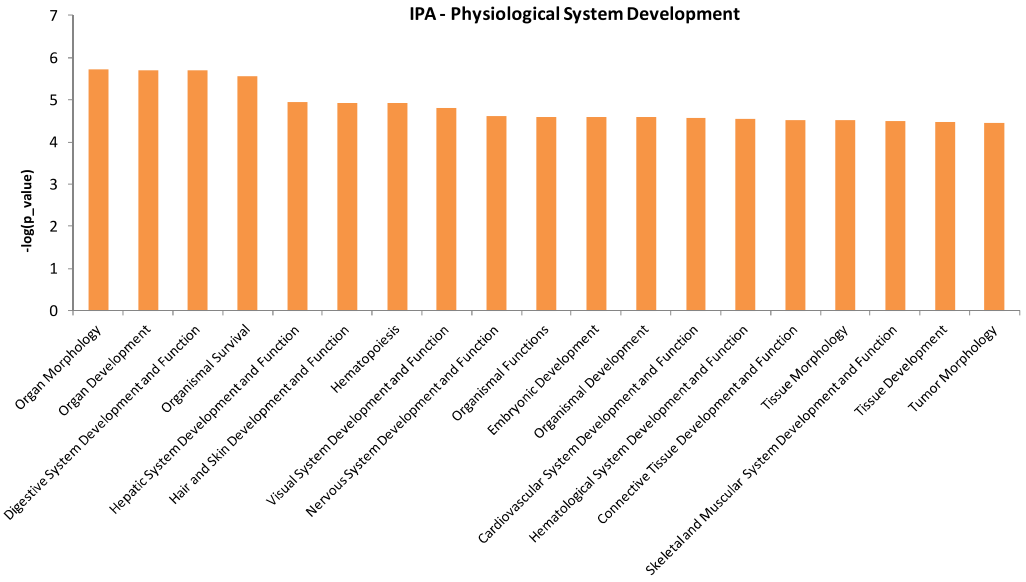
IPA (Ingenuity® Pathways Analysis) allows to perform a functional analysis, based on 3 categories: *Molecular and Cellular Functions*, *Physiological System Development and Functions* and *Diseases and Disorders.* The over-represented terms for these classifications are reported in Figure 4.15, Figure 4.16 and Figure 4.17, respectively.

Figure 4.15. Enrichment analysis of UT vs C comparison for the category *Molecular and Cellular Functions* of IPA annotation. Only the significantly enriched categories are shown (p-value<0.05).



Molecular and Cellular Functions category (Figure 4.15) identified as over-represented terms associated with DNA metabolism including replication, recombination, repair and gene expression. It also confirmed the involvement of basic cellular functions found in the previous evaluation based on Gene Ontology. They comprise cellular growth and proliferation, cell cycle, cell morphology, cellular organization, cell movement, cell to cell interactions and cellular development. Also post-translational modification and carbohydrate metabolism were found among the enriched terms.

Figure 4.16. Enrichment analysis of UT vs C comparison for the category *Physiological System Development* of IPA annotation. Only the significantly enriched categories are shown (p-value<0.05).



The evaluation of the *Physiological System Development and Functions* (Figure 4.16) category confirmed the previous observation on developmental processes in the analysis performed by BiNGO. Development seems to be altered at any level, both on embryonic and organismal stage and seems to involve many different tissues, organs and systems: digestive, nervous, cardiovascular, skeletal, muscular and visual systems, liver, connective tissue, hair, skin and blood.

Enrichment analysis of the category *Diseases and Disorders* (Figure 4.17) evidenced terms associated with disease affecting all tissues and organs reflecting the multiorgan and multisystem nature of Hunter syndrome.

Another annotation system used is that based on SP-PIR keywords implemented by DAVID tool.

Figure 4.18 shows the significantly enriched categories, which include extracellular matrix, cytoskeleton, secreted molecules, glycoprotein, chondroitin sulfate proteoglycan, heparin binding, lipoprotein and ubiquitin conjugation.

Figure 4.17. Enrichment analysis of UT vs C comparison for the category *Diseases and Disorders* of IPA annotation. Only the significantly enriched categories are shown (p-value<0.05).

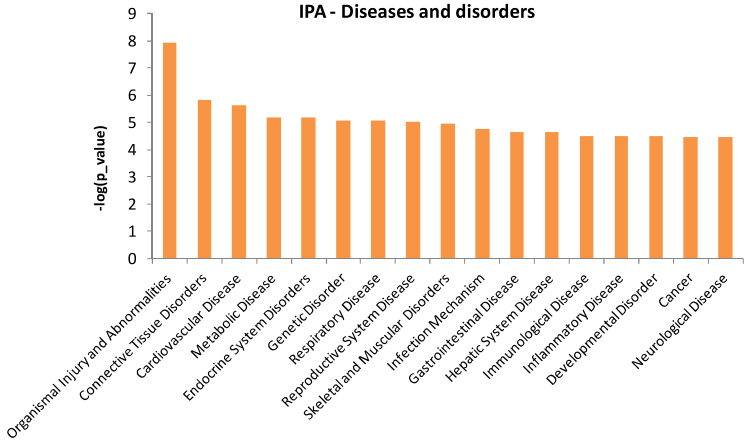
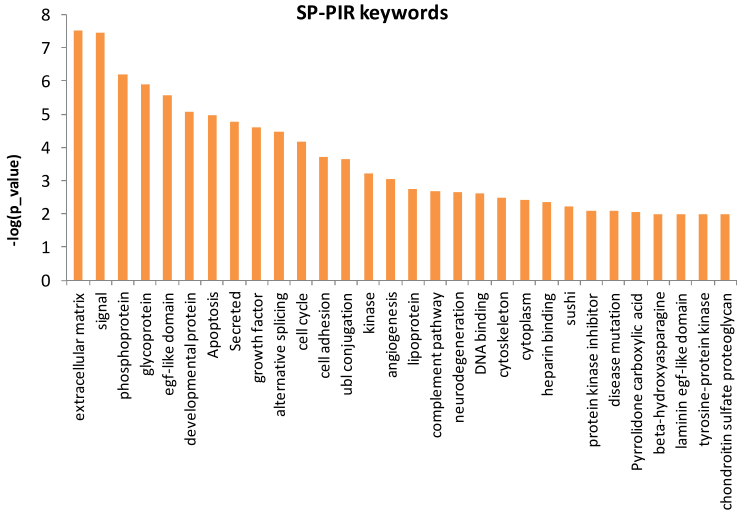


Figure 4.18. Enrichment analysis of UT vs C comparison for the category SP-PIR keywords. Only the significantly enriched categories are shown (p-value<0.05).



4.2.1.3 Pathway analysis

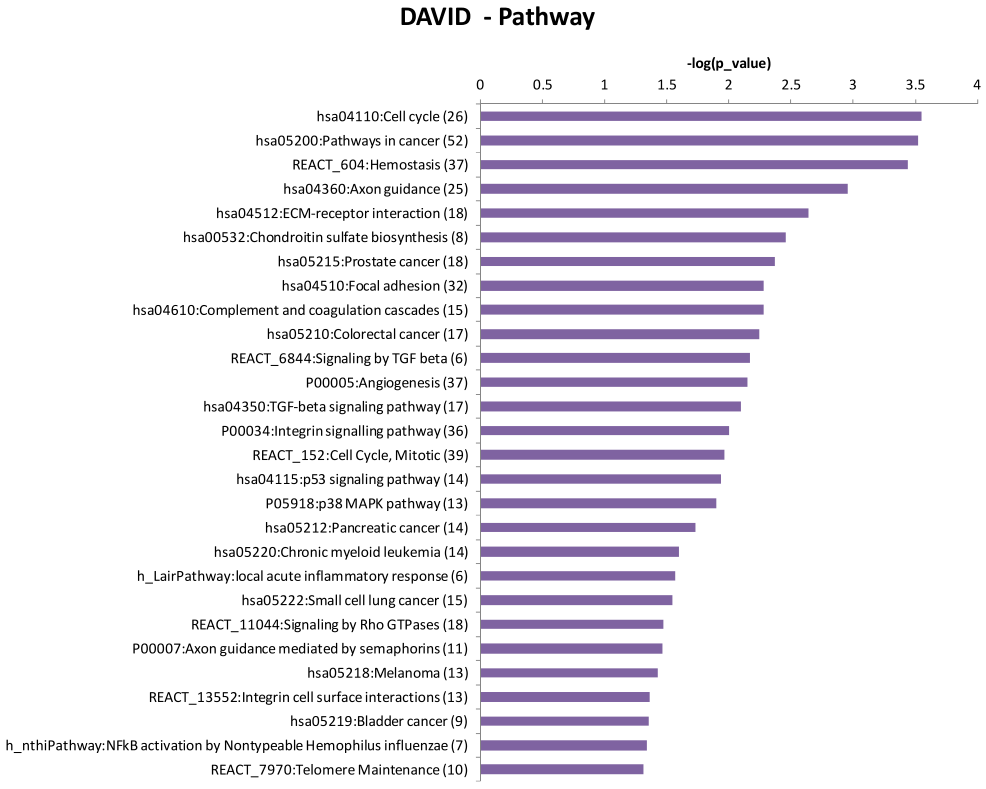
To give a less generalized insight on the pathologic alterations of the MPS II cells, Functional analysis was accompanied by a Pathway analysis. Similarly to functional evaluation, also in this

case an enrichment analysis was conducted with the difference that in this case categories to which gene lists are compared are pathways.

Shown below are the results of pathway analysis conducted by using the above mentioned tools DAVID (Figure 4.19) and PANTHER (Figure 4.20).

DAVID tool evidenced alteration in pathways involved in cell cycle and proliferation ("p38 MAPK"), cellular communication, ("focal adesion", "integrin signaling", "ECM receptor interaction"), immune response, such as complement and coagulation cascades, developmental processes (TGFβ signaling, axon guidance) and structural molecules biosynthesis (chondroitin sulfate).

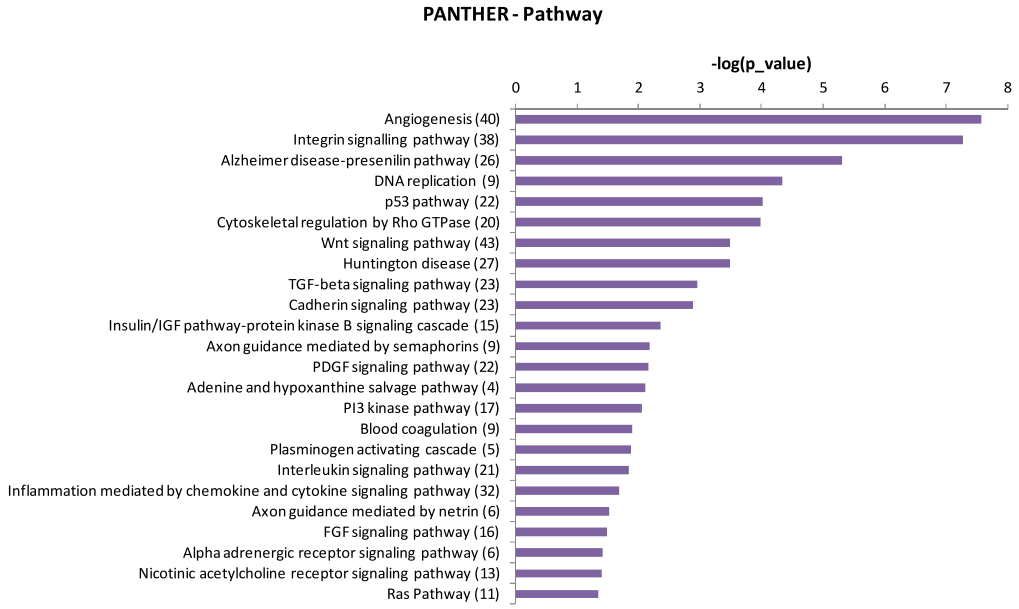
Figure 4.19. Pathway analysis of DEGs in UT vs C comparison obtained by DAVID tool. The annotation database Biocarta, KEGG and Reactome were selected. Only significantly enriched categories are shown (p-value<0.05). The number of genes associated to each category is reported in brackets.



Pathway analysis performed by PANTHER tool confirms what evidenced with DAVID, adding other important pathways involved in the same processes. Among these there are: "cytoskeletal regulation" and "cadherin signaling" implicated in cell-cell comunication and structure, "inflammation mediated by chemokine and cytokine" and "interleukin signaling" for

immune response, "Wtn signaling", "insulin/IGF pathway-protein kinase B", "PDGF signaling" and "FGF signaling" for growth factors mainly involved in development.

Figure 4.20 Pathway analysis of DEGs in UT vs C comparison obtained by PANTHER tool. The annotation database Biocarta, KEGG and Reactome were selected. Only significantly enriched categories are shown (p-value<0.05). The number of genes associated to each category is reported in brackets.



4.2.1.4 Promoter analysis

In addition to functional and pathway analysis of DEGs which are based on known functional information about genes, also promoter analysis has been conducted. This is based on a different type of information obtained from nucleotide sequence of gene regulatory regions. Starting from the analysis of these sequences, transcription factors binding sites were identified as well as putative transcription factors able to induce the observed gene expression patterns. Then the analysis has been taken further in order to gain insights into the key upstream signaling regulators influencing the activity of these transcription factors (TF). Finally a pathway analysis was performed on the obtained lists of putative transcription factors together with the upstream regulators in order to identify signaling pathway which could be responsible for the observed gene expression variations.

Coming from a different type of information, the results of this pathway analysis could be used to corroborate the hypothesis formulated from functional analysis.

Promoter analysis was performed using Explain™ data analysis system on DEGs of the UT vs C comparison, considering up- and down- regulated gene separately.

Table 4.8 shows all the TFs obtained from analysis of regulatory regions of up-regulated genes. With this approach the alteration observed in functional and pathway analysis were confirmed. Many of them are involved in cell-cycle regulation and cellular differentiation (SP1-4, ARID5B, EGR1-3) as well as immune response regulation (NFKB1, REL, ZEB1). Moreover, molecules activated as response to stress and inflammation, such as MAP2K6, or involved in development (BMP2) and apoptosis (CASP6 and XIAP) have been identified as key upstream molecules. The key node upstream molecules are shown in Table 4.9.

Table 4.8 TFs obtained from analysis of regulatory regions of up-regulated genes (only TFs with p-value <0.0001).

TF acting on up-regulated genes				
TF	**p-value**		**TF**	**p-value**
SP1	3.05E-11		ZEB1	2.03E-05
SP2	3.05E-11		HIC1	5.19E-05
SP3	3.05E-11		**ARID5B**	1.42E-04
SP4	3.05E-11		ASCL1	1.42E-04
TFAP2A	3.08E-10		MYF5	1.42E-04
TFAP2B	3.08E-10		MYF6	1.42E-04
TFAP2C	3.08E-10		MYOD1	1.42E-04
TFCP2	3.70E-07		MYOG	1.42E-04
NFKB1	6.02E-06		TCF12	1.42E-04
NFKB2	6.02E-06		TCF3	1.42E-04
REL	6.02E-06		TCF4	1.42E-04
RELA	6.02E-06		ZBTB7A	1.53E-04
NFYA	6.90E-06		LMO2	1.85E-04
NFYB	6.90E-06		**EGR1**	4.20E-04
NFYC	6.90E-06		EGR2	4.20E-04
ZHX1	6.90E-06		EGR3	4.20E-04

Table 4.9 Key node upstream molecules (up regulated genes).

KEY NODE UPSTREAM MOLECULES (up regulated genes)				
Input	**Z-Score**		**Input**	**Z-Score**
ITGAM	2.30702		**MAP2K6**	1.59935
MAP3K7IP1	2.24258		PPM1A	1.58753
MAPK14	2.22993		MOS	1.58033
NFYB	2.08471		CDC25B	1.56214
ACVR2A	1.75466		CDC25C	1.56214
BMP2	1.75466		MAX	1.56214
BMPR1A	1.75466		PBK	1.56214

MAP3K7	1.75466		**CASP6**	1.55119
XIAP	1.75466		**DUSP2**	1.53218
EIF2AK2	1.69952		DUSP7	1.51750
SP3	1.69621		DUSP9	1.51750
PPM1B	1.63435		**EGR1**	1.50038

In Table 4.10 pathways associated with above mentioned TFs and the corresponding upstream molecules are listed.

Table 4.10 Pathway (up regulated genes). Only pathways with p-value<0.001

Pathway (up-regulated genes)	p-value
hsa04010:MAPK signaling pathway	5.11E-11
hsa04621:NOD-like receptor signaling pathway	2.09E-04
hsa04620:Toll-like receptor signaling pathway	1.35E-03
hsa04622:RIG-I-like receptor signaling pathway	4.81E-03
hsa05222:Small cell lung cancer	7.69E-03
hsa04914:Progesterone-mediated oocyte maturation	8.21E-03
hsa04350:TGF-beta signaling pathway	8.48E-03
hsa04210:Apoptosis	8.48E-03
hsa04660:T cell receptor signaling pathway	1.53E-02
hsa05200:Pathways in cancer	1.97E-02
hsa05120:Epithelial cell signaling in Helicobacter pylori infection	4.34E-02

Table 4.11 shows all the TFs obtained from analysis of regulatory regions of down-regulated genes, while in Table 4.12 and Table 4.13 the Key node upstream molecules and pathway are listed. As for up-regulated gene, the analysis of TFs and upstream molecules probably regulating down-regulated genes, has evidenced molecules involved in cell cycle, differentiation, adhesion and cell-cell interaction such as GATA2, FOXD1,CD47, ITGA5, ITGA6, ITGA8, ITGB1-4, BCAR1, ITGAV, GAB1 and PTPN2 and in immune and inflammatory response as CEBPD, STAT3-6, IRF1-8, OSMR, IL6ST, IL10RB, IL1R1, SOCS3, IFNGR1, IRAK1 and TYK2. Moreover, many growth factors are also present (IGF1-2, EGF, PDGFRA-B, IRS1-2 and NGF).

Table 4.11 TF acting on down-regulated genes.

TF acting on down-regulated genes				
Input	**P-value**		**Input**	**P-value**
STAT3	6.66E-09		IRF5	5.90E-06
STAT4	6.66E-09		**IRF6**	5.90E-06
STAT5A	6.66E-09		FOXG1	6.41E-06
STAT5B	6.66E-09		POU3F2	7.40E-06
STAT6	6.66E-09		POU2F1	8.63E-06
NANOG	9.80E-08		CUX1	1.68E-05

HLTF	1.69E-07		ZBTB16	2.57E-05
FOXC1	1.15E-06		POU1F1	3.22E-05
FOXD1	1.15E-06		IRF7	6.65E-05
FOXD3	1.15E-06		IRF8	6.65E-05
FOXF1	1.15E-06		CEBPA	7.92E-05
FOXF2	1.15E-06		CEBPB	7.92E-05
FOXH1	1.15E-06		HOXA9	8.60E-05
FOXI1	1.15E-06		**MEIS1**	8.60E-05
FOXJ1	1.15E-06		STAT1	9.66E-05
FOXJ2	1.15E-06		STAT2	9.66E-05
FOXL1	1.15E-06		ISGF3G	9.95E-05
FOXO3	1.15E-06		PDX1	1.14E-04
FOXO4	1.15E-06		**AR**	1.16E-04
CEBPD	1.18E-06		**PBX1**	1.25E-04
CEBPE	1.18E-06		GATA1	1.64E-04
CEBPG	1.18E-06		**GATA2**	1.64E-04
FOXO1	2.47E-06		GATA3	1.64E-04
IRF1	5.90E-06		GATA5	1.64E-04
IRF2	5.90E-06		GATA6	1.64E-04
IRF3	5.90E-06		GATA4	2.95E-04
IRF4	5.90E-06			

Table 4.12 Key node upstream molecules (down-regulated genes).

KEY NODE UPSTREAM MOLECULES (down-regulated genes)											
Input	**Z-Score**		**Input**	**Z-Score**		**Input**	**Z-Score**		**Input**	**Z-Score**
KIT	2.90843		IFNAR2	2.35673		SYK	2.06352		UBE2L3	1.86475
KITLG	2.90843		CD72	2.35031		**IGF1**	2.06271		TNFRSF11A	1.86047
IL2RG	2.88690		IL4R	2.35031		PDGFRB	2.05618		**PTEN**	1.85973
IL6ST	2.88690		PTPN6	2.35031		**GAB1**	2.03956		PTPRA	1.85443
JAK1	2.88690		PTPN1	2.33965		**IRAK1**	2.02768		HCK	1.83964
OSMR	2.88690		**SOCS3**	2.30281		**PDGFRA**	2.02634		PRLR	1.83052
CD47	2.88088		IFNA1	2.29324		CD22	2.01177		IGF1R	1.78046
STAT6	2.75943		**IL1R1**	2.26652		IFNG	1.99514		TEC	1.76254
IL7R	2.68417		VTN	2.26234		CBL	1.97523		**LYN**	1.74833
IL10RA	2.58169		HTR2A	2.23983		ITGA1	1.97049		MS4A2	1.74833
IL10RB	2.58169		POU5F1	2.22464		ITGA5	1.97049		VAV1	1.72516
IL22RA1	2.58169		IL1RAP	2.20841		**ITGA6**	1.97049		SPI1	1.72490
CSF2RB	2.49688		FES	2.20712		ITGA8	1.97049		ECSIT	1.71100
IL5RA	2.49688		PTPRC	2.19935		ITGAL	1.97049		MAP3K1	1.71100
KDR	2.49129		CEACAM1	2.16783		ITGB1	1.97049		RAPGEF1	1.69684
VEGFA	2.49129		ITGAV	2.15942		**ITGB2**	1.97049		CRKL	1.69669
VEGFC	2.49129		**ITGB3**	2.15942		ITGB4	1.97049		NEDD9	1.69281
EPHA4	2.48206		SIRPA	2.15608		SOCS1	1.95454		BMX	1.65092
GHR	2.48206		INSR	2.14810		IRS1	1.92325		**BCAR1**	1.63933
JAK2	2.48206		SRC	2.12104		IRS2	1.90196		PTPRJ	1.61239
LEPR	2.48206		PTPN11	2.11714		ITGA2B	1.89899		FYN	1.59988
EPO	2.46152		SIT1	2.11714		EDA2R	1.87934		CSK	1.58830
EPOR	2.46152		SLAMF1	2.11714		IRAK2	1.87934		FGA	1.58830
PTPN2	2.45927		IRF1	2.10739		NGFR	1.87934		FGB	1.58830
TYK2	2.44723		**IGF2**	2.10600		TNFAIP3	1.87934		FGG	1.58830
IFNGR1	2.43758		EGF	2.10161		TRAF6	1.87934		JAK3	1.54157
IFNGR2	2.43758		EGFR	2.10161		ZNF274	1.87934			
NOSIP	2.42517		STAT1	2.10161		**NGF**	1.86475			
PRL	2.35719		PTPN21	2.10072		NTRK1	1.86475			
IFNAR1	2.35673		POU2F2	2.07076		SQSTM1	1.86475			

Table 4.13 Pathway (down-regulated genes).

Pathway (down-regulated genes)	P-Value
hsa04010:MAPK signaling pathway	7.68E-06
hsa05200:Pathways in cancer	4.67E-05
hsa04620:Toll-like receptor signaling pathway	5.38E-05
hsa04630:Jak-STAT signaling pathway	5.68E-04
hsa05221:Acute myeloid leukemia	6.13E-04
hsa04622:RIG-I-like receptor signaling pathway	1.24E-02
hsa04914:Progesterone-mediated oocyte maturation	2.07E-02
hsa04062:Chemokine signaling pathway	3.81E-02

4.2.1.5 Conclusions

The transcriptome analysis performed on MPS II fibroblasts and the comparison with healthy cells has revealed an extensive gene expression alteration associated with MPS II pathology. The observed variations involve many, mostly essential, cellular processes confirming the very complex cellular scenario emerging in the last years for LSD pathology as well as the central role of lysosome in cell biology.

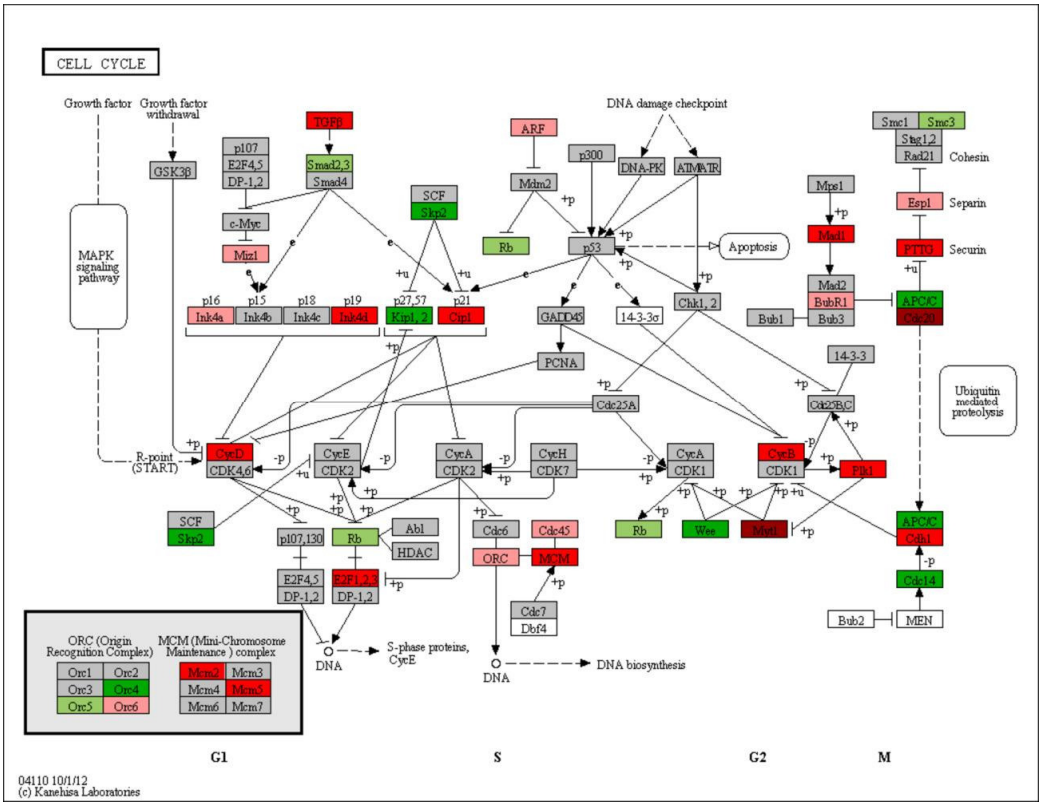
On the whole the variations resulted from this study could be summarized into alteration of:

1. basic cellular processes
2. metabolism
3. response to stimuli
4. developmental functions.

Basic cellular processes altered include cell cycle, apoptosis, ECM-cell interaction and actin-citoskeleton, whose pathways are shown in Figure 4.21, Figure 4.22, Figure 4.23, Figure 4.24. Color scale of the following maps is shown below:

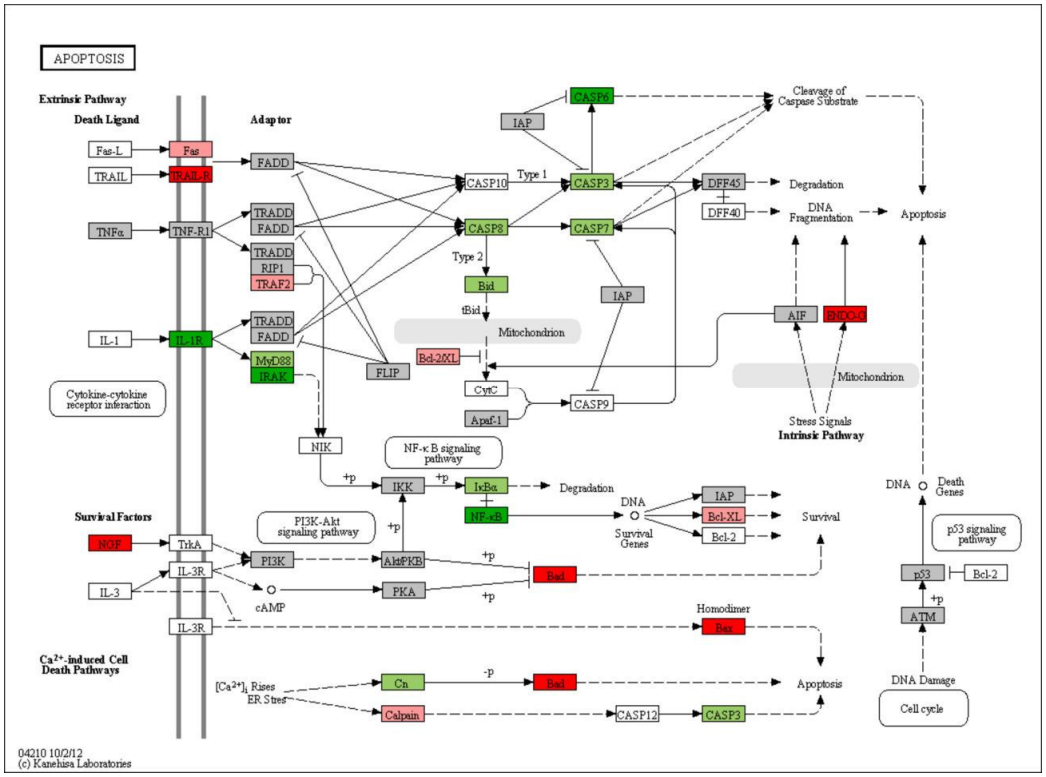
color scale	pvalue	log2ratio
	<0.001	>2
	<0.001	[2;1]
	<0.001	[1;0.7]
	<0.001	[-0.7;-1]
	<0.001	[-1; -2]
	<0.001	<-2
	other	
	not found	

Figure 4.21 KEGG Pathway Cell Cycle.



MPS II cells are interested by a wide alteration of the cell cycle, in particular CyclinB and Cyclin D are up-regulated. Moreover CDC20 that is required for two microtubule-dependent processes, nuclear movement prior to anaphase and chromosome separation, is also up-regulated.

Figure 4.22 KEGG Pathway Apoptosis.



Many genes involved in apoptosis result altered. Prolonged ER stress can result in the activation of BAD (log_2 ratio=1.50, p-value=2.22 E-15) leading to Ca^{++} cell death pathway. Also the intrinsic pathway was activated when various apoptotic stimuli trigger the release of cytochrome c from the mitochondria, as confirmed by the up-regulation of ENDO-G (log_2 ratio=1.54, p-value=1.27 E-08).

The interaction ECM-receptor (Fig 4.23) involves many structural protein such as fibronectin, OPN, Perlecan, Agrin and collagen which are up-regulated and laminin, CD47 and THBS which resulted down-regulated. To note Agrin, Perlecan and Sindecan are GAG containing-proteoglycan.

As for Aktin cytoskeleton, PFN1 (log_2 ratio=1.93, p-value=0) is among genes resulted as up-regulated; it encodes a member of the profilin family of small actin-binding proteins and plays an important role in actin dynamics by regulating actin polymerization in response to extracellular signals.

Figure 4.23 KEGG Pathway ECM-Receptor interaction.

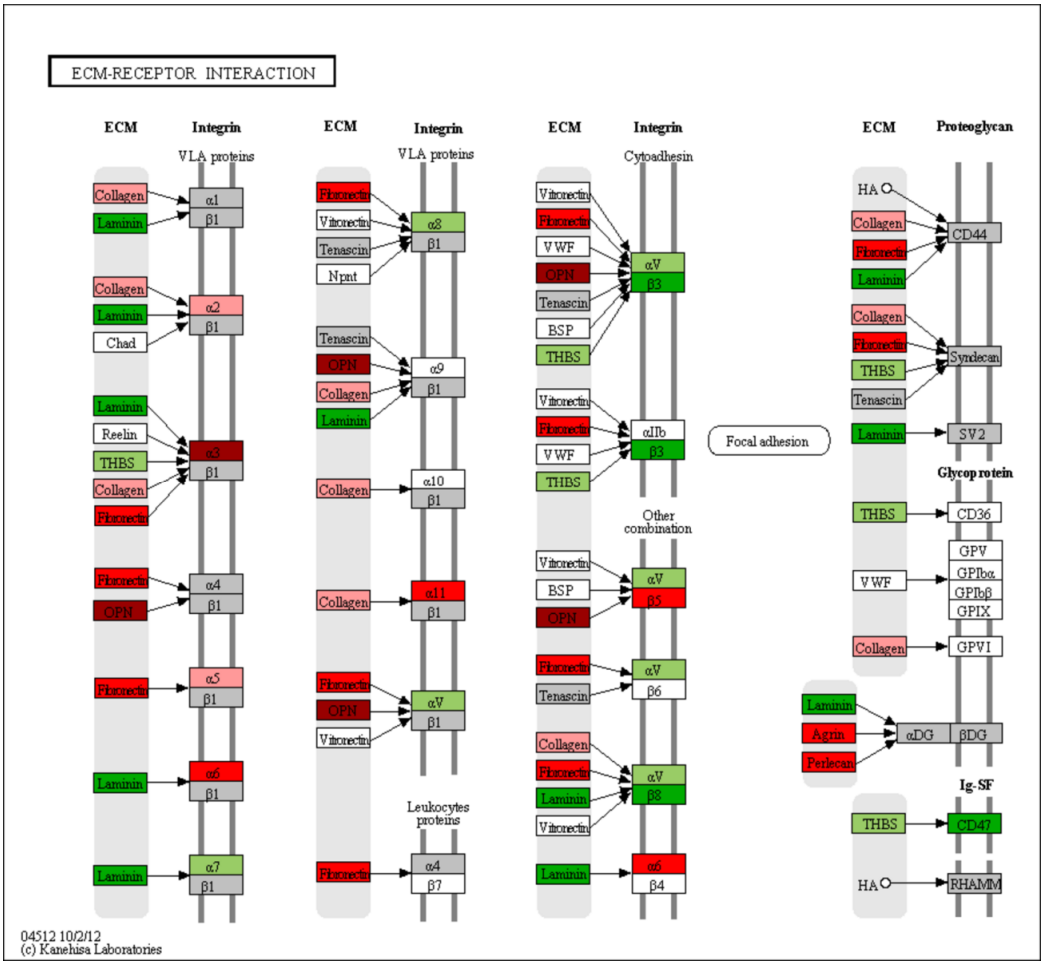
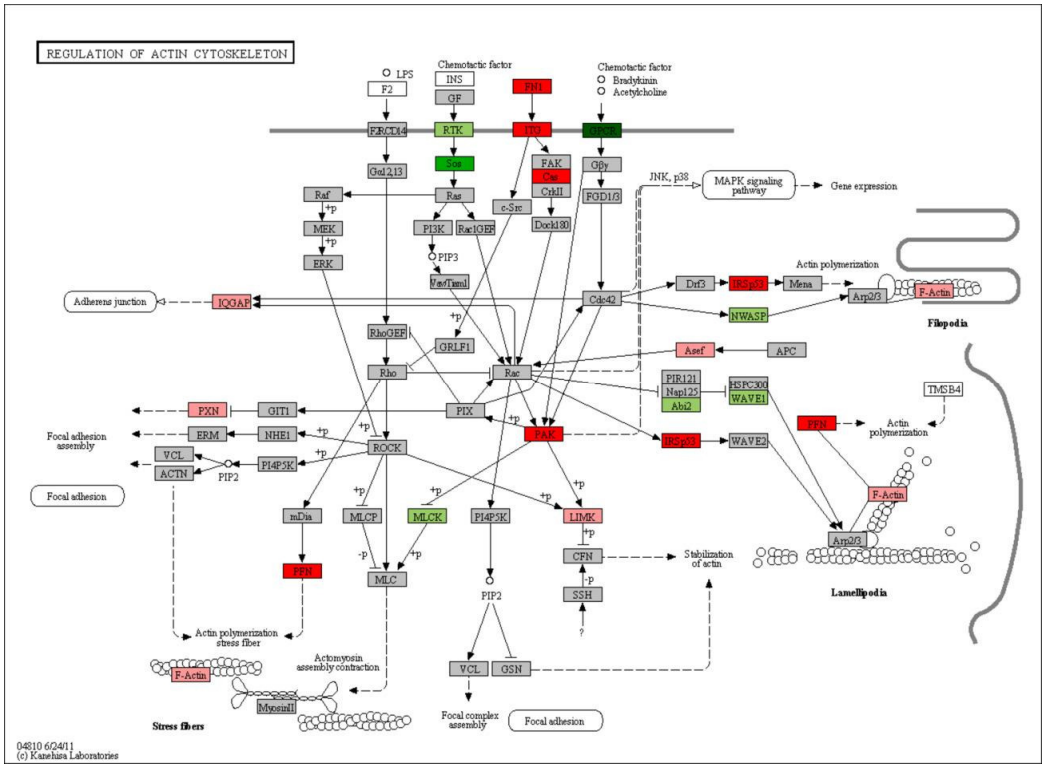


Figure 4.24 KEGG Pathway Regulation of Actin Cytoskeleton.

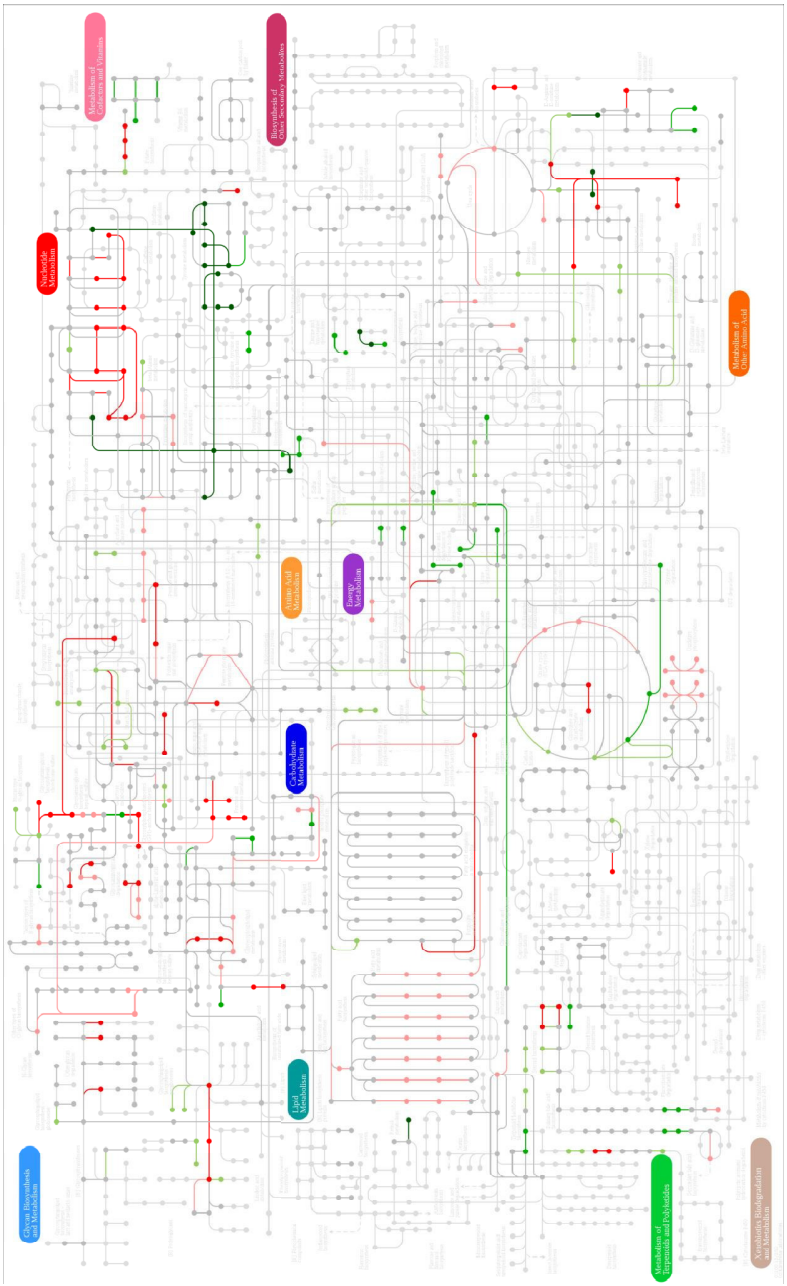


Also metabolic processes resulted altered at different levels. As seen by functional analysis (Figure 4.9, Table 4.5) main variations can be observed in polysaccharide, lipid and nucleotide metabolism. This observation is confirmed by mapping gene expression data on the global map of cell metabolism, as shown in Figure 4.25, which reveals also other alterations.

As for polysaccharide metabolism, BiNGO enrichment analysis (Figure 4.9, Table 4.5) reports, among the most represented terms, "proteoglycan metabolic process", "chondroitin sulfate proteoglycan metabolic process" and "polysaccharide biosynthetic process".

Alterations in polysaccharide, and in particular glycosaminoglycan (GAG) and proteoglycan, metabolism are expected considering that MPS II is due to a deficit of IDS, an enzyme of GAG degradation.

Figure 4.25 KEGG Pathway global metabolism



As shown in Figure 4.26 the enzymes catalyzing biochemical steps downstream of IDS resulted up-regulated in MPS II cells, probably as a part of a regulation mechanism sensitive to GAG concentration. Also enzymes of GAG synthesis resulted altered as shown in Figure 4.27, Figure 4.28, Figure 4.29 for heparan sulfate, chondroitin and dermatan sulfate and keratan sulfate, respectively. However it is difficult to understand which is the final effect of these variations.

Figure 4.26 KEGG Pathway Glycosamminoglycan degradation

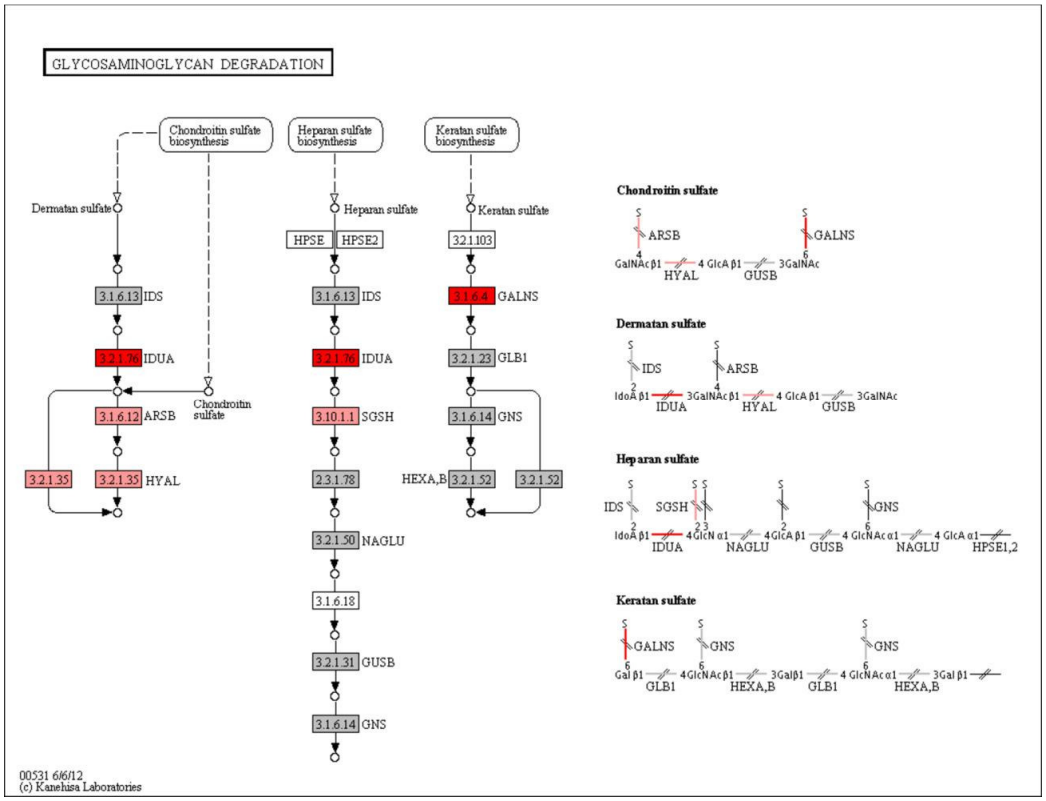


Figure 4.27 KEGG Map GAG Biosynthesis-Heparan Sulfate.

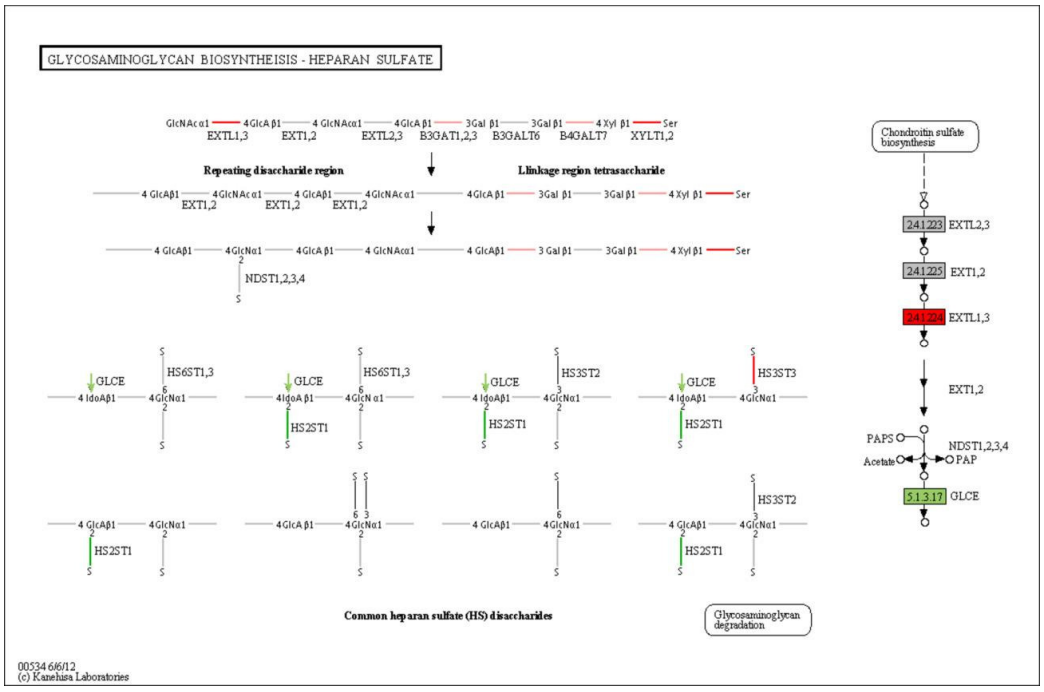


Figure 4.28 KEGG Map GAG Biosynthesis-Chondroitin Sulfate.

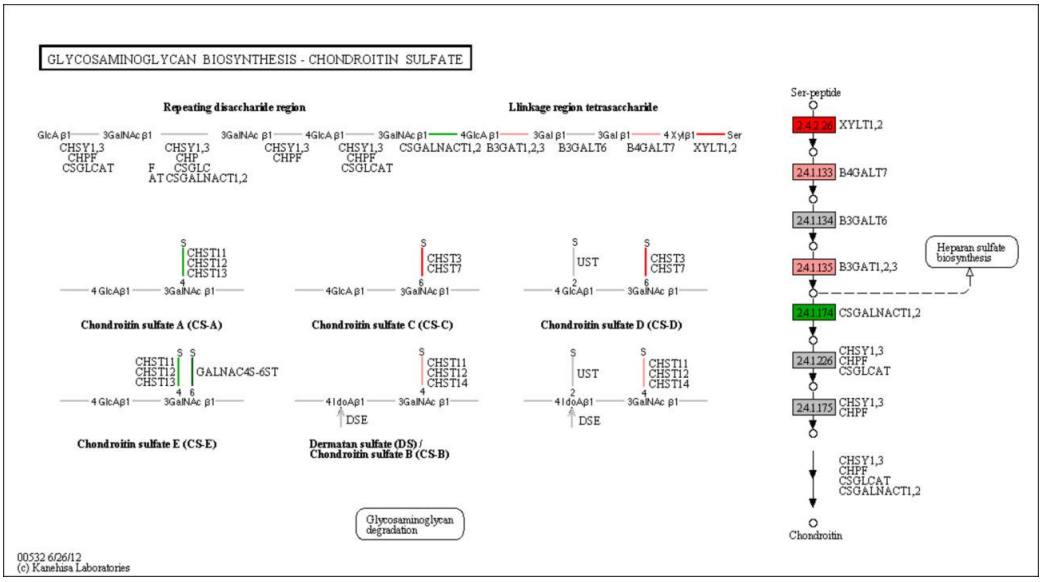
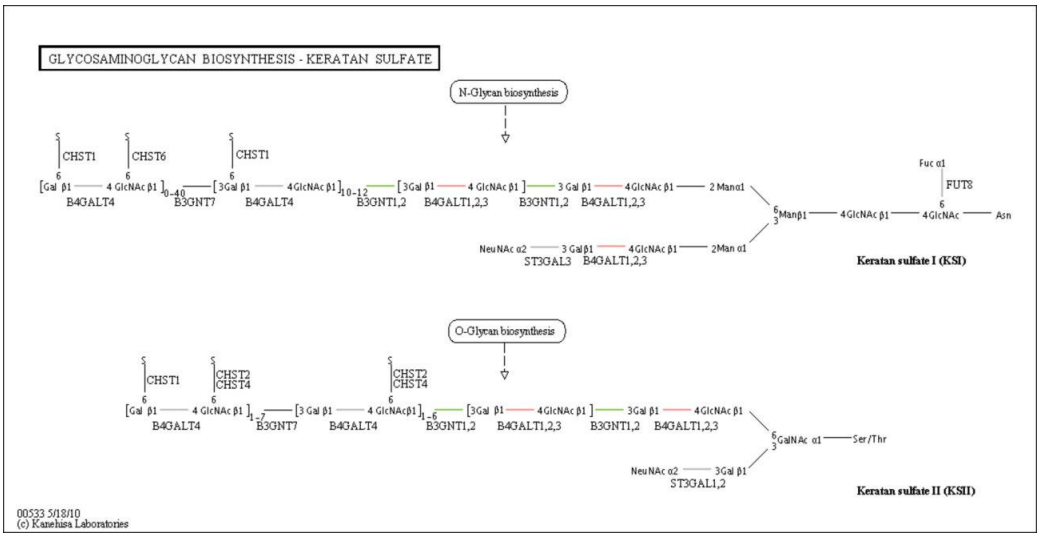


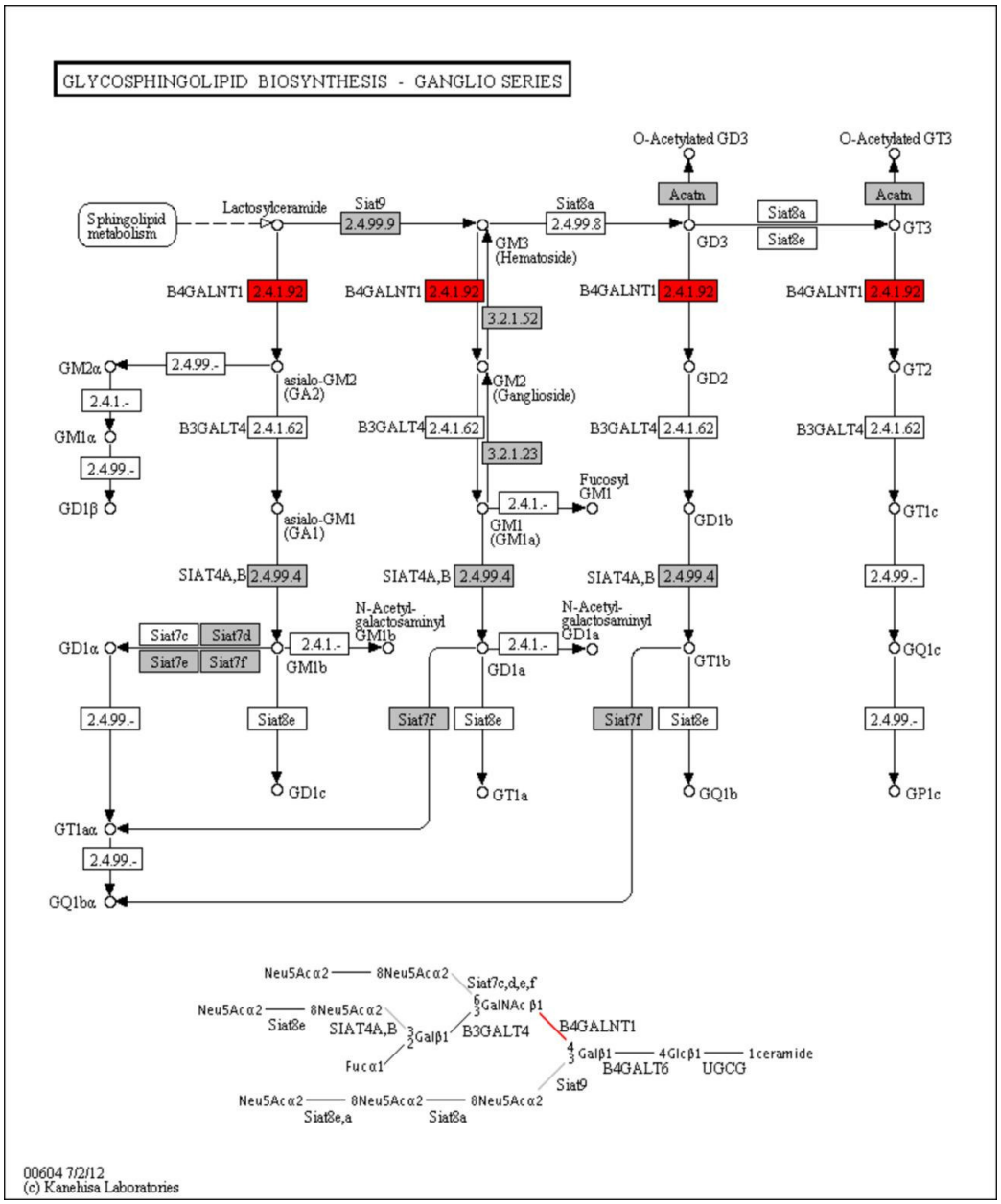
Figure 4.29 KEGG Map GAG Biosynthesis-Ketaran Sulfate.



Alteration in lipid metabolism highlighted by functional analysis refers to lipid, mainly phospholipid, biosynthetic process (Figure 4.9, Table 4.5). In this regards it should be noted that in MPS II in addition to accumulation of GAGs, which are the primary stored metabolites, also GM2 and GM3 gangliosides have been found as secondary stored metabolites. In this study B4GALNT1, an enzyme involved in the synthesis of GM2, has been found up-regulated in MPS II cells with respect to healthy control (log_2 ratio=1.84, p-value=0). This is shown in Figure

4.30, which illustrates glycosphingolipid biosynthesis pathway for ganglio series; unfortunately no information is available for other genes of the pathway.

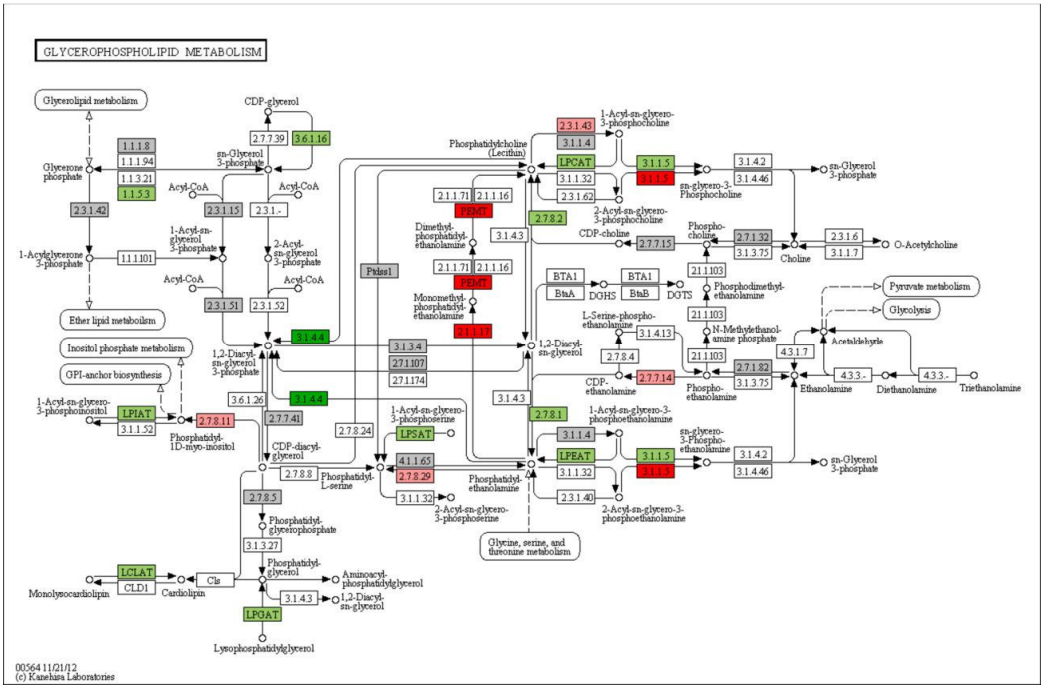
Figure 4.30 KEGG Map Glycosphingolipid biosynthesis- ganglio series.



Also other pathways of lipid metabolism have been found altered, among them, fatty acid metabolism and glycerophospholipid metabolism. This latter is shown in Figure 4.31**.** The most relevant variations involve PEMT (phosphatidylethanolamine N-methyltransferase), which catalyzes three sequential methylation reactions in the synthesis of phosphatidylcholine and

PLD1 (phosphatidylcholine-specific phospholipase D, 3.1.4.4 in Figure 4.31) which catalyzes the hydrolysis of phosphatidylcholine in phosphatidic acid and choline. PEMT is up-regulated in MPS II (log_2 ratio=1.50, p-value=9.99E-15), while PLD1 is down-regulated (log_2 ratio=1.74, p-value=0), thus suggesting an increasing of phosphatidylcholine level.

Figure 4.31KEGG Map Glycerophospholipid metabolism.



Modifications of phospholipid pathway were also found in other LSDs, as Sandhoff disease and in Gaucher disease. Interestingly in both diseases accumulation of GM2 occurs, in Sandhoff as primary storage, in Gaucher as secondary, together with GM3, like MPS II. In the first case phospholipids were found decreased in murine neurons as was phosphatidycholine in mouse brain. Also activities of two key enzymes in the synthesis of phosphatidylcholine and phosphatidylserine (CTP:phosphocholine cytidylyl-transferase and phosphatidylserine synthase, respectively) are decreased in the brain of Sandhoff mice, compared to wild type. It is thought that regulation of activity of these enzymes occurs post translationally. In contrast, in Gaucher mouse model the levels of various phospholipids are elevated. This is likely caused by a direct activation, by the accumulated glucosylceramide, of the rate limiting enzyme for the synthesis of phosphatidylcholine (CTP:phosphocholine cytidylyl-transferase) (Ballabio, Gieselmann 2009).

As for nucleotide metabolism, resulted altered in functional analysis (Figure 4.9, Table 4.5), both purine (Figure 4.32) and pyrimidine (Figure 4.33) metabolism seem to be affected by

variation. Moreover, observing the global map of cell metabolism also aminoacid metabolism, which did not emerge as altered from previous analysis, includes some differentially expressed genes. This was found for many aminoacid comprising arginine, proline, alanine, aspartate, cysteine, valine, leucine, isoleucine, tyrosine and tryptophan.

Figure 4.32KEGG Map Purine metabolism.

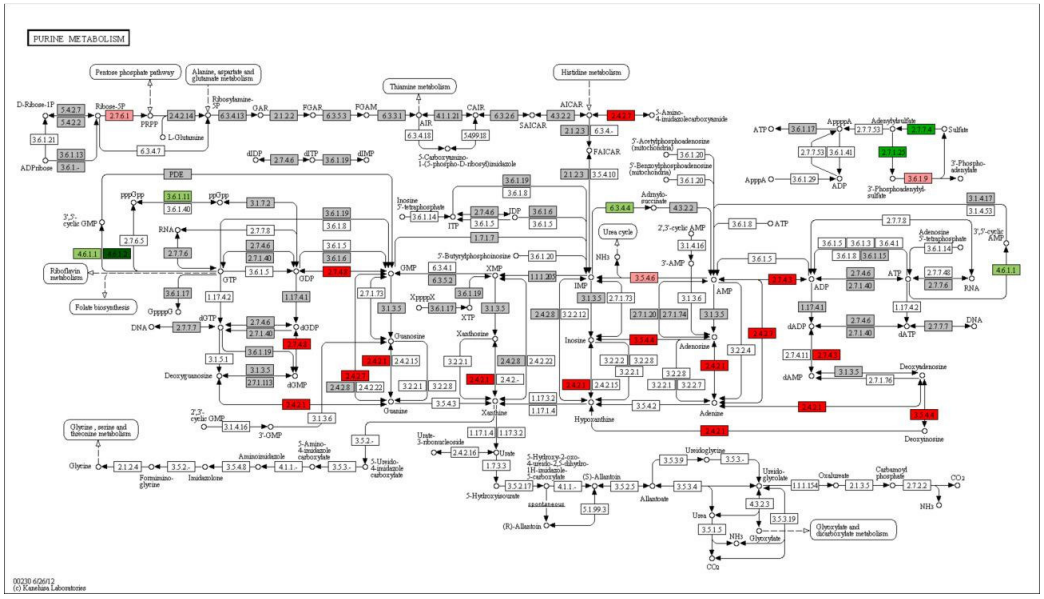
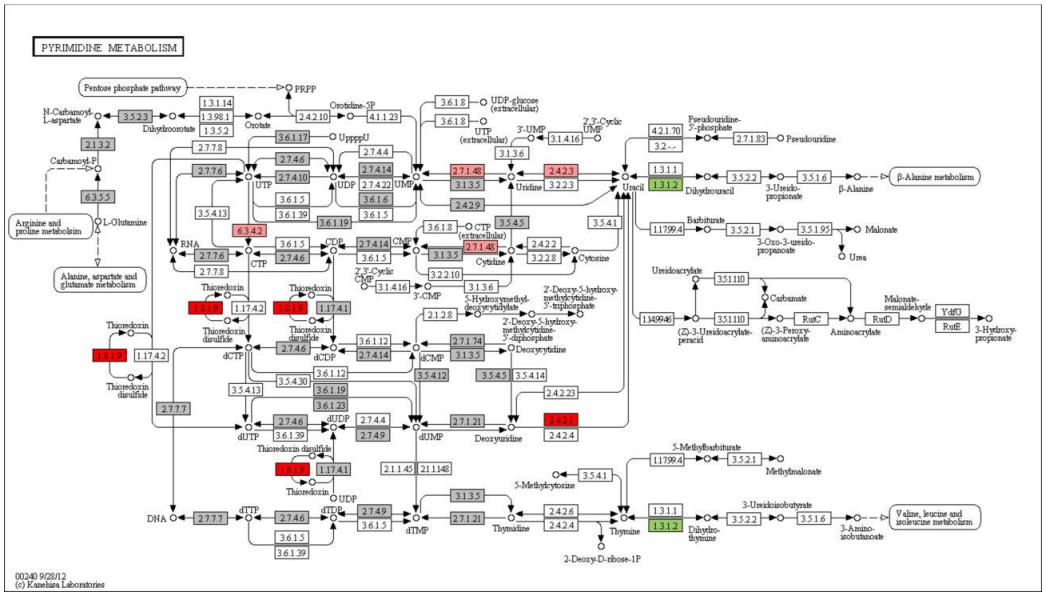


Figure 4.33 KEGG Map Pirimidyne metabolism.



As for the response to stimuli, functional and pathway analysis reported as overrepresented terms associated to cellular response to stress, comprising oxidative stress, wound healing, response to organic substances (lipid, cytokine and insulin) and immune response, in particular neutrophil mediated immunity.

Oxidative stress is thought to be a common mediator of apoptosis in several LSDs including all forms of NCLs, Tay-Sachs disease, Gaucher disease type I and II, NPC1 and 2 and nephropathic cystinosis (Zampieri et al. 2009).

In Gaucher disease, an up-regulation of apurinic endonuclease 1 (APE1) (a protein that repairs oxidative DNA damage) was observed in fibroblasts (Deganuto et al. 2007). Inducible nitric oxide synthase and nitrotyrosine were elevated in activated microglia/macrophages of GM1 and GM2 gangliosidoses (Jeyakumar et al. 2003a), and ROS was elevated in Fabry disease models (Shen et al. 2008) and in NPC1 fibroblasts. In the last case also gene microarray analysis confirmed oxidative stress hypothesis (Reddy, Ganley & Pfeffer 2006). In MPS IIIB, enhanced oxidative stress resulted in protein, lipid, and DNA oxidation (Villani et al. 2009), and an oxidative imbalance was found in MPS II (Reolon et al. 2009). In fibroblasts and brain extracts from CLN6 sheep, a significant increase in manganese-dependent superoxide dismutase activity was detected (Heine et al. 2003), and the expression of 4-hydroxynonenal was found increased in late infant and juvenile forms of NCL (Hachiya et al. 2006).

Here PRKD1 and ENC1 are up-regulated. PRKD1 is a serine/threonine kinase that regulates a variety of cellular functions, including membrane receptor signaling, transport at the Golgi, protection from oxidative stress at the mitochondria, gene transcription, and regulation of cell shape, motility, and adhesion. ENC1 plays a role in the oxidative stress response as a regulator of the transcription factor Nrf2.

The response to lipid which was evidenced in functional and pathway analysis include response to LPSs. In MPSs glycosaminoglycan (GAG) storage is not restricted to lysosomes but also occurs in the extracellular matrix (ECM), where they should be physiologically localized as components of ECM proteoglycans, and where GAG fragments with altered sulphatation accumulate. These fragments structurally resemble the bacterial endotoxin lypopolysaccharide (LPS), a ligand and activator of TLR4, whose binding leads to secretion of proinflammatory cytokines and immune response. Several genes involved in TLR signalling, including TLR4 itself, LPS binding protein and MyD88, are up-regulated in synovial cells from MPS VI cats and rats, and MPS VII dogs (Simonaro et al. 2008). Accordingly, chondrocytes display higher nitric oxide (NO) levels, secrete enhanced amount of proinflammatory cytokines such as IL-1β,TNF-α and

TGF-β. TLR pathway obtained for MPS II cells is shown in Figure 4.34. Both TLR4 and MyD88 are down-regulated, and many downstream element are altered with the final effect of altered production of IL-6 and IL-8.

Alteration of other cytokines also occurs, their expression profile is shown in Figure 4.35.

Figure 4.34 KEGG Map TOLL-like receptor signalling.

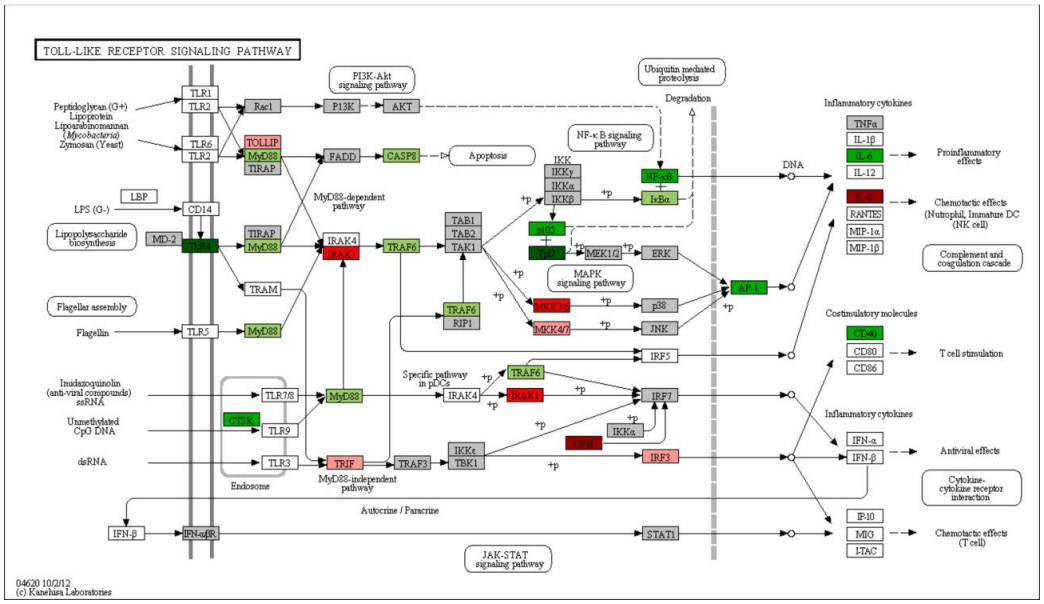
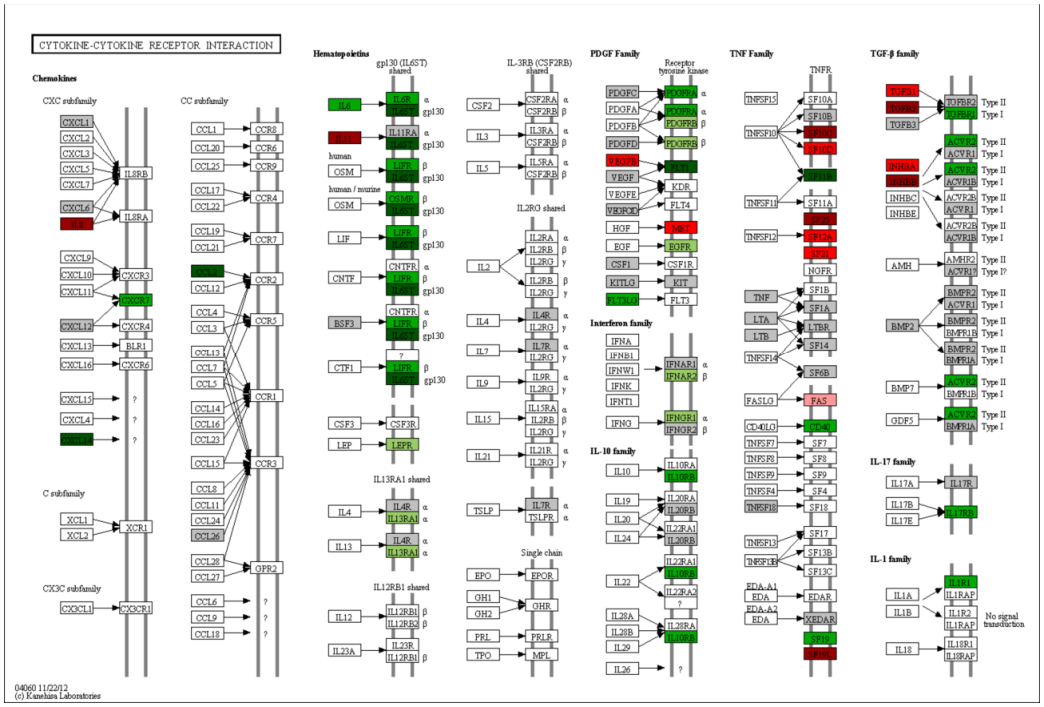


Figure 4.35 KEGG Map Cytokine- cytokine receptor interaction



According to functional analysis reported above, developmental processes seems to be altered at any level, both on embryonic and organismal stage and seems to involve many different tissues, organs and systems: skeletal, muscular, nervous, digestive and urinary systems, heart and vasculature, liver, connective tissue, hair, skin and blood (Figure 4.10, Table 4.6, Figure 4.16). Among the terms linked to anatomical structure formation, anterior/posterior pattern formation and appendage morphogenesis are the most represented (Figure 4.10). Figure 4.36 and Figure 4.37 represent the mapping of fibroblast gene expression data to developmental-related pathway, axon guidance for nervous system development and osteogenesis for skeletal system, respectively.

Figure 4.36 KEGG Map Axon guidance.

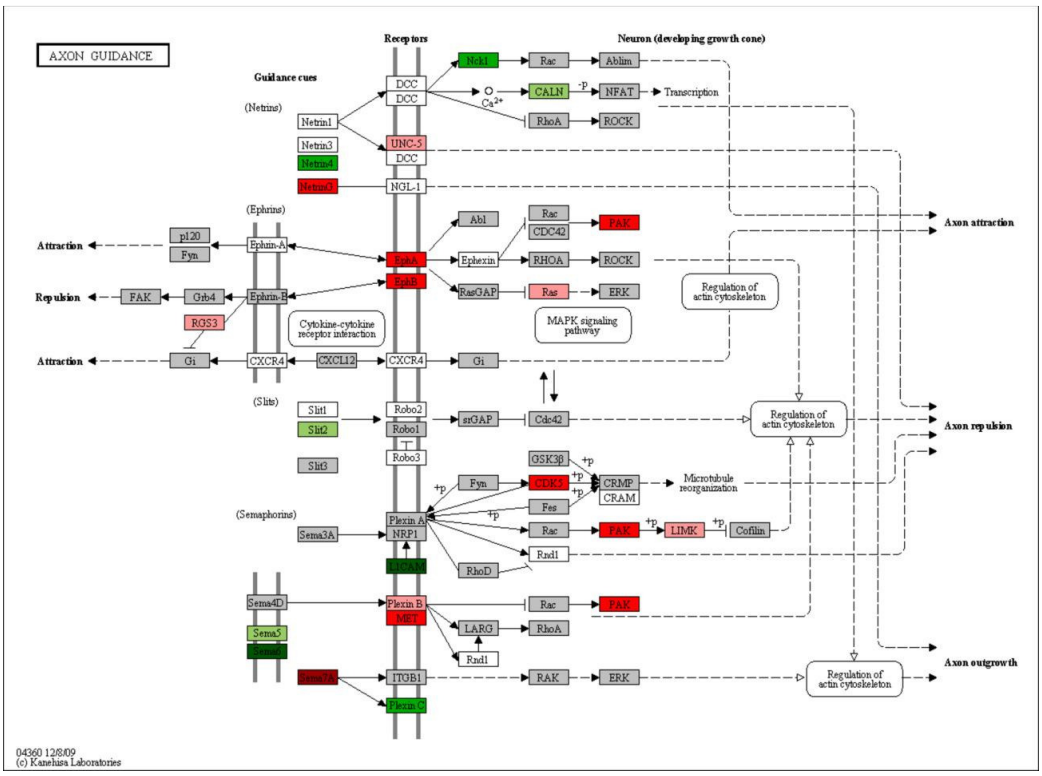
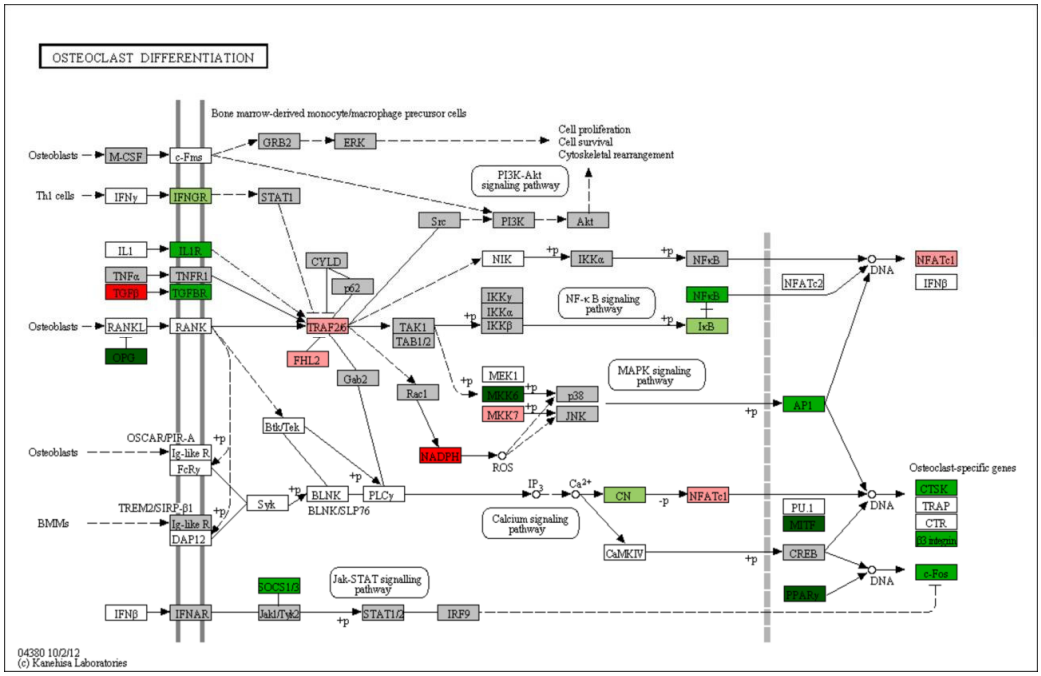


Figure 4.37 KEGG Map Osteoclast differentiation.



This finding can be explained with the widely demonstrated contribution of heparan sulfate proteoglycans (HSPGs) to many developmental processes. Thanks to their abilities to establish and mediate morphogen, growth factor, and cytokine gradients, facilitate signaling, provide structural stability to tissues, and act as molecular filters and barriers, these molecules play critical roles in the proliferation and differentiation of cell types. In nervous system, HSGP are involved in neural progenitor cells, neuronal migration, axon path-finding, synaptogenesis, neural plasticity and neural regeneration (Maeda et al. 2011). In the skeleton HSPG act as regulators of several dynamic aspects of skeletal development including patterning, differentiation, growth, and homeostasis. Moreover, heparan sulfate (HS) within the growth plate and bone marrow is also implicated as a critical component of the hematopoietic stem cell niche (Rodgers, San Antonio & Jacenko 2008).

HSPGs play these crucial roles by regulating key developmental signaling pathways, such as the Wnt, Hedgehog, transforming growth factor-β (TGFβ), and fibroblast growth factor (FGF) pathways (Moro et al. 2010).

Wnt, TGFβ and FGF pathways resulted as over-represented in pathway analysis (Figure 4.19, Figure 4.20); the corresponding expression patterns are shown in Figure 4.38 and Figure 4.39 for Wnt and TGFβ, respectively.

Figure 4.38 KEGG Map WNT signaling.

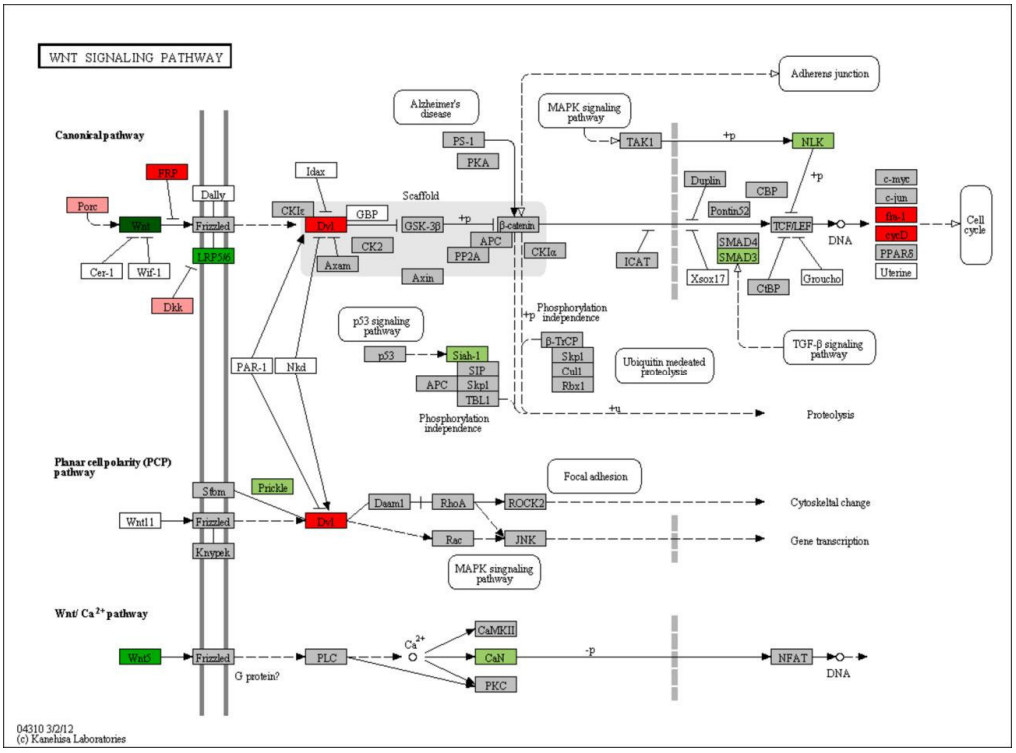
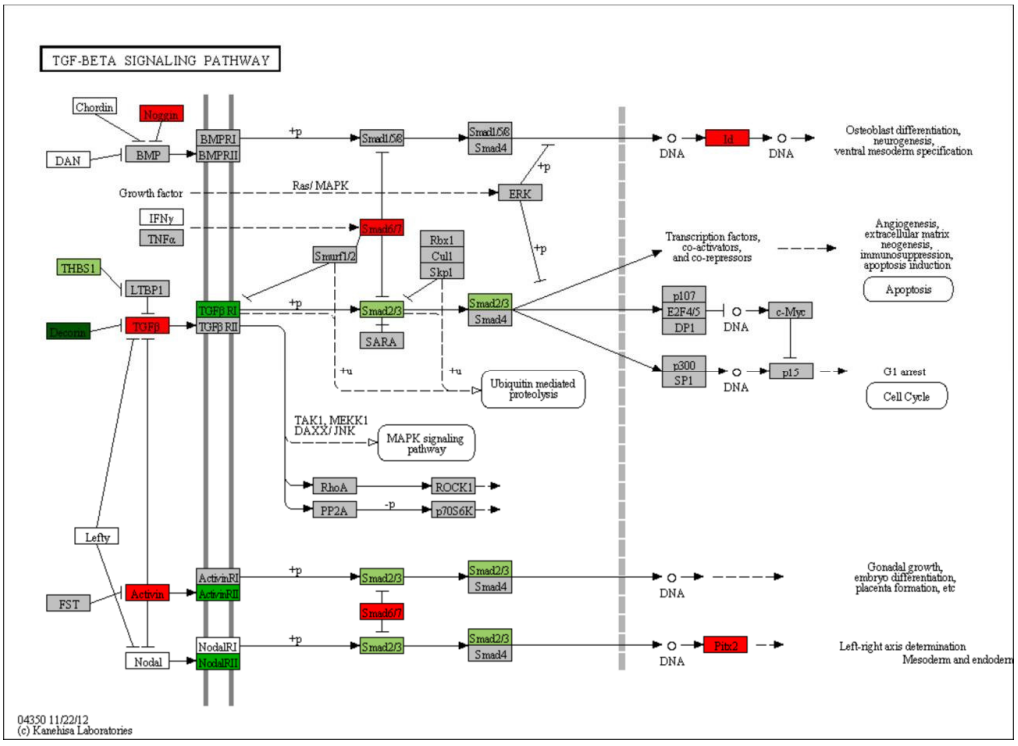


Figure 4.39 KEGG Map TGF-beta signaling.



In relation to this, it results very interesting to note that IDS was proved to play a critical role during vertebrate development, using zebrafish model (Moro et al. 2010). IDS transcript level and enzymatic activity are high already in unfertilized eggs and at later stages of development, an increase of IDS activity was observed during gastrulation. Moreover IDS-downregulation induces a significant percentage of lethality and severe developmental defects, including a misshapen head and trunk, abnormal craniofacial cartilages, disrupted body axis organization, reduced vasculature, as shown in Figure 4.40. A possible defect of the TGFβ signalling has been proposed to underlie this cartilage defect.

Figure 4.40 Zebrafish wild-type embryo (top) and morphant ids knockdown embrio (bottom), taken from (Moro et al. 2010) with permissions of the authors.

4.3 Search of molecular biomarkers

Gene expression data obtained from RNA-Seq transcriptome analysis of MPS II and healthy fibroblasts were also used for searching biomarkers of ERT therapeutic efficacy. These biomarkers were then evaluated in blood samples obtained from patients under treatment.

4.3.1.1 Selection of candidate biomarkers

In order to identify candidate biomarkers, transcripts were selected according to their expression profile in MPS II fibroblasts untreated and treated with IDS enzyme with respect to healthy control. In particular the following comparison were considered: untreated MPS II cells (UT) vs Healthy cells (C), MPS II cells treated with IDS for 24 h (T1) versus C and MPS II cells treated for 144 h (T2) vs C. To identify candidate markers of therapeutic efficacy, we selected genes that better reproduce the expression profile of a hypothetical ideal marker of therapeutic efficacy, that is having a significant difference of expression in Hunter cells respect

to healthy cells, decreasing with treatment. With the help of the software STEM, genes were clustered according to their expression ratios in the comparisons UT vs C, T1 vs C e T2 vs C, as showed in Figure 4.41, selected clusters are highlighted in blue.

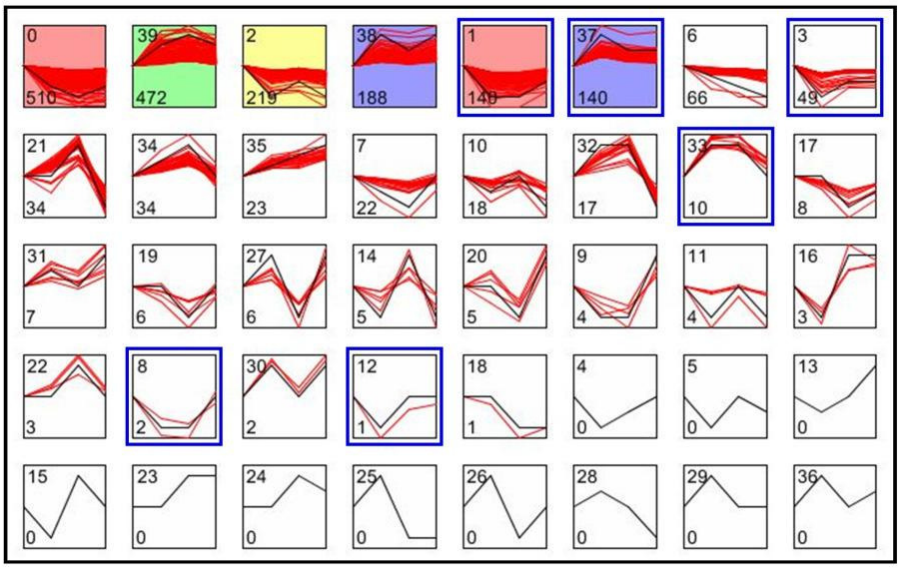
The selected clusters include overall 342 genes. To limit the list of candidate markers only to molecules easily detectable in patients, the obtained list of genes was further filtered according to gene expression levels in biological fluids.

Three different approaches were used, based on:

- expression level of transcript in whole blood
- presence of coded protein in plasma
- presence of coded protein in urine.

The first approach is based on the observation that blood cells express 80% of genes coded by human genome and share more than 80% of transcriptome of each of the subsequent tissues: brain, colon, kidney, liver, lung, prostate, spleen and stomach. Moreover, it is known that the expression profile of blood cells responds to physiological, pathological, and environmental stimuli with specific variation of expression profiles (Liew et al. 2006). In the last two strategies the so called ''Integrated Data Driven Proteomics method'' was used, according to which if the expression of a gene in a tissue is altered it is highly probable that the coded protein is differentially expressed and probably secreted in blood and urine (Chen et al. 2010).

Figure 4.41. Output of STEM software. For each profile is reported, in ordinate, log2(X/C) where X is, in order, the expression value in the samples C, UT, T1 and T2, and C the expression value in sample C. The number on the upper left of each square represents the profile ID, the one on the lower left the number of genes in the profile. The six selected profiles were highlighted in blue.



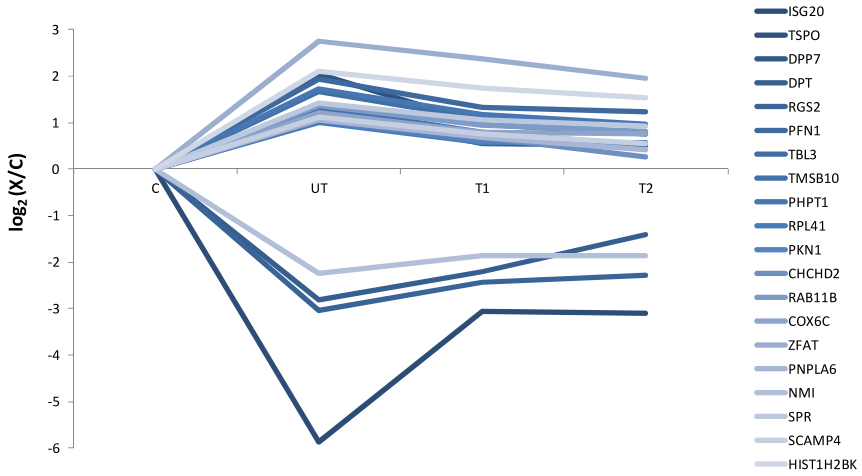
The application of this filter reduced the list of candidate biomarkers to 64 genes. Among these, the 20 genes showing the greatest variation in T1 vs UT comparison were selected; they are listed in Table 4.14 with the indication of the expression databases in which they are present. Figure 4.42 shows the expression profiles of the 20 selected markers.

Table 4.14. List of the 20 candidate markers with the indication of expression levels in the considered databases. w= expression value greater than 3 in the field whole blood in the database BioGPS, p= presence of coded protein in the 'HUPO Plasma Proteome Project database, u= presence of coded protein in plasma in MAPU Proteome database (urinary proteome).

GENE SYMBOL	PROTEIN	DATABASE	ENSEMBL_ID
CHCHD2	coiled-coil-helix-coiled-coil-helix domain containing 2	w	ENSG00000106153
COX6C	cytochrome c oxidase subunit VIc	w	ENSG00000164919
DPP7	dipeptidyl-peptidase 7	u	ENSG00000176978
DPT	dermatopontin	u	ENSG00000143196
HIST1H2BK	histone cluster 1, H2bk	w, p	ENSG00000197903
ISG20	interferon stimulated exonuclease gene 20kDa	w	ENSG00000172183
NMI	N-myc (and STAT) interactor	w	ENSG00000123609
PFN1	profilin 1	w, u, p	ENSG00000108518
PHPT1	phosphohistidine phosphatase 1	u	ENSG00000054148
PKN1	protein kinase N1	p	ENSG00000123143
PNPLA6	patatin-like phospholipase domain containing 6	p	ENSG00000032444

RAB11B	RAB11B, member RAS oncogene family	u	ENSG00000185236
RGS2	regulator of G-protein signaling 2, 24kDa	w	ENSG00000116741
RPL41	ribosomal protein L41	w	ENSG00000229117
SCAMP4	secretory carrier membrane protein 4	p	ENSG00000227500
SPR	sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)	u	ENSG00000116096
TBL3	transducin (beta)-like 3	p	ENSG00000183751
TMSB10	thymosin beta 10	w, u	ENSG00000034510
TSPO	translocator protein (18kDa)	w	ENSG00000100300
ZFAT	zinc finger and AT hook domain containing	p	ENSG00000066827

Figure 4.42. Expression profiles of the 20 selected markers. In ordinate is reported log2(X/C) where X is, in order, the expression value in the samples C, UT, T1 and T2, and C the expression value in sample C.



4.3.1.2 Clinical data collection

Identified candidate biomarkers were evaluated in a population of MPS II patients undergoing enzyme replacement therapy. These data are part of a wider study aimed to evaluate ERT efficacy by a 3 years follow up of 27 MPS II patients. Since evaluation of ERT efficacy is not the aim of the present study, the results of these evaluations will not be reported.

Data collection relative to the whole study was performed creating a relational database as specified in Chapter 3. From this database some parameters were extracted, according to their informative value and data availability, and used to perform correlation analysis with molecular data.

4.3.1.3 Quantitative PCR

Candidate biomarkers of therapeutic efficacy selected from transcriptome analysis have been evaluated by quantitative PCR in patients under ERT treatment. The analysis is still in progress; up to now 4 of the 20 candidate biomarkers have been evaluated in the samples available from 7 patients.

The 4 genes evaluated are DPT, ISG20, RGS2 and ZFAT.

DPT codes for dermatopontin, an extracellular matrix protein with possible functions in cell-matrix interactions and matrix assembly. It is postulated to modify the behaviour of TGF-beta through interaction with decorin.

ISG20 codes for an interferon-stimulated exonuclease with specificity for single-stranded RNA and, to a lesser extent for DNA. It is involved in the antiviral function of IFN against RNA.

RGS2 codes for a member of the regulator of G protein signaling (RGS) family which acts as GTPase activating protein (GAP) for G alpha subunits of heterotrimeric G proteins (Gi, Go and Gq alpha subtypes). It acts as a mediator of myeloid differentiation and may play a role in leukemogenesis.

ZFAT encodes a protein that likely binds DNA and functions as a transcriptional regulator involved in apoptosis and cell survival (http://www.genecards.org/).

These 4 genes were evaluated through RT-qPCR in blood samples from 7 patients at different time-points from the start of therapy as described in Chapter 3.

The expression values of these genes at different points were compared with pre-ERT samples, were available, and with an healthy control. Figure 4.43 shows, for each of the 4 genes selected and for each patient, the expression log ratio between each time-point of treatment and the pre-treatment value. Figure 4.44 shows the log ratio between the expression values of the 4 genes at each time-point with respect to the healthy control.

The pattern of gene expression evidenced through RT-qPCR presents a high degree of inter-individual variability.

Figure 4.43. Gene expression levels of 4 candidate markers at different time-points from the start of therapy with respect to pre-treatment. For each gene (DPT, ISG20, RGS2, ZFAT) and for each patient the log base 2 of the expression ratio between each time-point of treatment and the pre-treatment value is shown. Gene expression values have been obtained through RT-qPCR on blood samples.

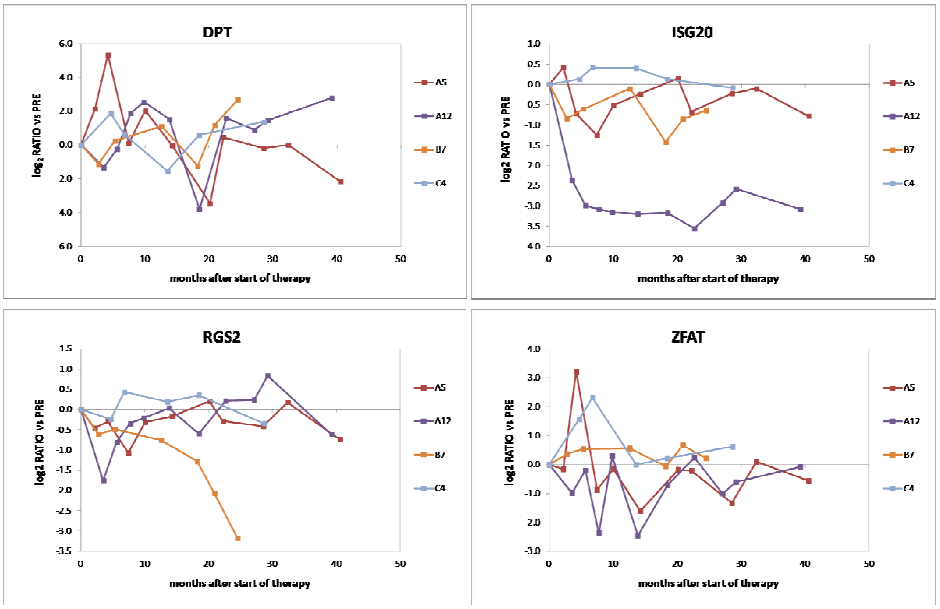
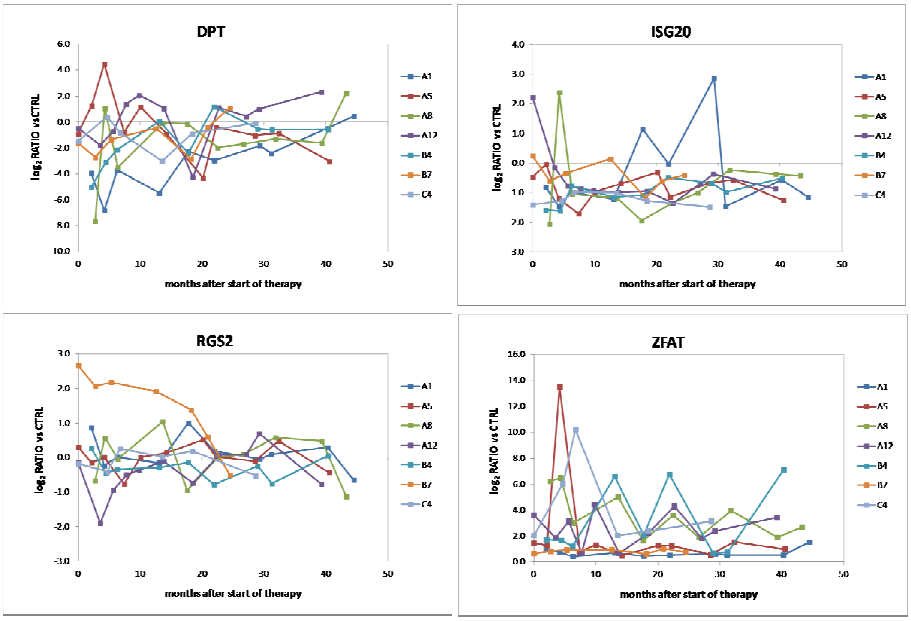


Figure 4.44. Gene expression levels of 4 candidate markers at different time-points from the start of therapy with respect to an healthy control. For each gene (DPT, ISG20, RGS2, ZFAT) and for each patient the log base 2 of the expression ratio between each time-point of treatment and an healthy control is shown. Gene expression values have been obtained through RT-qPCR on blood samples.



4.3.1.4 Correlation analysis

The obtained gene expression values were combined with the corresponding clinical data, as described in Chapter 3. A correlation analysis was performed between each considered clinical parameter (hepatomegaly, splenomegaly, valvulopathies, otitis media, hearing impairment, tonsillar hypertrophy, adenoid hypertrophy, sleep disorders, brain imaging abnormalities, cognitive disturbances, seizures and hydrocephalus) and the gene expression profile of each gene. Table 4.15 shows the result of the correlation analysis including the point biserial correlation coefficient (r_{bp}) and the p-value of Student's t-test for r_{bp} significance.

Table 4.15. Correlation analysis between gene expression profile of the candidate biomarkers (DPT, ISG20, RGS2, ZFAT) and clinical parameters; r_{bp}= point biserial correlation coefficient, p-value of Student's t-test for r_{bp} significance, *= p-value<0.05.

		r_{pb}	p (one - tailed)
Hepatomegaly	**DPT**	**-0.32**	0.0105*
	ISG20	-0.13	0.1784
	RGS2	-0.08	0.2789
	ZFAT	-0.21	0.0625
Splenomegaly	**DPT**	**0.37**	0.0210*
	ISG20	**-0.41**	0.0130*
	RGS2	-0.26	0.0805
	ZFAT	**0.45**	0.0060*
Otitis media	**DPT**	-0.1	0.2467
	ISG20	0.07	0.3132
	RGS2	-0.07	0.3062
	ZFAT	0.05	0.3677
Hearing impairment	**DPT**	-0.18	0.1032
	ISG20	-0.13	0.1759
	RGS2	-0.3	0.0162*
	ZFAT	**0.61**	<.0001*
Tonsillar hypertrophy	**DPT**	**0.51**	0.0006*
	ISG20	0.22	0.0603
	RGS2	0.29	0.0378*
	ZFAT	-0.06	0.3531
Adenoid hypertrophy	**DPT**	0.33	0.1373
	ISG20	-0.46	0.0576
	RGS2	-0.4	0.0889
	ZFAT	0.23	0.2231
Sleep disorders	**DPT**	-0.25	0.0314*
	ISG20	0.07	0.3131
	RGS2	0.11	0.2136
	ZFAT	**-0.37**	0.0026*
Cognitive disturbances	**DPT**	-0.24	0.0320*
	ISG20	-0.18	0.0850
	RGS2	-0.16	0.1137
	ZFAT	0.04	0.3864
Seizures	**DPT**	0.03	0.4331
	ISG20	-0.08	0.3389
	RGS2	**0.65**	<.0001*
	ZFAT	0.23	0.1110

Three of the parameters (valvulopathies, CNS abnormalities by brain imaging and hydrocephalus) presented the same outcome (presence or absence) for all evaluated data, therefore correlation analysis does not make sense. Data indicated with * represent significant correlation values. No strong correlations ($r_{bp>}0.7$) were observed; an intermediate correlation ($0.3 < r_{bp} < 0.7$) resulted for 8 pairs of gene/parameter evaluated, which are shown in bold type

in Table 4.15. The strongest correlations were observed between ZFAT and hearing impairment and between RGS2 and seizures, but also hepatomegaly, splenomegaly, tonsillar hypertrophy, sleep disorders and cognitive disturbances show correlation with at least one gene.

Extension of this evaluation to the other patients needs to be performed to confirm the obtained results. Also the analysis of the other candidate genes isolated from transcriptome analysis might indentify other potential biomarkers.

5 CONCLUSIONS

Hunter Syndrome (Mucopolysaccharidosis type II, MPS II) is a rare inherited metabolic disease belonging to the group of pathologies called Mucopolysaccharidoses (MPSs), and is due to the lack of activity of the lysosomal enzyme iduronate 2-sulfatase (IDS), involved in the degradation of the mucopolysaccharides heparan- and dermatan- sulphate. As a result such molecules accumulate within cell lysosomes and in the extracellular matrix of most tissues and organs, determining a severe pathological phenotype. Due to the housekeeping nature of IDS, enzymatic deficit affects most if not all the peripheral organ systems, also impairing cognitive functions in the severe forms of the disease.

MPSs belong to the wider group of the lysosomal storage disorders (LSDs), about 50 diverse lysosomal diseases, which, although characterized by different deposits of macromolecules, share several pathological signs and symptoms.

Studies conducted in the past few years to clarify biochemical and molecular pathways involved in lysosomal storage have provided some clues on LSDs pathogenesis; however, most of the primary molecular alterations causing the disease phenotypes as well as the resulting secondary changes remain unknown. This is particularly true in the case of Hunter syndrome, for which pathogenesis studies are quite limited almost absent.

On this basis, the present project has performed for the first time a high-throughput analysis of Hunter disease. A whole transcriptome analysis was conducted on fibroblasts derived from MPS II patients using RNA Sequencing by SOLiD System (Sequencing by Oligo Ligation and Detection). RNA-Seq is a full-length cDNA sequencing application that generates a comprehensive, quantitative, view of the entire cell transcriptome.

The comparison of the expression profile of Hunter vs healthy cells has highlighted alterations in basic cellular processes, metabolism, response to stimuli and developmental functions.

Cell cycle, replication, apoptosis and extra cellular matrix are widely compromise. Among metabolic processes, alterations in metabolism of both primary and one of the secondary stored metabolites, glycosaminoglycans and GM2 gangliosides respectively, are found, together with alterations in lipid, mainly fatty acid and phospholipid, metabolism. The latter seems to vary in such a way that an increase of phosphatidylcholine level could be suspected. Lipid metabolic alterations have been previously described in other LSDs, such as Sandhoff and

Gaucher disease; the present study has also evidenced gene expression variation in other metabolic pathways, not commonly associated with LSDs, such as nucleotide and aminoacid.

In Hunter cells many responses elicited in stress conditions, and also been associated to LSDs, seem to be active, comprising oxidative stress, cellular response to organic substances (lipid, cytokine and insulin) and immune response.

Cellular response to lipid seems to be due to a non physiological activation of Toll- like TLR receptor. As seen in other MPSs, GAG storage is not restricted to lysosomes but also occurs in the extracellular matrix (ECM), where they should be physiologically localized as components of ECM proteoglycans, and were GAG fragments with altered sulphatation accumulate. This fragments structurally resemble the bacterial endotoxin lypopolysaccharide (LPS), a ligand and activator of TLR4, whose binding leads to secretion of proinflammatory cytokines and immune response. Toll-like receptor signal pathway results altered in this study, together with the consequent production of cytokines and the immune response.

An unexpected finding is the reported alteration of developmental function. According to our data developmental processes seems to be altered at any level, both on embryonic and organismal stage and seems to involve many different tissues, organs and systems: skeletal, muscular, nervous, digestive and urinary systems, heart and vasculature, liver, connective tissue, hair, skin and blood. This finding can be explained with the widely demonstrated contribution of heparan sulfate proteoglycans (HSPGs) to many developmental processes. Thanks to their abilities to establish and mediate morphogens, growth factors and cytokine gradients, facilitate signaling, provide structural stability to tissues, and act as molecular filters and barriers, these molecules play critical roles in the proliferation and differentiation of cell types, especially in nervous system and skeleton. HSPGs play these crucial roles by regulating key developmental signalling pathways, such as the Wnt, Hedgehog, transforming growth factor-β (TGFβ), and fibroblast growth factor (FGF) pathways. In this study Wnt, TGFβ and FGF pathways resulted altered in MPS II cells. In relation to this, it results very interesting to note that IDS enzyme, deficient in MPS II, was proved to play a critical role during vertebrate development, using zebrafish model (Moro et al. 2010). IDS-downregulation induces a significant percentage of lethality and severe developmental defects, including a misshapen head and trunk, abnormal craniofacial cartilages, disrupted body axis organization, reduced vasculature. Interestingly, gene associated to body axic organization have been found misregulated in MPS II cells by the present analysis.

Overall these observations reveal an extensive gene expression alteration associated with MPS II pathology. The observed variations involve many, mostly essential, cellular processes confirming the very complex cellular scenario emerging in the last years for LSD pathology as well as the central role of lysosome in cell biology. A deep comprehension of these mechanisms would shed light on important cellular functions and would help to identify potential therapeutic targets as well as biomarkers of diagnosis, prognosis and therapeutic efficacy.

In addition to this transcriptomic characterization of MPS II cells, also a research of biomarkers for therapeutic efficacy of Enzyme Replacement Therapy in MPS II was performed. At the moment, a part from symptomatic treatments, the most significant therapeutic approach applied to MPS II is, in fact, represented by the ERT. However until now a valid and reliable biomarker of ERT efficacy for MPS II does not exist. The only biochemical marker taken into consideration at the moment is the urinary GAG excretion, which however does not reflect the disease burden and the efficacy of therapy (Clarke et al. 2012). Although application of ERT has obtained variable and very subjective results in the different patients analyzed, at present same therapeutic dose, time-schedule and modality of administration are adopted for all MPS II patients. Instead, it might be necessary to adjust both doses and timing, based on a personal treatment follow-up, thus favouring the design of a personalized therapeutic approach. Recent studies have indicated that serum heparin-cofactor II-thrombin complex (HCII-T) may be used as biomarker in the group of MPSs that accumulate dermatan-sulphate (Randall et al. 2006). More recently, a study on few ERT-treated MPS II patients revealed poor correlation between urinary GAG and serum HCII-T, while it evidenced a strong correlation between HCII-T and the presence of anti-IDS antibodies (Clarke et al. 2012). However, no correlations of the HCII-T profile with the clinical outcomes of the patients investigated was found. Hence the need to identify new biomarkers able to monitor the clinical status of the patient, the pathology progression and the response to treatment.

With this aim, in parallel with pathogenetic characterization, trancriptome analysis was used also as starting point for the research of molecular biomarker of ERT efficacy. MPS II cell were treated *in vitro* with the therapeutic IDS enzyme and collected 24 and 144 h post-treatment. Their transcriptional profile has been studied to characterize the early cellular response to the enzyme supply. Candidate biomarkers were selected based on their expression profiles during treatment with respect to healthy control, and filtered in according to the expression in biological fluids. 20 candidate biomarkers have been isolated and some of them have been

afterwards evaluated by using Real Time PCR, in blood samples obtained from Hunter patients under ERT. From this population clinical data were collected for a 3 years follow-up allowing to perform a correlation analysis between clinical parameters and the gene expression profile of each gene. The study is still in progress and up to now 4 (DPT, ISG20, RGS2, ZFAT genes) of the 20 candidate biomarkers have been evaluated. However a moderate correlation was fount for 8 pairs of gene/parameter evaluated. The strongest correlations were observed between ZFAT and hearing impairment and between RGS2 and seizures, but also hepatomegaly, splenomegaly, tonsillar hypertrophy, sleep disorders and cognitive disturbances show correlation with at least one gene. In particular DPT codes for dermatopontin, an extracellular matrix protein with possible functions in cell-matrix interactions and matrix assembly, which probably modifies the behaviour of TGF-beta through interaction with decorin. Being both extra cellular matrix and TGFβ altered in MPS II cells, we hypothesize that DPT could be a good biomarker, also having a role in MPS II pathogenesis.

Extension of this evaluation to the other patients needs to be performed to confirm the obtained results. Moreover, the analysis of the other candidate genes isolated from transcriptome analysis might indentify other potential biomarkers. This could allow a personalized therapeutic approach that will permit to improve clinical practice for patients in treatment, to optimize therapeutic efficacy and reduce or eliminate possible risks of under/overdosage.

6 REFERENCES

Alzheimer, C. & Werner, S. 2002, "Fibroblast growth factors and neuroprotection", *Advances in Experimental Medicine and Biology,* vol. 513, pp. 335-351.

Araya, K., Sakai, N., Mohri, I., Kagitani-Shimono, K., Okinaga, T., Hashii, Y., Ohta, H., Nakamichi, I., Aozasa, K., Taniike, M. & Ozono, K. 2009, "Localized donor cells in brain of a Hunter disease patient after cord blood stem cell transplantation", *Molecular genetics and metabolism,* vol. 98, no. 3, pp. 255-263.

Bach, G., Eisenberg, F.,Jr, Cantz, M. & Neufeld, E.F. 1973, "The defect in the Hunter syndrome: deficiency of sulfoiduronate sulfatase", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 70, no. 7, pp. 2134-2138.

Ballabio, A. & Gieselmann, V. 2009, "Lysosomal disorders: from storage to cellular damage", *Biochimica et biophysica acta,* vol. 1793, no. 4, pp. 684-696.

Baehner, F., Schmiedeskamp, C., Krummenauer, F., Miebach, E., Bajbouj, M., Whybra, C., Kohlschutter, A., Kampmann, C. & Beck, M. 2005, "Cumulative incidence rates of the mucopolysaccharidoses in Germany", *Journal of inherited metabolic disease,* vol. 28, no. 6, pp. 1011-1017.

Beck, M. 2010, "Therapy for lysosomal storage disorders", *IUBMB life,* vol. 62, no. 1, pp. 33-40.

Ben Turkia, H., Tebib, N., Azzouz, H., Abdelmoula, M.S., Ben Chehida, A., Chemli, J., Monastiri, K., Chaabouni, M., Sanhagi, H., Zouari, B., Kaabachi, N. & Ben Dridi, M.F. 2009, "Incidence of mucopolysaccharidoses in Tunisia", *La Tunisie medicale,* vol. 87, no. 11, pp. 782-785.

Bishop, J.R., Schuksz, M. & Esko, J.D. 2007, "Heparan sulphate proteoglycans fine-tune mammalian physiology", *Nature,* vol. 446, no. 7139, pp. 1030-1037.

Bondeson, M.L., Malmgren, H., Dahl, N., Carlberg, B.M. & Pettersson, U. 1995, "Presence of an IDS-related locus (IDS2) in Xq28 complicates the mutational analysis of Hunter syndrome", *European journal of human genetics : EJHG,* vol. 3, no. 4, pp. 219-227.

Brante, G. 1952, "Gargoylism; a mucopolysaccharidosis", *Scandinavian Journal of Clinical and Laboratory Investigation,* vol. 4, no. 1, pp. 43-46.

Burton, B.K. & Giugliani, R. 2012, "Diagnosing Hunter syndrome in pediatric practice: practical considerations and common pitfalls", *European journal of pediatrics,* vol. 171, no. 4, pp. 631-639.

Campagna, D., Albiero, A., Bilardi, A., Caniato, E., Forcato, C., Manavski, S., Vitulo, N. & Valle, G. 2009, "PASS: a program to align short sequences", *Bioinformatics (Oxford, England),* vol. 25, no. 7, pp. 967-968.

Campeau, P.M., Rafei, M., Boivin, M.N., Sun, Y., Grabowski, G.A. & Galipeau, J. 2009, "Characterization of Gaucher disease bone marrow mesenchymal stromal cells reveals an altered inflammatory secretome", *Blood,* vol. 114, no. 15, pp. 3181-3190.

Campos, D. & Monaga, M. 2012, "Mucopolysaccharidosis type I: current knowledge on its pathophysiological mechanisms", *Metabolic brain disease,* vol. 27, no. 2, pp. 121-129.

Cantz, M., Chrambach, A., Bach, G. & Neufeld, E.F. 1972, "The Hunter corrective factor. Purification and preliminary characterization", *The Journal of biological chemistry,* vol. 247, no. 17, pp. 5456-5462.

Cardone, M., Polito, V.A., Pepe, S., Mann, L., D'Azzo, A., Auricchio, A., Ballabio, A. & Cosma, M.P. 2006, "Correction of Hunter syndrome in the MPSII mouse model by AAV2/8-mediated gene delivery", *Human molecular genetics,* vol. 15, no. 7, pp. 1225-1236.

Castaneda, J.A., Lim, M.J., Cooper, J.D. & Pearce, D.A. 2008, "Immune system irregularities in lysosomal storage disorders", *Acta Neuropathologica,* vol. 115, no. 2, pp. 159-174.

Chen, D., Zhao, M. & Mundy, G.R. 2004, "Bone morphogenetic proteins", *Growth factors (Chur, Switzerland),* vol. 22, no. 4, pp. 233-241.

Chen, R., Sigdel, T.K., Li, L., Kambham, N., Dudley, J.T., Hsieh, S.C., Klassen, R.B., Chen, A., Caohuu, T., Morgan, A.A., Valantine, H.A., Khush, K.K., Sarwal, M.M. & Butte, A.J. 2010, "Differentially expressed RNA from public microarray data identifies serum protein biomarkers for cross-organ transplant rejection and other conditions", *PLoS computational biology,* vol. 6, no. 9, pp. e1000940.

Clarke, L.A., Hemmelgarn, H., Colobong, K., Thomas, A., Stockler, S., Casey, R., Chan, A., Fernoff, P. & Mitchell, J. 2012, "Longitudinal observations of serum heparin cofactor II-thrombin complex in treated Mucopolysaccharidosis I and II patients", *Journal of inherited metabolic disease,* vol. 35, no. 2, pp. 355-362.

Cox, T.M. & Cachon-Gonzalez, M.B. 2012, "The cellular pathology of lysosomal diseases", *The Journal of pathology,* vol. 226, no. 2, pp. 241-254.

Constantopoulos, G., Iqbal, K. & Dekaban, A.S. 1980, "Mucopolysaccharidosis types IH, IS, II, and IIIA: glycosaminoglycans and lipids of isolated brain cells and other fractions from autopsied tissues", *Journal of neurochemistry,* vol. 34, no. 6, pp. 1399-1411.

Deganuto, M., Pittis, M.G., Pines, A., Dominissini, S., Kelley, M.R., Garcia, R., Quadrifoglio, F., Bembi, B. & Tell, G. 2007, "Altered intracellular redox status in Gaucher disease fibroblasts and impairment of adaptive response against oxidative stress", *Journal of cellular physiology,* vol. 212, no. 1, pp. 223-235.

Dhami, R. & Schuchman, E.H. 2004, "Mannose 6-phosphate receptor-mediated uptake is defective in acid sphingomyelinase-deficient macrophages: implications for Niemann-Pick disease enzyme replacement therapy", *The Journal of biological chemistry,* vol. 279, no. 2, pp. 1526-1532.

Dorfman, A. & Lorincz, A.E. 1957, "Occurrence of Urinary Acid Mucopolysaccharides in the Hurler Syndrome", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 43, no. 6, pp. 443-446.

Dumeaux, V., Olsen, K.S., Nuel, G., Paulssen, R.H., Borresen-Dale, A.L. & Lund, E. 2010, "Deciphering normal blood gene expression variation--The NOWAC postgenome study", *PLoS genetics,* vol. 6, no. 3, pp. e1000873.

Elleder, M., Sokolova, J. & Hrebicek, M. 1997, "Follow-up study of subunit c of mitochondrial ATP synthase (SCMAS) in Batten disease and in unrelated lysosomal disorders", *Acta Neuropathologica,* vol. 93, no. 4, pp. 379-390.

Ernst, J. & Bar-Joseph, Z. 2006, "STEM: a tool for the analysis of short time series gene expression data", *BMC bioinformatics,* vol. 7, pp. 191.

Farfel-Becker, T., Vitner, E., Dekel, H., Leshem, N., Enquist, I.B., Karlsson, S. & Futerman, A.H. 2009, "No evidence for activation of the unfolded protein response in neuronopathic models of Gaucher disease", *Human molecular genetics,* vol. 18, no. 8, pp. 1482-1488.

Filocamo, M. & Morrone, A. 2011, "Lysosomal storage disorders: molecular basis and laboratory testing", *Human genomics,* vol. 5, no. 3, pp. 156-169.

Fratantoni, J.C., Hall, C.W. & Neufeld, E.F. 1968, "Hurler and Hunter syndromes: mutual correction of the defect in cultured fibroblasts", *Science (New York, N.Y.),* vol. 162, no. 853, pp. 570-572.

Friso, A., Tomanin, R., Alba, S., Gasparotto, N., Puicher, E.P., Fusco, M., Hortelano, G., Muenzer, J., Marin, O., Zacchello, F. & Scarpa, M. 2005, "Reduction of GAG storage in MPS II mouse model following implantation of encapsulated recombinant myoblasts", *The journal of gene medicine,* vol. 7, no. 11, pp. 1482-1491.

Friso, A., Tomanin, R., Zanetti, A., Mennuni, C., Calvaruso, F., La Monica, N., Marin, O., Zacchello, F. & Scarpa, M. 2008, "Gene therapy of Hunter syndrome: evaluation of the efficiency of muscle electro gene transfer for the production and release of recombinant iduronate-2-sulfatase (IDS)", *Biochimica et biophysica acta,* vol. 1782, no. 10, pp. 574-580.

Friso, A., Tomanin, R., Salvalaio, M. & Scarpa, M. 2010, "Genistein reduces glycosaminoglycan levels in a mouse model of mucopolysaccharidosis type II", *British journal of pharmacology,* vol. 159, no. 5, pp. 1082-1091.

Froissart, R., Da Silva, I.M. & Maire, I. 2007, "Mucopolysaccharidosis type II: an update on mutation spectrum", *Acta paediatrica (Oslo, Norway : 1992).Supplement,* vol. 96, no. 455, pp. 71-77.

Froissart, R., Moreira da Silva, I., Guffon, N., Bozon, D. & Maire, I. 2002, "Mucopolysaccharidosis type II--genotype/phenotype aspects", *Acta paediatrica (Oslo, Norway : 1992).Supplement,* vol. 91, no. 439, pp. 82-87.

Gallagher, J.T. 2006, "Multiprotein signalling complexes: regional assembly on heparan sulphate", *Biochemical Society transactions,* vol. 34, no. Pt 3, pp. 438-441.

Ginzburg, L. & Futerman, A.H. 2005, "Defective calcium homeostasis in the cerebellum in a mouse model of Niemann-Pick A disease", *Journal of neurochemistry,* vol. 95, no. 6, pp. 1619-1628.

Gritti, A., Frolichsthal-Schoeller, P., Galli, R., Parati, E.A., Cova, L., Pagano, S.F., Bjornson, C.R. & Vescovi, A.L. 1999, "Epidermal and fibroblast growth factors behave as mitogenic regulators for a single multipotent stem cell-like population from the subventricular region of the adult mouse forebrain", *The Journal of neuroscience : the official journal of the Society for Neuroscience,* vol. 19, no. 9, pp. 3287-3297.

Gucev, Z.S., Tasic, V., Sinigerska, I., Kremensky, I., Tincheva, R., Pop-Jordanova, N., Danilovski, D., Hofer, D. & Paschke, E. 2011, "Hunter syndrome (Muccopolysaccharridosis Type II) in Macedonia and Bulgaria", *Prilozi / Makedonska akademija na naukite i umetnostite, Oddelenie za bioloski i medicinski nauki = Contributions / Macedonian Academy of Sciences and Arts, Section of Biological and Medical Sciences,* vol. 32, no. 2, pp. 187-198.

Hachiya, Y., Hayashi, M., Kumada, S., Uchiyama, A., Tsuchiya, K. & Kurata, K. 2006, "Mechanisms of neurodegeneration in neuronal ceroid-lipofuscinoses", *Acta Neuropathologica,* vol. 111, no. 2, pp. 168-177.

Heine, C., Tyynela, J., Cooper, J.D., Palmer, D.N., Elleder, M., Kohlschutter, A. & Braulke, T. 2003, "Enhanced expression of manganese-dependent superoxide dismutase in human and sheep CLN6 tissues", *The Biochemical journal,* vol. 376, no. Pt 2, pp. 369-376.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. 2007, "qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data", *Genome biology,* vol. 8, no. 2, pp. R19.

Holt, J.B., Poe, M.D. & Escolar, M.L. 2011, "Natural progression of neurological disease in mucopolysaccharidosis type II", *Pediatrics,* vol. 127, no. 5, pp. e1258-65.

Huang da, W., Sherman, B.T. & Lempicki, R.A. 2009, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources", *Nature protocols,* vol. 4, no. 1, pp. 44-57.

Hunter, C. 1917, "A Rare Disease in Two Brothers", *Proceedings of the Royal Society of Medicine,* vol. 10, no. Sect Study Dis Child, pp. 104-116.

Jennings, J.J.,Jr, Zhu, J.H., Rbaibi, Y., Luo, X., Chu, C.T. & Kiselyov, K. 2006, "Mitochondrial aberrations in mucolipidosis Type IV", *The Journal of biological chemistry,* vol. 281, no. 51, pp. 39041-39050.

Jeyakumar, M., Smith, D.A., Williams, I.M., Borja, M.C., Neville, D.C., Butters, T.D., Dwek, R.A. & Platt, F.M. 2004, "NSAIDs increase survival in the Sandhoff disease mouse: synergy with N-butyldeoxynojirimycin", *Annals of Neurology,* vol. 56, no. 5, pp. 642-649.

Jeyakumar, M., Thomas, R., Elliot-Smith, E., Smith, D.A., van der Spoel, A.C., d'Azzo, A., Perry, V.H., Butters, T.D., Dwek, R.A. & Platt, F.M. 2003a, "Central nervous system inflammation is a hallmark of pathogenesis in mouse models of GM1 and GM2 gangliosidosis", *Brain : a journal of neurology,* vol. 126, no. Pt 4, pp. 974-987.

Jeyakumar, M., Thomas, R., Elliot-Smith, E., Smith, D.A., van der Spoel, A.C., d'Azzo, A., Perry, V.H., Butters, T.D., Dwek, R.A. & Platt, F.M. 2003b, "Central nervous system inflammation is a hallmark of pathogenesis in mouse models of GM1 and GM2 gangliosidosis", *Brain : a journal of neurology,* vol. 126, no. Pt 4, pp. 974-987.

Jmoudiak, M. & Futerman, A.H. 2005, "Gaucher disease: pathological mechanisms and modern management", *British journal of haematology,* vol. 129, no. 2, pp. 178-188.

Jung, S.C., Park, E.S., Choi, E.N., Kim, C.H., Kim, S.J. & Jin, D.K. 2010, "Characterization of a novel mucopolysaccharidosis type II mouse model and recombinant AAV2/8 vector-mediated gene therapy", *Molecules and cells,* vol. 30, no. 1, pp. 13-18.

Kanazawa, T., Nakamura, S., Momoi, M., Yamaji, T., Takematsu, H., Yano, H., Sabe, H., Yamamoto, A., Kawasaki, T. & Kozutsumi, Y. 2000, "Inhibition of cytokinesis by a lipid metabolite, psychosine", *The Journal of cell biology,* vol. 149, no. 4, pp. 943-950.

Khan, S.A., Nelson, M.S., Pan, C., Gaffney, P.M. & Gupta, P. 2008, "Endogenous heparan sulfate and heparin modulate bone morphogenetic protein-4 signaling and activity", *American journal of physiology.Cell physiology,* vol. 294, no. 6, pp. C1387-97.

Kim, S.J., Zhang, Z., Hitomi, E., Lee, Y.C. & Mukherjee, A.B. 2006, "Endoplasmic reticulum stress-induced caspase-4 activation mediates apoptosis and neurodegeneration in INCL", *Human molecular genetics,* vol. 15, no. 11, pp. 1826-1834.

Klein, D., Yaghootfam, A., Matzner, U., Koch, B., Braulke, T. & Gieselmann, V. 2009, "Mannose 6-phosphate receptor-dependent endocytosis of lysosomal enzymes is increased in sulfatide-storing kidney cells", *Biological chemistry,* vol. 390, no. 1, pp. 41-48.

Kobayashi, T., Beuchat, M.H., Lindsay, M., Frias, S., Palmiter, R.D., Sakuraba, H., Parton, R.G. & Gruenberg, J. 1999, "Late endosomal membranes rich in lysobisphosphatidic acid regulate cholesterol transport", *Nature cell biology,* vol. 1, no. 2, pp. 113-118.

Korkotian, E., Schwarz, A., Pelled, D., Schwarzmann, G., Segal, M. & Futerman, A.H. 1999, "Elevation of intracellular glucosylceramide levels results in an increase in endoplasmic reticulum density and in functional calcium stores in cultured neurons", *The Journal of biological chemistry,* vol. 274, no. 31, pp. 21673-21678.

Krabbi, K., Joost, K., Zordania, R., Talvik, I., Rein, R., Huijmans, J.G., Verheijen, F.V. & Ounap, K. 2012, "The live-birth prevalence of mucopolysaccharidoses in Estonia", *Genetic testing and molecular biomarkers,* vol. 16, no. 8, pp. 846-849.

Kuo, W.J., Digman, M.A. & Lander, A.D. 2010, "Heparan sulfate acts as a bone morphogenetic protein coreceptor by facilitating ligand-induced receptor hetero-oligomerization", *Molecular biology of the cell,* vol. 21, no. 22, pp. 4028-4041.

Lee, O.J., Kim, S.J., Sohn, Y.B., Park, H.D., Lee, S.Y., Kim, C.H., Ko, A.R., Yook, Y.J., Lee, S.J., Park, S.W., Kim, S.H., Cho, S.Y., Kwon, E.K., Han, S.J. & Jin, D.K. 2012, "A study of the relationship between clinical phenotypes and plasma iduronate-2-sulfatase enzyme activities in Hunter syndrome patients", *Korean journal of pediatrics,* vol. 55, no. 3, pp. 88-92.

Lieberman, A.P., Puertollano, R., Raben, N., Slaugenhaupt, S., Walkley, S.U. & Ballabio, A. 2012, "Autophagy in lysosomal storage disorders", *Autophagy,* vol. 8, no. 5, pp. 719-730.

Liew, C.C., Ma, J., Tang, H.C., Zheng, R. & Dempsey, A.A. 2006, "The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool", *The Journal of laboratory and clinical medicine,* vol. 147, no. 3, pp. 126-132.

Lin, H.Y., Lin, S.P., Chuang, C.K., Niu, D.M., Chen, M.R., Tsai, F.J., Chao, M.C., Chiu, P.C., Lin, S.J., Tsai, L.P., Hwu, W.L. & Lin, J.L. 2009, "Incidence of the mucopolysaccharidoses in Taiwan, 1984-2004", *American journal of medical genetics.Part A,* vol. 149A, no. 5, pp. 960-964.

Lloyd-Evans, E., Morgan, A.J., He, X., Smith, D.A., Elliot-Smith, E., Sillence, D.J., Churchill, G.C., Schuchman, E.H., Galione, A. & Platt, F.M. 2008, "Niemann-Pick disease type C1 is a sphingosine storage disease that causes deregulation of lysosomal calcium", *Nature medicine,* vol. 14, no. 11, pp. 1247-1255.

Lloyd-Evans, E., Pelled, D., Riebeling, C., Bodennec, J., de-Morgan, A., Waller, H., Schiffmann, R. & Futerman, A.H. 2003, "Glucosylceramide and glucosylsphingosine modulate calcium mobilization from brain microsomes via different mechanisms", *The Journal of biological chemistry,* vol. 278, no. 26, pp. 23594-23599.

Lorincz, A.E. 1978, "The mucopolysaccharidoses: advances in understanding and treatment", *Pediatric annals,* vol. 7, no. 2, pp. 104-122.

Luiro, K., Kopra, O., Blom, T., Gentile, M., Mitchison, H.M., Hovatta, I., Tornquist, K. & Jalanko, A. 2006, "Batten disease (JNCL) is linked to disturbances in mitochondrial, cytoskeletal, and synaptic compartments", *Journal of neuroscience research,* vol. 84, no. 5, pp. 1124-1138.

Maeda, N., Ishii, M., Nishimura, K. & Kamimura, K. 2011, "Functions of chondroitin sulfate and heparan sulfate in the developing brain", *Neurochemical research,* vol. 36, no. 7, pp. 1228-1240.

Maere, S., Heymans, K. & Kuiper, M. 2005, "BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks", *Bioinformatics (Oxford, England),* vol. 21, no. 16, pp. 3448-3449.

Maiuri, M.C., Zalckvar, E., Kimchi, A. & Kroemer, G. 2007, "Self-eating and self-killing: crosstalk between autophagy and apoptosis", *Nature reviews.Molecular cell biology,* vol. 8, no. 9, pp. 741-752.

Manara, R., Priante, E., Grimaldi, M., Santoro, L., Astarita, L., Barone, R., Concolino, D., Di Rocco, M., Donati, M.A., Fecarotta, S., Ficcadenti, A., Fiumara, A., Furlan, F., Giovannini, I., Lilliu, F., Mardari, R., Polonara, G., Procopio, E., Rampazzo, A., Rossi, A., Sanna, G., Parini, R. & Scarpa, M. 2011, "Brain and spine MRI features of Hunter disease: frequency, natural

evolution and response to therapy", *Journal of inherited metabolic disease,* vol. 34, no. 3, pp. 763-780.

Marks, D.L. & Pagano, R.E. 2002, "Endocytosis and sorting of glycosphingolipids in sphingolipid storage disease", *Trends in cell biology,* vol. 12, no. 12, pp. 605-613.

Martin, R., Beck, M., Eng, C., Giugliani, R., Harmatz, P., Munoz, V. & Muenzer, J. 2008a, "Recognition and diagnosis of mucopolysaccharidosis II (Hunter syndrome)", *Pediatrics,* vol. 121, no. 2, pp. e377-86.

Meikle, P.J., Hopwood, J.J., Clague, A.E. & Carey, W.F. 1999, "Prevalence of lysosomal storage disorders", *JAMA : the journal of the American Medical Association,* vol. 281, no. 3, pp. 249-254.

Mizushima, N., Ohsumi, Y. & Yoshimori, T. 2002, "Autophagosome formation in mammalian cells", *Cell structure and function,* vol. 27, no. 6, pp. 421-429.

Moro, E., Tomanin, R., Friso, A., Modena, N., Tiso, N., Scarpa, M. & Argenton, F. 2010, "A novel functional role of iduronate-2-sulfatase in zebrafish early development", *Matrix biology : journal of the International Society for Matrix Biology,* vol. 29, no. 1, pp. 43-50.

Muenzer, J. 2011, "Overview of the mucopolysaccharidoses", *Rheumatology (Oxford, England),* vol. 50 Suppl 5, pp. v4-12.

Muenzer, J., Wraith, J.E., Beck, M., Giugliani, R., Harmatz, P., Eng, C.M., Vellodi, A., Martin, R., Ramaswami, U., Gucsavas-Calikoglu, M., Vijayaraghavan, S., Wendt, S., Puga, A.C., Ulbrich, B., Shinawi, M., Cleary, M., Piper, D., Conway, A.M. & Kimura, A. 2006, "A phase II/III clinical study of enzyme replacement therapy with idursulfase in mucopolysaccharidosis II (Hunter syndrome)", *Genetics in medicine : official journal of the American College of Medical Genetics,* vol. 8, no. 8, pp. 465-473.

Mulloy, B. & Rider, C.C. 2006, "Cytokines and proteoglycans: an introductory overview", *Biochemical Society transactions,* vol. 34, no. Pt 3, pp. 409-413.

Nelson, J. 1997, "Incidence of the mucopolysaccharidoses in Northern Ireland", *Human genetics,* vol. 101, no. 3, pp. 355-358.

Nelson, J., Crowhurst, J., Carey, B. & Greed, L. 2003, "Incidence of the mucopolysaccharidoses in Western Australia", *American journal of medical genetics.Part A,* vol. 123A, no. 3, pp. 310-313.

Pan, C., Nelson, M.S., Reyes, M., Koodie, L., Brazil, J.J., Stephenson, E.J., Zhao, R.C., Peters, C., Selleck, S.B., Stringer, S.E. & Gupta, P. 2005, "Functional abnormalities of heparan sulfate in mucopolysaccharidosis-I are associated with defective biologic activity of FGF-2 on human multipotent progenitor cells", *Blood,* vol. 106, no. 6, pp. 1956-1964.

Pelled, D., Lloyd-Evans, E., Riebeling, C., Jeyakumar, M., Platt, F.M. & Futerman, A.H. 2003, "Inhibition of calcium uptake via the sarco/endoplasmic reticulum Ca2+-ATPase in a mouse

model of Sandhoff disease and prevention by treatment with N-butyldeoxynojirimycin", *The Journal of biological chemistry,* vol. 278, no. 32, pp. 29496-29501.

Pina-Aguilar, R.E., Zaragoza-Arevalo, G.R., Rau, I., Gal, A., Alcantara-Ortigoza, M.A., Lopez-Martinez, M.S. & Santillan-Hernandez, Y. 2012, "Mucopolysaccharidosis type II in a female carrying a heterozygous stop mutation of the iduronate-2-sulfatase gene and showing a skewed X chromosome inactivation", *European journal of medical genetics,* .

Poorthuis, B.J., Wevers, R.A., Kleijer, W.J., Groener, J.E., de Jong, J.G., van Weely, S., Niezen-Koning, K.E. & van Diggelen, O.P. 1999, "The frequency of lysosomal storage diseases in The Netherlands", *Human genetics,* vol. 105, no. 1-2, pp. 151-156.

Randall, D.R., Sinclair, G.B., Colobong, K.E., Hetty, E. & Clarke, L.A. 2006, "Heparin cofactor II-thrombin complex in MPS I: a biomarker of MPS disease", *Molecular genetics and metabolism,* vol. 88, no. 3, pp. 235-243.

Ransohoff, R.M. & Perry, V.H. 2009, "Microglial physiology: unique stimuli, specialized responses", *Annual Review of Immunology,* vol. 27, pp. 119-145.

Ravikumar, B., Berger, Z., Vacher, C., O'Kane, C.J. & Rubinsztein, D.C. 2006, "Rapamycin pre-treatment protects against apoptosis", *Human molecular genetics,* vol. 15, no. 7, pp. 1209-1216.

Reddy, J.V., Ganley, I.G. & Pfeffer, S.R. 2006, "Clues to neuro-degeneration in Niemann-Pick type C disease from global gene expression profiling", *PloS one,* vol. 1, pp. e19.

Reolon, G.K., Reinke, A., de Oliveira, M.R., Braga, L.M., Camassola, M., Andrades, M.E., Moreira, J.C., Nardi, N.B., Roesler, R. & Dal-Pizzol, F. 2009, "Alterations in oxidative markers in the cerebellum and peripheral organs in MPS I mice", *Cellular and molecular neurobiology,* vol. 29, no. 4, pp. 443-448.

Rodgers, K.D., San Antonio, J.D. & Jacenko, O. 2008, "Heparan sulfate proteoglycans: a GAGgle of skeletal-hematopoietic regulators", *Developmental dynamics : an official publication of the American Association of Anatomists,* vol. 237, no. 10, pp. 2622-2642.

Sands, M.S. & Haskins, M.E. 2008, "CNS-directed gene therapy for lysosomal storage diseases", *Acta paediatrica (Oslo, Norway : 1992).Supplement,* vol. 97, no. 457, pp. 22-27.

Sano, R., Annunziata, I., Patterson, A., Moshiach, S., Gomero, E., Opferman, J., Forte, M. & d'Azzo, A. 2009, "GM1-ganglioside accumulation at the mitochondria-associated ER membranes links ER stress to Ca(2+)-dependent mitochondrial apoptosis", *Molecular cell,* vol. 36, no. 3, pp. 500-511.

Schulze-Frenking, G., Jones, S.A., Roberts, J., Beck, M. & Wraith, J.E. 2011, "Effects of enzyme replacement therapy on growth in patients with mucopolysaccharidosis type II", *Journal of inherited metabolic disease,* vol. 34, no. 1, pp. 203-208.

Shen, J.S., Meng, X.L., Moore, D.F., Quirk, J.M., Shayman, J.A., Schiffmann, R. & Kaneski, C.R. 2008, "Globotriaosylceramide induces oxidative stress and up-regulates cell adhesion

molecule expression in Fabry disease endothelial cells", *Molecular genetics and metabolism,* vol. 95, no. 3, pp. 163-168.

Sillence, D.J. & Platt, F.M. 2004, "Glycosphingolipids in endocytic membrane transport", *Seminars in cell & developmental biology,* vol. 15, no. 4, pp. 409-416.

Simonaro, C.M., D'Angelo, M., He, X., Eliyahu, E., Shtraizent, N., Haskins, M.E. & Schuchman, E.H. 2008, "Mechanism of glycosaminoglycan-mediated bone and joint disease: implications for the mucopolysaccharidoses and other connective tissue diseases", *The American journal of pathology,* vol. 172, no. 1, pp. 112-122.

Smith, D., Wallom, K.L., Williams, I.M., Jeyakumar, M. & Platt, F.M. 2009, "Beneficial effects of anti-inflammatory therapy in a mouse model of Niemann-Pick disease type C1", *Neurobiology of disease,* vol. 36, no. 2, pp. 242-251.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L. & Ideker, T. 2011, "Cytoscape 2.8: new features for data integration and network visualization", *Bioinformatics (Oxford, England),* vol. 27, no. 3, pp. 431-432.

Staretz-Chacham, O., Lang, T.C., LaMarca, M.E., Krasnewich, D. & Sidransky, E. 2009, "Lysosomal storage disorders in the newborn", *Pediatrics,* vol. 123, no. 4, pp. 1191-1207.

Sukegawa-Hayasaka, K., Kato, Z., Nakamura, H., Tomatsu, S., Fukao, T., Kuwata, K., Orii, T. & Kondo, N. 2006, "Effect of Hunter disease (mucopolysaccharidosis type II) mutations on molecular phenotypes of iduronate-2-sulfatase: enzymatic activity, protein processing and structural analysis", *Journal of inherited metabolic disease,* vol. 29, no. 6, pp. 755-761.

Suzuki, M., Sugimoto, Y., Ohsaki, Y., Ueno, M., Kato, S., Kitamura, Y., Hosokawa, H., Davies, J.P., Ioannou, Y.A., Vanier, M.T., Ohno, K. & Ninomiya, H. 2007, "Endosomal accumulation of Toll-like receptor 4 causes constitutive secretion of cytokines and activation of signal transducers and activators of transcription in Niemann-Pick disease type C (NPC) fibroblasts: a potential basis for glial cell activation in the NPC brain", *The Journal of neuroscience : the official journal of the Society for Neuroscience,* vol. 27, no. 8, pp. 1879-1891.

Tagami, S., Inokuchi Ji, J., Kabayama, K., Yoshimura, H., Kitamura, F., Uemura, S., Ogawa, C., Ishii, A., Saito, M., Ohtsuka, Y., Sakaue, S. & Igarashi, Y. 2002, "Ganglioside GM3 participates in the pathological conditions of insulin resistance", *The Journal of biological chemistry,* vol. 277, no. 5, pp. 3085-3092.

Takada, T., Katagiri, T., Ifuku, M., Morimura, N., Kobayashi, M., Hasegawa, K., Ogamo, A. & Kamijo, R. 2003, "Sulfated polysaccharides enhance the biological activities of bone morphogenetic proteins", *The Journal of biological chemistry,* vol. 278, no. 44, pp. 43229-43235.

Tanaka, A., Okuyama, T., Suzuki, Y., Sakai, N., Takakura, H., Sawada, T., Tanaka, T., Otomo, T., Ohashi, T., Ishige-Wada, M., Yabe, H., Ohura, T., Suzuki, N., Kato, K., Adachi, S., Kobayashi, R., Mugishima, H. & Kato, S. 2012, "Long-term efficacy of hematopoietic stem cell transplantation on brain involvement in patients with mucopolysaccharidosis type II: a

nationwide survey in Japan", *Molecular genetics and metabolism,* vol. 107, no. 3, pp. 513-520.

Tessitore, A., del P Martin, M., Sano, R., Ma, Y., Mann, L., Ingrassia, A., Laywell, E.D., Steindler, D.A., Hendershot, L.M. & d'Azzo, A. 2004, "GM1-ganglioside-mediated activation of the unfolded protein response causes neuronal death in a neurodegenerative gangliosidosis", *Molecular cell,* vol. 15, no. 5, pp. 753-766.

Tessitore, A., Faella, A., O'Malley, T., Cotugno, G., Doria, M., Kunieda, T., Matarese, G., Haskins, M.E. & Auricchio, A. 2008, "Biochemical, pathological, and skeletal improvement of mucopolysaccharidosis VI after gene transfer to liver but not to muscle", *Molecular therapy : the journal of the American Society of Gene Therapy,* vol. 16, no. 1, pp. 30-37.

Timms, K.M., Bondeson, M.L., Ansari-Lari, M.A., Lagerstedt, K., Muzny, D.M., Dugan-Rocha, S.P., Nelson, D.L., Pettersson, U. & Gibbs, R.A. 1997, "Molecular and phenotypic variation in patients with severe Hunter syndrome", *Human molecular genetics,* vol. 6, no. 3, pp. 479-486.

Tomanin, R., Zanetti, A., Zaccariotto, E., D'Avanzo, F., Bellettato, C.M. & Scarpa, M. 2012, "Gene therapy approaches for lysosomal storage disorders, a good model for the treatment of mendelian diseases", *Acta Paediatrica (Oslo, Norway : 1992),* vol. 101, no. 7, pp. 692-701.

Tuschl, K., Gal, A., Paschke, E., Kircher, S. & Bodamer, O.A. 2005, "Mucopolysaccharidosis type II in females: case report and review of literature", *Pediatric neurology,* vol. 32, no. 4, pp. 270-272.

Vainio, S., Bykov, I., Hermansson, M., Jokitalo, E., Somerharju, P. & Ikonen, E. 2005, "Defective insulin receptor activation and altered lipid rafts in Niemann-Pick type C disease hepatocytes", *The Biochemical journal,* vol. 391, no. Pt 3, pp. 465-472.

Vellodi, A. 2005, "Lysosomal storage disorders", *British journal of haematology,* vol. 128, no. 4, pp. 413-431.

Villani, G.R., Di Domenico, C., Musella, A., Cecere, F., Di Napoli, D. & Di Natale, P. 2009, "Mucopolysaccharidosis IIIB: oxidative damage and cytotoxic cell involvement in the neuronal pathogenesis", *Brain research,* vol. 1279, pp. 99-108.

Vitner, E.B., Platt, F.M. & Futerman, A.H. 2010, "Common and uncommon pathogenic cascades in lysosomal storage diseases", *The Journal of biological chemistry,* vol. 285, no. 27, pp. 20423-20427.

Walkley, S.U. & Vanier, M.T. 2009, "Secondary lipid accumulation in lysosomal disease", *Biochimica et biophysica acta,* vol. 1793, no. 4, pp. 726-736.

Walton, R.M. & Wolfe, J.H. 2007, "Abnormalities in neural progenitor cells in a dog model of lysosomal storage disease", *Journal of neuropathology and experimental neurology,* vol. 66, no. 8, pp. 760-769.

Wang, L., Feng, Z., Wang, X., Wang, X. & Zhang, X. 2010, "DEGseq: an R package for identifying differentially expressed genes from RNA-seq data", *Bioinformatics (Oxford, England),* vol. 26, no. 1, pp. 136-138.

Wang, Z., Gerstein, M. & Snyder, M. 2009, "RNA-Seq: a revolutionary tool for transcriptomics", *Nature reviews.Genetics,* vol. 10, no. 1, pp. 57-63.

Wei, H., Kim, S.J., Zhang, Z., Tsai, P.C., Wisniewski, K.E. & Mukherjee, A.B. 2008, "ER and oxidative stresses are common mediators of apoptosis in both neurodegenerative and non-neurodegenerative lysosomal storage disorders and are alleviated by chemical chaperones", *Human molecular genetics,* vol. 17, no. 4, pp. 469-477.

Wilson, P.J., Morris, C.P., Anson, D.S., Occhiodoro, T., Bielicki, J., Clements, P.R. & Hopwood, J.J. 1990, "Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 87, no. 21, pp. 8531-8535.

Wraith, J.E., Scarpa, M., Beck, M., Bodamer, O.A., De Meirleir, L., Guffon, N., Meldgaard Lund, A., Malm, G., Van der Ploeg, A.T. & Zeman, J. 2008, "Mucopolysaccharidosis type II (Hunter syndrome): a clinical review and recommendations for treatment in the era of enzyme replacement therapy", *European journal of pediatrics,* vol. 167, no. 3, pp. 267-277.

Young, I.D. & Harper, P.S. 1981, "Psychosocial problems in Hunter's syndrome", *Child: care, health and development,* vol. 7, no. 4, pp. 201-209.

Zampieri, S., Mellon, S.H., Butters, T.D., Nevyjel, M., Covey, D.F., Bembi, B. & Dardis, A. 2009, "Oxidative stress in NPC1 deficient cells: protective effect of allopregnanolone", *Journal of Cellular and Molecular Medicine,* vol. 13, no. 9B, pp. 3786-3796.