Metachromatic Leukodystrophy: A Bioinformatics Approach through Protein-Protein Interaction Network Analysis

A Major Project dissertation submitted

in partial fulfilment of the requirement for the degree of

Master of Technology

In

Bioinformatics

Submitted by

Shri Ram

(2K11/BIO/17)

Delhi Technological University, Delhi, India

Under the supervision of

Dr. B.D. Malhotra

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road,
Delhi-110042, India

DECLARATION

I hereby declare that this project titled "**METACHROMATIC LEUKODYSTROPHY: A BIOINFORMATICS APPROACH THROUGH PROTEIN-PROTEIN INTERACTION NETWORK ANALYSIS**", submitted to Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), Delhi is the original and independent work carried out by me under the guidance of Dr. Punit Kaur, Additional Professor, Head of the Department, Department of Biophysics, All India Institute of Medical Sciences, New Delhi, in partial fulfilment of the requirements for the award of Degree of Master of Technology in Bioinformatics and this dissertation or part of it has not been submitted by me for any other degree /diploma in any University/ Institute.

Shri Ram
2k11/BIO/17

ACKNOWLEDGMENT

It really gives me great pleasure to express my deep sense of gratitude to all those who were really involved and helped me to complete my project work. With utmost sincerity, I am grateful to Dr. B.D. Malhotra, Professor, Department of Biotechnology, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), Delhi for giving me the permission and opportunity to do this project work.

I express my gratitude and indebtedness to Dr. Punit Kaur, Additional Professor, Head of the Department, Department of Biophysics, A.I.I.M.S, New Delhi, for her invaluable guidance and supervision, which has enabled me to complete the research work and to write this thesis. She has constantly steer-headed and supervised my work all throughout with her valuable guidance, suggestions, encouragement and care for which I am in dearth of words to convey my gratitude.

I am indebted to Dr. Parul Sharma who provided me both logistic and moral support through the highs and lows of my work. Her empathetic, cheerful and affectionate behaviour is unforgettable. I express my gratitude towards her valuable assistance, untiring efforts and constant help in entire work.

I wish to thank Ms. Kamna Gosain for helping and supporting me in this project. I would also like to thank Dr. Manoj Kumar and Mrs. Umay Kulsum for their valuable help and support.

I would also like to thank all of my trainee mates, who made my time and work more enjoyable.

And I express my sincere thanks to my brother Mr. Shrikrishan and friend Mr. Nitish Sharma for their everlasting support and encouragement. Last but not the least, this work would not be enlightened without the blessings of Loving God and my parents

Shri Ram

2k11/BIO/17

CONTENTS

TOPIC **PAGE NO**

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

MLD	Metachromatic Leukodystrophy
CNS	Central Nervous System
PNS	Peripheral Nervous System
ASA	Arylsulfatase A
GM2	Type of Gangliosidoses
GM3	Monosialodihexosylganglioside
MBP	Myelin Basic Protein
PLP	Proteolipid Protein
SCS	Schwann Cells
SAP-1	Sphingolipid Activator Protein-1
ADHD	Attention-Deficit Hyperactivity Disorder
MRI	Magnetic Resonance Image
ERT	Enzyme Replacement Therapy
SRT	Substrate Reduction Therapy
EET	Enzyme Enhancement Therapy
SQL	Structured Query Language
MiMI	Molecular Interactions from Michigan Molecular Interactions
BiNGO	Biological Networks Gene Ontology Tool
FWER	Family-wise Error Rate
FDR	False Discovery Rate

ClusterONE	Clustering with Overlapping Neighborhood Expansion
EPC	Edge Percolated Component
MNC	Maximum Neighborhood Component
DMNC	Density Of Maximum Neighborhood Component
MCC	Maximal Clique Centrality
BN	Bottleneck
GAD	Genetic Association Database
RDBMS	Relational Database Management System
DAVID	Databasefor Annotation, Visualization and Integrated Discovery
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
BIND	Biomolecular Interaction Network Database
DIP	Database of Interacting Proteins
MINT	Molecular Interaction Database
BioGRID	Biological General Repository for Interaction Datasets
HPRD	Human Protein Reference Database
OMIM	Online Mendelian Inheritance in Man
BH	Benjamini Hochberg
HECT	Homologous to the E6-AP Carboxyl Terminus
ARVC	Arrhythmogenic Right Ventricular Cardiomyopathy

Metachromatic Leukodystrophy: A Bioinformatics Approach through Protein-Protein Interaction Network Analysis

Shri Ram

Delhi Technological University, Delhi, India

1. ABSTRACT

System networks help in identifying new drug targets which in turn will generate more in-depth understanding of the underlying mechanism of diseases. Network based approach to study the pathogenesis of a disease is emerging as an important paradigm for analysis of biological systems. In this work the interaction network for Metachromatic Leukodystrophy (MLD) was built using Cytoscape and the analysis of network was carried out to find the drug targets. The most functional and highly interconnected sub-networks in the network were found which could help in understanding the mechanism of MLD.

Metachromatic Leukodystrophy (MLD) is one of a group of genetic disorders called the leukodystrophies. These diseases impair the growth or development of the myelin sheath, the fatty covering that acts as an insulator around nerve fibers. MLD is caused by a deficiency of the enzyme arylsulfatase A (ARSA). Lacking of complete knowledge of gene Co-expression network and pathogenesis for Metachromatic Leukodystrophy motivated us to develop an interaction network for this disorder. The *in-silico* prediction of potential interactions between nodes (genes) and target genes are of core importance for the identification of new drugs or novel targets for existing drugs. One of the aims of the study was to identify the functional and highly interconnected nodes in the network. Identifying the important sub-networks in the system could provide useful insights into the underlying molecular mechanism for Metachromatic Leukodystrophy. Another aim of the study was to identify candidate genes with high centrality score and perform their disease and pathway enrichment analysis to find out the disease classes in which these genes are involved and the pathways

they are affecting. An additional aspect of study was chromosome enrichment analysis of candidate genes to calculate the distribution of genes across different chromosomes.

Clique analysis was performed on the network and the highly interconnected clusters were found. SMAD9, PSAP, BMPR2, ARSA, UBE3A which may be the functional modules and can be identified as highly interconnected sub-graphs in the network. Most important potential drug targets found were TAF1, SMAD2, BRCA1, HNF4A, AR, SMAD9, CDC2, RB1, UBC, CDK2, UBB, PSAP, CDC23, MYC, MNAT1, CCNH, CDK7.

2. INTRODUCTION

Metachromatic Leukodystrophy (MLD) is an autosomal-recessive inherited lysosomal disorder, which belongs to the family of leukodystrophies as well as the sphingolipidoses as it affects the metabolism of sphingolipids. The MLD was discovered by Scholz Greenfield in 1925. MLD is rare, about 1 in 100,000 people has this disorder. MLD is also known as Scholz's disease or Arylsulfatase A deficiency **[**Barboura I, Ferchichi S, 2010]. Deficiency of the enzyme arylsulfatase-A (ARSA), or, more rarely, of its activator protein saposin-B leads to MLD [Holtschmidt H *et al.*, 1991, Aicardı` J (ed),1998, Von Figura K,2001]. In partial deficiency of ARSA, known as ARSA pseudodeficiency, which is sometimes also found in non-affected individuals, ARSA enzyme activity that is 5% to 20% of normal controls does not cause MLD [Arvan L Fluharty,2011].

The function of enzyme ARSA is to metabolise sulfatides by hydrolysing the 3-O ester bond of galactosyl and lactosyl sulfatides. The deficiency of enzyme ARSA leads to the accumulation of these substrates in metachromatic granules in oligodendrocytes, macrophages and some subtypes of neurons in the CNS, in Schwann cells and macrophages in the peripheral nervous system (PNS), and in visceral organs such as adrenal cortex, liver, pancreas, testes, kidney, sweat glands, , gallbladder and rectal tissue.

Four types of diseases are currently distinguished (according to the age at onset):

- Late infantile
- Early juvenile
- Late juvenile
- Adult forms

The late infantile form of the disease becomes evident in the second year of life, usually at 15–24 months. The juvenile form usually begins between 4 and 15 years of age and is further subdivided into early juvenile and late juvenile depending on whether the onset is before or after 6 years of age. Adult forms refer to the disease type which begins after 16 years of age. Adult-onset MLD progresses more slowly than the late infantile and juvenile forms, with a protracted course of a decade or more.

Late infantile and early juvenile cases are the most frequent, having a more severe phenotype and rapid progression. Characteristic symptoms of disease variants, because of the involvement of both CNS and PNS, include muscle wasting and weakness, muscle rigidity, developmental delays, convulsions, impaired swallowing, paralysis, dementia, dysphagia, blindness and deafness, seizures, diffuse atrophy and weakness with areflexia. Prognosis is severe, leading to death a few years after the onset of symptoms. In late juvenile cases, cognitive difficulties generally precede gait disturbances and the progression is slower. In the adult form, behavioral and intellectual changes are often the presenting symptoms, followed by spastic paraparesis and incontinence[Hyde TM,1992, Aicardı` J (ed),1998].

The infantile form, which presents a fairly constant clinical and pathological pattern, was first described by Greenfield (1933) who recorded two cases, and four further examples were reported by Russell Brain and Greenfield (1950). Jervis (1960) found 10 cases in the literature with pathological observations and described two more. To these should be added one recorded by Jacobi (1947) in which metachromasia was subsequently demonstrated by Diezel (1957), and another briefly reported by Peiffer and Hirsch (1955). Other cases have recently been described by Norman, Urich, and Tingey (1960) and by Hagberg, Sourander, Svennerholm, and Voss (1960) [J. W. Black and J. N. Cumings,1961].

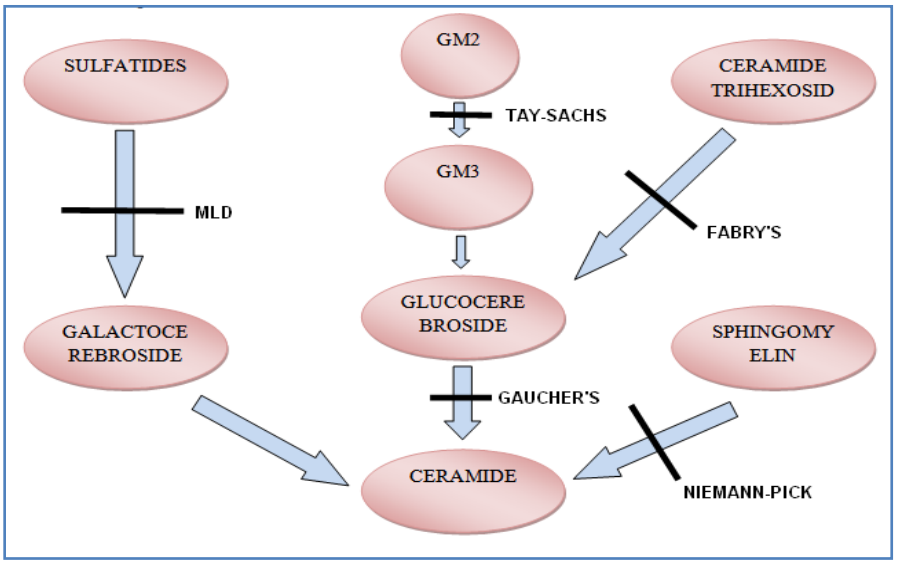


Figure 1: A diagram of Metachromatic Leukodystrophy and some other inborn errors of metabolism, showing the pathways and the names of the diseases when deficient.

At present, there is no successful treatment available to overcome the deterioration and loss of function caused by metachromatic leukodystrophy (MLD). In case of late infantile and early juvenile forms of the disease, bone marrow or cord blood transplantation may stabilize neurocognitive function [Krivit W,2004, Martin PL, Carter SL,2006] however, symptoms of motor function loss frequently progress. Mildly symptomatic and asymptomatic late juvenile and adult-onset forms are slower in progression, so are more likely to be stabilized with bone marrow transplantation.

In this study, we performed a system analysis of gene co-expression network in Metachromatic Leukodystrophy to investigate the topological characteristics of modulated genes in biological interaction network. In an effort to understand the molecular basis of genetic disease, it is important to discover their casual genes. Genes related to same disease are also known to have protein products that physically interact. A class of computational approaches has recently been proposed that exploit these two sources of information, physical interaction networks and linkage intervals, to predict associations between genes and diseases. The study of gene co-expression network is essential to define the molecular networks that contribute to maintain homeostasis of an organism's body functions. Disruptions in the gene networks results in disease in both human and animals. In effort to understand underlying mechanism of Metachromatic Leukodystrophy, Cytoscape, its plug-ins and various databases were used.

3. REVIEW OF LITERATURE

The word Leukodystrophy is coined from the Greek word "leuko" meaning white which refers to the white matter of the nervous system and "dystrophy" meaning inadequate development or growth. in medicinal terminology "dystrophy" imply to a condition which is progressive; that is, as the patient gets older the condition tends to get worse.

German neurologist Alzheimer published the first description of a patient with MLD in the year 1910. Later on, Metachromatic granules in liver, kidney and testis apart from brain were reported by Witte in 1921. Earlier MLD was known as a diffuse brain sclerosis. In 1938 Einarson and Neel renamed it as MLD [von Figura *et al*., 2001]. ASA deficiency in MLD patients was reported by Austin in 1963. Later on, Mehl and Jatzkewitz studied sulfatide metabolism and showed a block in the metabolism of sulfatide [Mehl and Jatzkewitz, 1965]. In 1975, very low levels of ASA were reported in some of the healthy individuals with a family history of MLD. Afterwards it was recognized that there is a moderately common allele of the ASA gene that leads to low expression of the enzyme [von Figura *et al*., 2001].

Metachromatic Leukodystrophy (MLD) is an uncommon type of diffuse cerebral sclerosis, characterized by the accretion of fats called sulfatides in cells, principally cells of the nervous system. [J.W.Black, J.N. Cumings,1961]. MLD is also known as Scholz's disease or Arylsulfatase A deficiency [Barboura I, Ferchichi S, 2010]. It is basically an autosomal-recessive neurometabolic disorder which belongs to the family of leukodystrophies as well as the sphingolipidoses. The accretion of fats results in progressive demolition of white matter of the brain, which consists of nerve fibers covered by myelin sheath. Myelin is a substance that insulates and protects nerves.

Metachromatic leukodystrophy is inherited in an autosomal recessive pattern and is caused by deficiency of enzyme arylsulfatase-A (ARSA), or, more rarely, of its activator protein saposin-B. Mutations in the ARSA and PSAP genes result in a decreased ability of enzyme arylsulfatase-A (ARSA), and protein saposin-B to break down sulfatides, resulting in the accretion of these substances in cells. Surplus sulfatides are toxic to the nervous system. This accretion progressively destroys myelin-producing cells, leads to the destruction of nervous system function [Holtschmidt H *et al.*, 1991, Aicardı` J (ed),1998, Von Figura K,2001].

The clinical features include progressive weakening of motor skills such as the ability to walk and intellectual functions. Loss of sensation in the extremities, incontinence, seizures, blindness, paralysis, inability to speak, and hearing loss is also observed. In due course patient lose consciousness of their surroundings and become unresponsive [R Macfaul *et al.*, 1982]. The signs of this disorder are most easily confused with similar symptoms of other degenerative nerve conditions such as Krabbe disease, childhood disintegrative disorder, and Gaucher's disease.

Some cases with partial deficiency of enzyme arylsulfatase A and no evident symptoms of MLD are also reported. This partial deficiency of enzyme ARSA is known as ARSA pseudo-deficiency. In case of Pseudo-deficiency, low enzyme activity (5% to 20% of normal controls) is reported but sulfatide is processed normally so MLD symptoms do not exist. [Arvan L Fluharty,2011].

3.1 GENETICS

Metachromatic leukodystrophy (MLD) is inherited in an autosomal recessive pattern. If both copies of the gene in each cell have mutations than probability to have this disorder is quit soaring. If individual carry one copy of the mutated gene than he/she typically do not show signs and symptoms of the MLD. The inheritance probabilities of Metachromatic leukodystrophy per birth are as follows:

- Case 1 : When One parent is affected and one is free of MLD
    - ✓ 0% children will have MLD because of normal gene is inherited from other parent.
    - ✓ 100% children will be carriers of MLD
- Case 2: When One parent is a carrier and the other is free of MLD
    - ✓ 50% children will be carriers of MLD
    - ✓ 50% children will be free from MLD

❖ Case 3: When Both parents are carriers:

- ✓ 25% children will be affected with MLD.
- ✓ 50% children will be carriers of MLD.
- ✓ 25% children will be free from MLD.

3.2 EPIDEMIOLOGY

Epidemiological surveys shows the incidence of metachromatic leukodystrophy per 100 000 live births varied from 0.6 to 2.5 [Ługowska A *et al.,* 2011]. The incidence of metachromatic leukodystrophy is estimated to occur in 1 in 40,000 to 1 in 160,000 individuals worldwide. The epidemiological studies revealed the non-random geographical distribution of metachromatic leukodystrophy prevalence. The increased frequency of MLD observed in Habbanite community (a small group of Jews who immigrated to Israel from southern Arabia), in which consanguineous marriages are relatively common (genetically isolated populations), has reported an incidence of metachromatic leukodystrophy of 1.3% and a 17% carrier frequency (Zlotogora *et al*.,1980).

The geographical variability of metachromatic leukodystrophy is shown in **Table 1**. This gives an impression that the prevalence of metachromatic leukodystrophy is highly variable. However, it should be noted that ost of the population-based studies have shown metachromatic leukodystrophy incidence being in range from 1:40,000 to 1:160,000.

Table1: Geographical variability of Metachromatic Leukodystrophy in some countries/communities.

Countries/communities	**Prevalence of MLD**	**Reference**
Northern Sweden	1:40000	Gustavson and Hagberg 1971
United States	1:100000	Kolodny and Fluharty 2001; Bonkowsky *et al* 2010
Habbanite community	1:75	Zlotogora *et al.,* 1980
Navajo Nation	1:2500	Genetics Home Reference. Reviewed September 2007
Arab groups in Israel	1:8000	Genetics Home Reference. Reviewed September 2007

3.3 CLINICAL MANIFESTATIONS:

Metachromatic Leukodystrophy (MLD) is a neurometabolic, autosomal recessive disease. It is a rare disorder and the diagnosis is often missed or delayed because it is not the first cause that comes to mind when a patient presents with symptoms.MLD is usually classified according to the age of onset:

- Late infantile
- Early juvenile
- Late juvenile
- Adult forms.

The late infantile and early juvenile forms are the most common in nature.

3.3.1 LATE INFANTILE:

This form becomes evident in the second year of life, usually at 15–24 months. Hagberg subdivided the clinical course into 4 stages (Hagberg *et al.*, 1962).

- **Stage I** , Sign and symptoms are stuble and easily overlooked during the first stage, which starts between 15-24 months of age. A child who has previously learned to walk becomes unsteady and requires support to stand. Hypotonia and flaccid weakness and occur in the legs or all 4 limbs. Muscle stretch reflexes may be diminished or absent. This stage lasts from a few months to a year or more.
- **Stage II**, In this stage child experienced problem to stand. However child can sit but no longer stand; cognitive deterioration begins; ataxia and nystagmus are noted; spasticity develops in the legs and possibly the arms and speech is also impaired due to a combination of dysarthria and aphasia. Intermittent pain in the arms and legs may also be reported. This stage may last only a few months
- **Stage III**, Quadriplegic and bedridden is characteristic features of this stage. Difficulty in maintaining normal breathing and dystonic posturing also observed .Speech is no longer distinct but the child may still be able to smile and respond.

- **Stage IV,** This stage lasts from a few months to several years. The child became deaf, blind, speechless and volitional movement.

3.3.2 EARLY JUVENILE

If age of onset is before 6 years of age it is characterized as early juvenile stage.
characteristic features of this form are :

- Impaired school performance
- Behavioral disturbances, such as confusion or bizarre behavior, followed by incontinence.
- Dysarthria , and extrapyramidal dysfunction.

The disease progresses to states that correspond to stages 3 and 4 of the late infantile form (Gordon 1978). Late infantile and early juvenile cases are the most frequent, having a more severe phenotype and rapid progression.

3.3.3 LATE JUVENILE

In this case age of onset is marked between 6 years of age to 12 years of age. The late juvenile form progresses more slowly than does the late infantile and early juvenile variants. Duration of the illness may be longer than 20 years (Mahmood *et al.,* 2010). In late juvenile cases, cognitive difficulties generally precede gait disturbances.

3.3.4 ADULT FORMS

The adult form age of onset can range from 12 to over 70 years. In the adult form, behavioral and intellectual changes are often the presenting symptoms, followed by spastic paraparesis and incontinence [Hyde TM,1992, Aicardı` J (ed),1998].
Other characteristic features of this form are as follows:

- Manifesting as a change in personality
- Poor school or work performance.
- Anxiety and emotional liability.
- Poor memory, and disorganized thinking
- Depression.
- Schizophrenia-like psychosis,

❖ Feelings of depersonalization etc

Table 2: Natural history of Metachromatic Leukodystrophy (MLD) subtypes

Source: GeneticsIN Medicine, Volume 13, Number 5, May 2011

	Late infantile type	**Juvenile type**	**Adult type**
General	40–60% of MLD cases, onset 6 months to 4 years, uniform disease course, and death within 5 years of onset.	20–35% of MLD cases, onset 4 to 16 years, more variable disease progression, and death within 10–20 years of onset.	15–25% of MLD cases, onset following puberty, variable velocity of progression, death within 5–20 years of onset
Initial symptoms	Peripheral neuropathy, muscle weakness, appendicular hyptonia, and hypo- or areflexia	Decline in school performance, emotional, behavioral disturbances, dysarthria/ataxia, and hyperreflexia	Decline in school/job performance, emotional lability, disorganized thinking, and hallucinations /delusions
Subsequent symptoms	Mental regression, visual and auditory impairment, loss of ambulation, evolution to hypertonia, dysarthria/ ataxia, bulbar paresis/dysphagia, and seizure disorder (25%)	Mental regression, optic atrophy, loss of ambulation, spastic quadriplegia, incontinence, bulbar paresis/dysphagia, and seizure disorder (50%)	Clumsiness, incontinence, spastic quadriplegia, choreiform movements, dystonia, bulbar dysfunction, and seizures rare
End stage	Vegetative state, complete loss of interaction, and death from aspiration pneumonia	Vegetative state and decerebrate posturing	Vegetative state

3.4 SULFATIDE BIOCHEMISTRY

Sulfatide accumulation is a major concern in case of Metachromatic Leukodystrophy. So in order to have a deep insight of the work, it is necessary to have an overview about sulfatide. The term sulfatide which is actually sulfur containing glycolipid is coined by Thudichum. 87.5% of the total lipids of myelin are sulfatide [Norton and Poduslo, 1982]. In 1933, Blix found that sulfatide is a mixture of equimolar amounts of cerebronic acid, sphingosine, galactose and sulfate [von Figura *et al*., 2001]. In sulfatide, an ester linkage is found between C-3 hydroxyl of galactose and sulfate [Stoffyn and Stoffyn, 1963]. C-18 sphingosine is the base of sulfatide [Stoffyn, 1966]. High proportion of long chain fatty acids and of fatty acids having 2-hydroxy groups are found in sulfatide and galactocerebroside. These two glycolipids constitute a major portion of 2- hydroxy fatty acids found in brain. After brain, kidney is second in relative abundance of sulfatides, but sulfatide concentration in kidney is around only 10% of the brain.

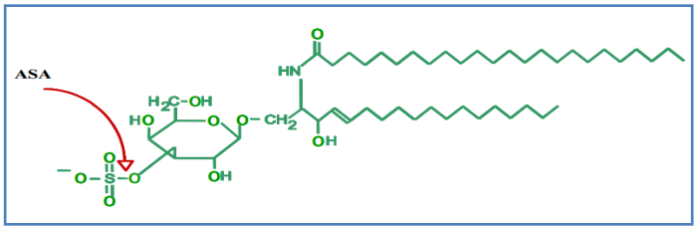


Figure 2: Chemical structure of Cerebroside 3-sulfate (Sulfatide). Sulfatide is a sulfate ester of cerebroside. Sulfatide has a sphingosine back bone with 3-O–beta–sulfo-D-galactose head group.

Synthesis of sulfatide involves sulfation of galactosylceramide (also designated as galactocerebroside) by a reaction with 3'-phosphoadenosine -5'-phosphosulfate (PAPS). Microsomal sulfotransferase catalyses the reaction [Farrell, 1974]. Biosynthesis of galactosylceramide which is the precursor of sulfatide is catalysed by UDP galactose : ceramide galactosyltransferase (CGT). Sulfates galactosylceramide to sulfatide is coverted with the help of 3'-phosphoadenosine-5'-phosphosulfate-cerebroside sulfotransferase (CST). During myelination its synthesis is maximum and in adults it proceeds more slowly. Microscopically, in affected cells from MLD patients, sulfatide is visualized as weird shaped massive inclusions.

Hydrophilic and hydrophobic interactions are found in sulfated glycolipids. Electrical neutrality of membranes is maintained through their anionic charge interactions with inorganic cations or organic amines. Sulfatide is bound to myelin basic protein (MBP) and proteolipid protein (PLP) by strong ionic interactions as it is located at the surface of the myelin membrane. Insulation feature of membranebilayer is due to galactosylceramides and sulfatides [Arvanitis *et al*., 1992; Norton and Poduslo, 1973; Vacher *et al*., 1989].Sulfatides works as a cofactor for Na/K ATPase in active sodium transport [Rintoul and Welti, 1989]. Sulfatides have very high affinity for various cellular adhesion molecules like laminin, thrombospondin, tenascin R etc. Sulfatides The anticoagulant activity in serum is also due to sulfatides.

3.5 MOLECULAR KNOWLEDGE ON MLD

Mutations in the ARSA gene may lead to the deficiency of ARSA causing MLD. The ARSA gene is located on human chromosome 22 at location 22q13. It is about 32 kb long with a coding sequence of 1521 bp. Eight exons encoding the 507 amino acid enzyme are present in the ARSA gene[Stein et al, 1989].

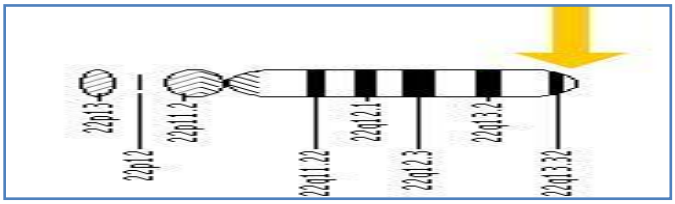


Figure 3: Showing chromosomal location of ARSA gene on the long (q) arm of chromosome 22 at position 13.33.

The ARSA gene is transcribed into three mRNA species, a major species of 2.1 kb, and two minor species of 3.7 and 4.8 kb. More than 90 missense mutations and polymorphisms have been identified in the ARSA gene [Kelly J Perkins *et al.,* 2005]. The most common mutation causing MLD, 459 +1 G--A, is a G--A transition destroying the splice donor site of exon 2. This mutation is associated with a severe phenotype of MLD and has never been found in the homozygous state in patients with juvenile or adult MLD [Joel Zlotogora *et al.*, 1994]. However, most common mutations in the general population: the IVS2 + 1G> A mutation which associated with the late infantile form, the P426L mutation is frequently

associated with the juvenile form, I179S mutation is associated with adult form and pseudodeficiency: N350 and 1524 +95 A → G (poly A-) [Clouter-Mackie & Gagnier, 2003].

Till date, total eight MLD-causing ARSA alleles have been characterized [Gieselmann *et al*. 1991] which can be divided into two groups. One of them encode for low residual enzyme activities and in other alleles is associated with the absence of enzyme activity. Genotype-phenotype correlation study has been done in reference to distribution of these alleles among patients. This study reveals that patients with two alleles associated with no residual activity suffer from the severe, late-infantile type of the disease and those with one or two alleles with low enzyme activity have the intermediate juvenile or mild adult type of the disease, respectively [Polten *et al*., 1991]. Biochemical data have been presented that support this genotype-phenotype correlation [Kappler *et al*., 1991; Leinekugel *et al*., 1992].

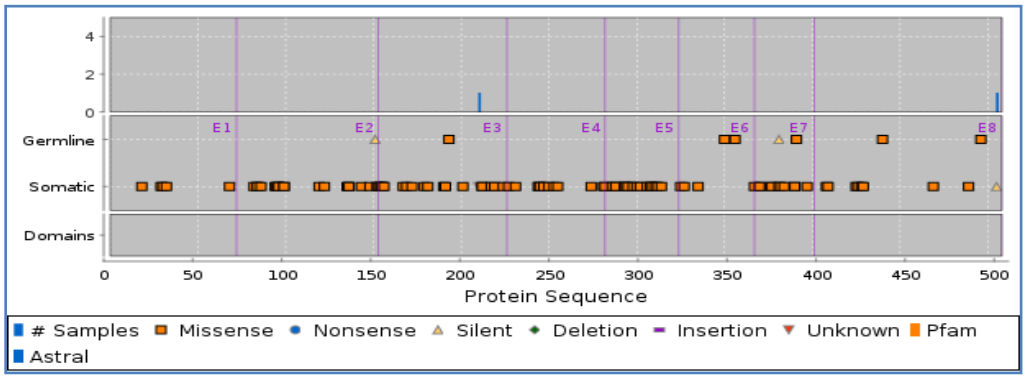


Figure 4: Distribution of mutations in ARSA gene.

Source: RCG Database http://rcgdb.bioinf.unisb.de/MutomeWeb/SimpleQuery?query=ARSA#space_5

MLD is found to be mainly caused by somatic mutations in ARSA gene. There are around 98 somatic mutations in ARSA reported and no germline mutations are found till date.

3.6 ARSA FUNCTION

The principle function of ARSA enzyme is the hydrolysis of sulfatides commonly called cerebroside-3-sulfate or 3-O sulfogalactosylceramide in galactocerebroside and sulfate. The successful working of ARSA necessarily depends on the presence of saposine B which forms a complex with the substrate [Ilhem Barboura *et al.*, 2010]. Myelin sheath is made up of membrane lipids like galactosyl-3-sulfate ceramide (cerebroside sulfate or sulfatide) and

ARSA catalyzes the first step in the degradation pathway of these membrane lipids. Insufficiency of this enzyme causes the storage of the substrate in lysosomes and an increase of sulfatide in the myelin membranes of affected individuals. Oligodendrocytes and Schwann cells of the nervous system are mainly effected due to accumulation of sulfatides but storage in many other organs is also reported[Joachim Kreysing *et al.*,1993] .

Housekeeping enzyme ASA'S expression can be seen in almost all tissues. Liver, placenta and urine has been reported as the major sources for the enzyme. The enzyme has a stumpy isoelectric point.

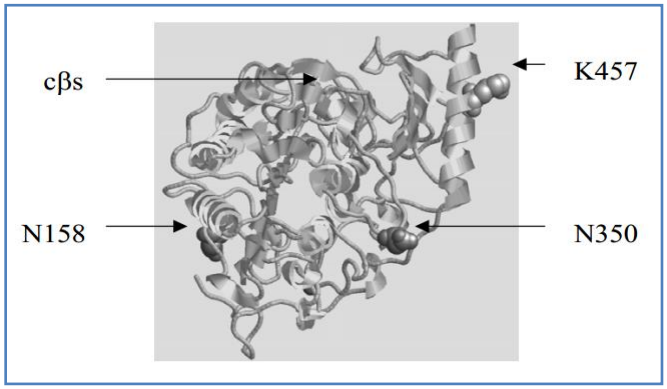


Figure 5: Three dimensional structure of housekeeping enzyme-ASA. Synthesized as a 62-kDa polypeptide, ASA consists of three N-linked oligosaccharideside chains,from which two (N-158 and N-350) are accessible by phosphotransferase and are needed for the activity of ASA. The N-terminal helix of the enzyme contains lysine which is represented here as K457 and Cβs represents the central β-pleated sheet [Yaghootfam *et al*., 2003].

Sulfatides are the major physiological substrates present in various sulfates broken up by ASA. Other than this, lactosyl ceramide 3-sulfate, seminolipid and psychosine sulfate are three major types of 3-O galactosyl sulfates cleaved by ASA.

ASA contains N-oligosaccharide side chains at each of the three potential N-glycosylation sites because of the hydrolyses of signal peptide in the lumen of the endoplasmic reticulum. ASA is recognized as lysosomal enzyme by phosphotransferase as it enters into the Golgi apparatus in fully folded form [von Figura et.al., 2001]. The presence of mannose-6-phosphate residues on the first and the third oligosaccharide side chains and the ineffective phosphorylation of only second N-glycosylation site has been reported after the in-depth examination of the three N-linked oligosaccharide side chains of ASA.

3.7 MYELIN AND SCHWANN CELLS

3.7.1 MYELIN

Oligodendrocytes and Schwann cells are the chief factories producing highly specialised plasma membrane i.e Myelin. Myelin acts as an insulator by wrapping itself around axon and thus is responsible for the fast axonal transmission of the action potential. One of the reasons behind proper functioning of the nervous system is the specific molecular organisation of myelin. Cholesterol and the glycosphingolipids, galactosylceramide and sulfatide forms the major portion of the lipid composition in multi lamellar structure of Myelin. Importance of these lipids in the formation and maintenance of myelin has been clinically proved [Bosio *et al*., 1996; Coetzee *et al*., 1996; Honke *et al*., 2002)].

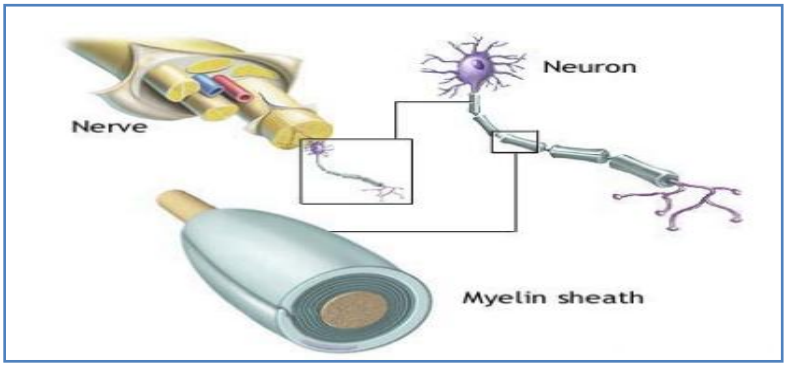


Figure 6: Neuron and Myelin. Diagrammatic representation of the nerve cell neuron and myelin insulating axon.

Source: http://www.nlm.nih.gov/medlineplus/ency/images/ency/fullsize/9682.jpg

Neurological disorders usually occur due to changes in myelin architechture which may be a result of a hereditary or acquired disease. Alterations in genes encoding myelin protein or myelin lipid metabolism may cause Genetic diseases[Anderson *et al*., 1998; Nave, 1994; Werner *et al*., 1998] . Disturbance in myelin lipid metabolism particularly causes two major lysosomal storage diseases - Krabbe disease and MLD [Gieselmann, 2003; Suzuki, 2003; von Figura *et al*., 2001; Wenger *et al*., 2001]. ßgalactocerebrosidase deficiency causes Krabbe disease, also known as Globoid cell Leukodystrophy and results in accumulation of galactosylceramide. One of the treatments of MLD may include the prevention of demyelination. This can help in the treatment of the disease to a great extent.

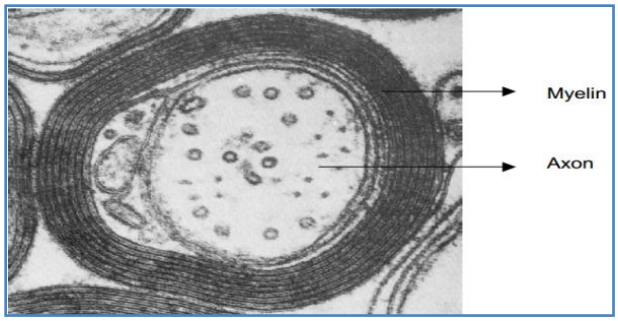


Figure 7: Ultrastructure of myelin. Electron micrograph of axon enwrapped by myelin.
Source: www.cytochemistry.net

3.7.2 SCHWANN CELLS (SCS)

Till date, an appropriate model for cell culture which can store sulfatides is unavailable and it is still unclear how at the molecular level the metabolism of a cell is affected by sulfatide storage [Gieselmann *et al*., 2003]. However only minute information is available for the demyelination machinery and the abnormal metabolic pathway of the causative molecule. Deficiency of particular lysosomal enzyme which causes metabolic defects in the cell can be cross corrected by an exogenously applied enzyme and this has also been reported by certain previous cell culture experiments [Neufeld, 1991]. Degradation of the substrate present in the lysosomes is caused by the internalization of the enzyme by mannose 6-phosphate receptors into the lysosomal compartment. In glial and neuronal cells, endocytosis of lysosomal enzymes is dependent on mannose 6-phosphate [Schluff *et al*., 1998; Stewart *et al*., 1997].

SCs, which sustain neurons, play an important role in the development of axons, its renaissance and myelination, are the principal accessory cells of the peripheral nervous system. Peripheral nervous system's SCs and central nervous system's oligodendrocytes form Myelin. It helps in the salutatory nerve conduction [Guenard *et al*., 1994]. Demyelinating and remyelinating lesions and characteristic lamellar inclusions in SCs and macrophages were reported in the low ASA activity in leukocytes and fibroblasts culture of a 38 year old man [Fressinaud *et al*., 1992]. Later, for MLD patients with peripheral neuropathy, similar reports were published [Coulter-Mackie *et al*., 2002; Hansen *et al*., 1994; Martinez *et al*., 1975].

In many laboratories, pure neonatal SCs are prepared using numerous methods. Till date, spontaneous transfection or immortalisation using oncogenes more than fifteen SC lines have been produced [Bolin *et al*., 1992; Boutry *et al*., 1992; Chen *et al*., 1987; Goda *et al*., 1991; Jirsova *et al*., 1997; Li *et al*., 1996; Porter *et al*., 1987; Ridley *et al*., 1989; Tennekoon *et al*., 1987; Thi *et al*., 1998; Toda *et al*., 1994; Watabe *et al*., 1995; Watabe *et al*., 1990). In each cell line, the extent of phenotype expression and differentiation differ from one other [Hai *et al*., 2002]. SCs from ASA KO and wild type (WT) mice need to be isolated for an in vitro cell culture system for MLD and this became possible with the technology which was established using above mentioned reports.

3.8 PROSAPOSIN

In humans PSAP gene encodes for Prosaposin (also known as PSAP) protein. Prosaposin is a highly preserved glycoprotein and it plays an important role in numerous biologically significant functions like the development of the nervous system and the reproductive system.

When Prosaposin is broken down, it produces four smaller proteins : Saposin A, B, C and D. Saposin is an contraction for Sphingolipid Activator **P**r**O**[**S**]te**IN**s [Morimoto S *et al.*, 1990]. Cysteine residues and glycosylation sites are at almost similar places in each domain of the prosaposin and each domain is found to be approximately 80 amino acid residues long.. PSAP not only carries out the neurotrophic activities but also serves as an integral membrane protein and secretory protein.

3.8.1 FAMILY MEMBERS

- **Saposin A**: Before isolation, Saposin A was considered as an N-terminal domain in the prosaposin cDNA. It has been reported that enzymatic degradation of 4-methlyumbelliferyl-β-glucoside, glucocerebroside, and galactocerebroside is motivated by Saposin A [Morimoto S *et al.*, 1989].

- **Saposin B**: It was the first small prosaposin protein to be discovered. Arylsulfatase A hydrolyses sulfatides but for the first time it was found that ARSA requires a heat stable factor Saposin B for this hydrolysis. It is recognized by a list of names like sulfatide

activator protein, sphingolipid activator protein-1 (SAP-1), dispersin, GM1 ganglioside activator, and nonspecific[O'Brien JS, Kishimoto Y, 1991]. It has been experimentally shown that many enzymes get activated when saposin interacts with the substrates instead of interacting with enymes themselves.

- **Saposin C**: Glycosylceramidase hydrolyses glycocerebroside and galactoslyceramidase degrades galactocerebroside with the help of Saposin C. This was the second small prosaposin protein to be discovered.
- **Saposin D**: Till date, not much is known about Saposin D, more exploration needs to be done. cDNA sequence of prosaposin lead to the prediction of this protein. Saposin-D is a specific sphingomyelin phosphodiesterase activator [Kishimoto Y, 1992].

3.8.2 STRUCTURE

Eighty amino acid residues long every saposin has six uniformly positioned cytesines, two prolines, and a glycosylation site (two in saposin A, one each in saposins B, C, and D) [Kishimoto Y, 1992]. Saposins which are enormously dense and tightly disulfide-linked molecules have intense heat-stability, plenty of disulfide linkages, and resistance to most proteases. α-helical structure of each saposin which is seen in all (especially with the first region), is maximal at a pH of 4.5 and thus is important for stimulation[Kishimoto Y, 1992]. The first 24 amino acids of N-terminal of saposin gives it a β-sheet configuration [O'Brien JS, Kishimoto Y , 1991].

The PSAP gene is located on human chromosome 10 at location 10q21.The molecular location of this gene on chromosome 10 is found at base pairs 73,576,054 to 73,611,081. In particular, on chromosome 10 PSAP gene is located from base pair 73,576,054 to base pair 73,611,081.

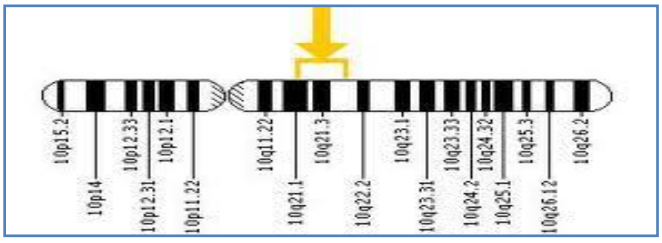


Figure 8: Showing chromosomal location of PSAP gene on the long (q) arm of chromosome 10 between positions 21 and 22

To understand the origination of the locus that codes the four highly conserved sphingolipids activator proteins or saposins, sequence analysis of chromosomal DNA which helped in characterizing the gene for prosaposin, was carried out. It comprises of 13 exons ranging from 57 to 1200 bp and introns range from 91 to 3812 bp in length. Among the four saposins, the regions encoding saposins A, B, and D each have three exons and region encoding saposin C have only two. The 3′ untranslated region, the four mature prosaposin proteins and carboxy terminus of the signal peptide were present in the sequence which was used. [Efrat Gavrieli Rorman *et al.,* 1992].

3.8.3 FUNCTION

The cell's recycling centers called lysosomes are the primary home for saposins. Sphingolipids (fatty substances) are broken down by saposins with the help of lysosomal enzymes.

- **Saposin A-D:** In lysosomal section, the catabolism of glycosphingolipids with tiny oligosaccharide groups is carried out primarily with the help of Saposins A-D
- **Saposin B:** Saposin B protein breaks down sphingolipids with the help of several enzymes. Saposin B protein when coupled with enzyme arylsulfatase A carries out tts most critical biological role. The nervous system's white matter which comprises of nerve fibers insulated with myelin, contains huge amount of sphingolipids and a subgroup of these sphingolipids called sulfatides further is cleaved by enzyme Saposin B. Transportation of lipids to the external surface of the cell may be carried out by Saposin B.
- **Saposin C:** Enzyme beta-glucocerebrosidase break down another sphingolipid glucocerebroside with the help of saposin C protein.
- **Saposins A and D:** These are also found to deal with sphingolipids.

3.9 SYMPTOMS

1. HYPERTONIA

Resting muscle tone: The muscles of the body are normally in a state of tension or resistance. Brain modulates this state by sending messages to the muscles which are contracting and those which are relaxing. Resting state of tension is required for a person to move and

respond. Abnormal tone occurs due to disrupted or incorrect messages from the brain. Damage to the brain or spinal cord results in this abnormal tone. Hypertonic muscles, also referred to as Spastic, feel hard or stiff. In this state the muscles and joints feel resistant to movement.

2. BEHAVIOR PROBLEMS

Behavioral problems often occurs in the adult form of Metachromatic Leukodystrophy and this is the first sign which appears during the teenage years or later. Symptoms may include alcoholism, drug abuse, or difficulties at school or work, anger & defiance, arguing & fighting, self-esteem, power, struggles, substance abuse & risky behavior, disrespect, aggression, outbursts & temper tantrums.

3. DECREASED COGNITIVE FUNCTION

Cognitive function describes a person's alertness and orientation, memory and attention span. In case of MLD, person's ability to interact with his physical environment goes down and leads to a state of confusion. Memory which is the most important cognitive ability is also diminished. In terms of attention span, patients' ability to make decisions, reasoning, his knowledge and use of language also gets declined.

4. INCONTINENCE

Incontinence is inability to control the excretion of urine or feces. Urination system relies on normal function in pelvic and abdominal muscles, diaphragm, and control nerves. In case of babies' nervous systems are too immature for urinary control. Incontinence in adults is assumed to be a disorder. In case of MLD, incontinence is seen due to malfunctioning of nervous system.

5. IRRITABILITY

Inconsolability/Irritability makes young ill children fussy, whiny, and fretful. This may be an early symptom of any serious problems. Irritability does not indicate any specific illness, it only shows that something is wrong with the patient. Children with MLD become increasingly agitated and irritable as it gets older.

6. SEIZURE

Disorders associated with central nervous system often results in changes in the physical behavior of the patient and this change is referred to as seizure s. In MLD, seizure causes the patient's muscles to contract and relax repeatedly. There are many different types of seizures and symptoms may vary from person to person.

7. DIFFICULTY WITH SWALLOWING

Mouth, throat area, and esophagus (the muscular tube that moves food to the stomach) are the majorly involved in the complex act of Swallowing and functioning of these body parts is controlled by different nerves and muscles. Swallowing is basically a voluntary control, but it also involves a lot of involuntary control. In MLD, patient finds difficulty with swallowing food as it gets stuck in the throat, or from the neck down to just above the abdomen behind the breastbone (sternum).

8. LOSS OF NORMAL MOVEMENT AND COORDINATION

This is the first symptom which is often seen in early and late infantile forms of MLD. In this, the most significant loss of movement skills (motor skills) usually starts between 12 to 18 months age. The first signs often include a decrease of body control and especially decrease in the ability to crawl or walk normally. At first, this loss of abilities occurs rapidly and then continues more gradually.

3.10 DIAGNOSIS

In case of Metachromatic leukodystrophy (MLD), diagnosis is often missed or delayed because it shares many common symptoms with other prone diseases and MLD itself is a rare disorder. Cerebral palsy or other causes of developmental delay often leads to a state of confusion and makes it difficult to diagnose MLD in young children. Batten disease, attention-deficit hyperactivity disorder (ADHD), or adolescent/puberty-related mood/behavioral changes also cause the same problem in case of older children. In adults, psychological conditions are often suspected. In case of MLD, a variety of tests must be performed to correctly diagnose the disease.

1. MRI

First indicative test that is performed is magnetic resonance image (MRI) of the brain. In MRI, demyelination is detected around the nerve cells which is basically the presence or absence of myelin sheaths around the nerve cells.

2. BLOOD AND URINE TEST

It is a formal diagnosis test for MLD because MRI test is not absolute in nature and abnormal MRI test results may be due to other reasons. It mainly consists of two tests :

- Measuring blood arylsulfatase A (ASA) enzyme levels
- Measuring urine sulfatide levels

Low levels of ASA are also found in a condition called ASA pseudodeficiency so it is not considered as a definitive test for MLD. So the diagnosis of MLD must be confirmed with other tests, including urinary sulfatide measurement. An increased amount of sulfatide in the urine indicates that the ASA enzyme is not breaking down sulfatides.

3. NERVE CONDUCTION VELOCITY TEST

The diagnosis of MLD may also include an evoked potential test and a nerve conduction velocity test. In this test, measurement of nerve stimulation due to brain's activity shows a delayed or impaired response because of the demyelination that occurs in MLD. In case of nerve conduction velocity test, the speed of signals which travels along the nerves is measured. It only measures electrical signals along the nerves in the arms and legs. A slower-than-normal signal may indicate loss of myelin, suggesting MLD.

4. DNA TEST

Diagnosis of MLD also includes molecular analysis, also called DNA testing. In this test, DNA profiling especially for ARSA gene is done and compared with normal one.With the help of this test, ARSA allele, responsible for MLD can be identified. This can be useful if the patient has a family history of MLD.

5. SCREENING TEST

Screening test is more popular to diagnose many genetic disorders nowadays. In this test disease can be diagnosed with the help of single drop of blood. It is more efficient test as compared to other one but it is not available for diagnosis of MLD. It is still in development phase.

3.11 TREATMENT

There is no effective treatment or cure available for MLD. Present care focus on treating symptoms and preserving the patient's quality of life. Currently treatment is only limited to pain and symptom management. However, there are two options available for Presymptomatic late infantile MLD patients, as well as those with juvenile or adult MLD, displaying mild to moderate symptoms.

- Bone marrow transplantation
- Stem cell transplantation

Even these two methods are in developmental phase and under investigation to see if it may slow down progression of the disease or stop its progression in the central nervous system.

- Gene therapy,
- Enzyme replacement therapy (ERT),
- Substrate reduction therapy (SRT), and
- Potentially enzyme enhancement therapy (EET)

These four methods may have future potential to cure or better management of the MLD. These methods are in process of investigation.

3.12 PROGNOSIS

Metachromatic Leukodystrophy is a progressive disorder and gets worsened over time. Eventually, patient loses all mental and muscle functions. Life span is variable for every individual and depends on at what age the condition has started. Course of the disease usually runs three to twenty years or more.

3.13 PREVENTION

If person have a family history of MLD, Reproductive planning in conjunction with genetic counseling is principle method for the prevention of metachromatic leukodystrophy. Carrier of MLD may be identified with the help of arylsulfatase A deficiency assay [Zlotogora *et al.* 1980], and affected fetus can be identified prenatally [Eto *et al.,* 1982]. Prenatal diagnosis and carrier identification is necessary to distinguish metachromatic leukodystrophy from the benign pseudodeficiency condition. Mutation analysis provides the most accurate method for prenatal diagnosis and carrier identification and helps in preimplantation genetic diagnosis [Braude *et al.,* 2002].

4. METHODOLOGY

In an effort to understand the interaction between various genes and gene pathways which are involved or getting affected in Metachromatic Leukodystrophy, the gene-gene interaction network map was built using Cytoscape and the network was analyzed using various plug-ins in Cytoscape. The genes responsible for MLD from the various databases were searched. Journals with published literature and reports on MLD were also referred. There are around 70 genes interacting with the candidate genes (ARSA and PSAP), found through literature and various database searches. In this study fallowing software and their corresponding applications was used to create and analyze the protein interaction network for MLD.

4.1 CYTOSCAPE

Cytoscape is an open source software platform for integrating and visualizing bimolecular interaction networks with high-throughput expression data, and integrating these networks with annotations and other state data into a unified conceptual framework. Cytoscape was designed at the Institute of Systems Biology in Seattle in 2002. Now, it is developed by an international consortium of open source developers [Paul Shannon *et al.*, 2003]. Cytoscape was initially made public in July, 2002 (v0.8).As of March 2013, there are three versions of Cytoscape: Cytoscape 2.x, Cytoscape 3.x, and cytoscape.js. Cytoscape 2.x is a stable, production version of Cytoscape. In current study we have used Cytoscape version 2.8.3.

Cytoscape's base is in Systems Biology, where it is used for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. It is a project dedicated to building open-source network visualization and analysis software. A software "Core" provides basic functionality to layout and query the network and to visually integrate the network with state data. The Core is extensible through a plug-in architecture; align rapid development of additional computational analysis and features [Raman 2010].

A network (graph), which includes genes, proteins, and molecules which are represented as nodes and interactions represented as links, i.e. edges between nodes, is the central organizing metaphor of Cytoscape. Cytoscape is continuously progressing, both in core feature set and developer community. Molecular networks can be effectively visualized by

using various Cytoscape applications. In order to reduce the complexity of network, subsets of nodes and edges can be selected according to variety of criteria.

4.2 MiMI PLUG-IN

This plug-in is a java based interactive visualization tool for analyzing protein interactions and their biological effects. It is freely available in public domain. It integrates Cytoscape with MiMI database, a database that uses an intelligent deep-merging approach to integrate data from multiple protein interaction database. The MiMI database has data on 119,880 molecules, 330,153 interactions, and 579 complexes. The MiMI plug-in communicates with the MiMI database using a PHP server that acts as a database interface. The interface is implemented using stored procedures to modularize the code and avoid embedding of SQL queries in PHP scripts. In this way, the database query is isolated from the Cytoscape MiMI plug-in. This allows changes and upgrades to be made to the database without affecting the MiMI plug-in [Jing Gao *et al.,* 2009]. MiMI plug-in allows to access the integrated molecular data assembled in the database and retrieve graphics which contains protein/gene interactions.

4.3 BiNGO (The Biological Networks Gene Ontology Tool)

Among various Gene Ontology categories, in order to determine which are significantly statistically overrepresented in a cluster of genes or sub-graph of biological network, this java based plug-in is used. The functional themes of a gene cluster on the GO hierarchy can be mapped using BiNGO and by taking advantage of Cytoscape's versatile visualization environment BiNGO returns this mapping as a graph. Sub-graphs of biological networks or list of genes pasted as text can be used as input.

BiNGO features include:

- Assessing underrepresentation or overrepresentation of GO categories
- Hypergeometric or binomial test for overrepresentation
- Batch mode : analyze several clusters simultaneously using same settings
- Multiple testing correction using Bonferroni (FWER) or Benjamini&Hochberg (FDR) correction
- Wide range of organisms

4.4 NETWORK ANALYZER

Molecular interactions through statistical analysis of a network can be studied with the help of this java based visualization plug-in. This plug-in provides overall network topology via computing following parameters:

- Average number of neighbors
- Number of connected pair of nodes
- Node degrees
- Average clustering coefficient
- Topological coefficient
- Shortest path lengths
- Betweenness centrality
- Closeness centrality
- Stress centrality

It plots distribution curves based on Correlation coefficient and R-square value and these results can be saved as images or text files.

4.5 ClusterONE (Clustering with Overlapping Neighborhood Expansion)

ClusterONE, available as a standalone command-line application, or as a plug-in to Cytoscape**,** is a graph clustering algorithm. It is mainly useful for detecting protein complexes in protein-protein interaction networks with associated confidence values. It has the ability to handle weighted graphs and it can also generate overlapping clusters.

ClusterONE works by "growing" dense regions out of small seeds (typically one or two vertices), driven by a quality function called cohesiveness. The network quality can be assessed by the number of internal edges divided by the sum of the number of internal and boundary edges. This quality measure is based on the fact that a well-defined network should have many internal edges and only a few boundary edges. This quality measure is known as cohesiveness.

ClusterONE search for groups of high cohesiveness. Greedy algorithm is used to achieve this objective. Single seed vertex (a small set of vertices that are strongly bound together), is selected and step by step group is extended with new vertices addition so that the newly

added vertex always increases the cohesiveness of a group as much as possible. If removing a vertex from the group increases its cohesiveness, it is also permitted. When it is not possible to increase the cohesiveness of the group by adding another external boundary vertex or removing an internal boundary vertex, the process terminates.

4.6 CYTOHUBBA

Due to the availability of plenty of high throughput experimental data, one of the critical issues of post genome era is to reconstruct the proteomic and genomic interacting networks. Identification of important nodes or hubs in these interactomes is a key to explore significant characteristics inside biochemical pathways or complex networks. These essential nodes/hubs may serve as potential drug-targets for developing novel therapy of human diseases [Chung-Yen Lin *et al.*, 2008].

We can explore important nodes/hubs and fragile motifs in an interactome network by several topological algorithms such as:

- Degree
- Edge Percolated Component (EPC)
- Maximum Neighborhood Component (MNC)
- Density of Maximum Neighborhood Component (DMNC)
- Maximal Clique Centrality (MCC)
- Bottleneck (BN)
- EcCentricity
- Radiality
- Betweenness
- Stress

Since the analysis methods of hubba are based on topology, it can also be used on other kinds of networks to explore the essential nodes.

4.7 VENN AND EULER DIAGRAMS

This is a java based data visualization tool which provides a diagram view and a detailed view for comparing two or more Cytoscape groups at a time. It consists of a symmetrical layout that supports two to four Cytoscape groups. For two and three Cytoscape groups,

Venn diagram views use two and three circles respectively. For four or more than four Cytoscape groups, a symmetrical construction of four ellipses is used. The Euler diagram view uses an area-proportional layout that supports two or more Cytoscape groups. It is similar to a Venn diagram but does not display all possible intersections.

4.8 GENETIC ASSOCIATION DATABASE (GAD)

Human genetic association studies of complex diseases and disorders has compiled in Genetic Association Database. It is freely available in public domain. Summary data extracted from published papers in peer reviewed journals on candidate gene and GWAS studies are assembled in GAD. The objective of database is rapid identification of medically relevant polymorphism from the large volume of polymorphism and mutational data, in the context of standardized nomenclature.

Additional molecular reference numbers and links are also integrated with study data and recorded in the context of official human gene nomenclature. Characteristic feature of this database is that, It is gene centered, each record is a record of a gene or marker. It is freely available, anyone may access this database and anyone may also submit records.

There are three main components of GAD:

- Web interface,
- Perl modules and
- Database, which uses the Oracle RDBMS.

The database has three layers; gene and disease data are organized into a large fact table in a middle layer with dimensional views on the top layer. The bottom layer contains the tools for adding, editing, batch loading and downloading data to and from the database [Kevin G Becker *et al.*, 2004].

4.9 THE DATABASEFOR ANNOTATION, VISUALIZATION AND INTEGRATED DISCOVERY (DAVID)

High throughput genomics, proteomics and bioinformatics approaches have generated a huge amount of data. To extract valuable information from this data, functional annotation of differentially expressed genes is a necessary and critical step in his era. Biological knowledge is distributed in nature and frequently requires researchers to navigate through numerous web-accessible databases gathering information one gene at a time. DAVID has provided a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes.

Characteristic features of DAVID:

- Convert gene identifiers from one type to another.
- Identify enriched biological themes, particularly GO terms.
- Cluster redundant annotation terms.
- Discover enriched functional-related gene groups.
- Visualize genes on BioCarta & KEGG pathway maps.
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list.
- Link gene-disease associations.

Tools:

- Functional Annotation
- Gene Functional Classification
- Gene ID Conversion
- Gene Name Batch Viewer

DAVID Analyze results and the graphical displays remain dynamically linked to primary data and external data repositories, thereby furnishing in-depth as well as broad-based data coverage. The functionality provided by DAVID accelerates the analysis of genome-scale datasets by facilitating the transition from data collection to biological meaning [Dennis G Jr,*et.al*,2003].

In this study, we collected and manually curated over 40-50 review and other articles involved in Metachromatic Leukodystrophy and built a protein interaction network using Cytoscape 2.8.3. Various plug-ins were used to analyze manually curated network. All the associated biological pathways such as metabolic, lysosome-acid hydrolases, glycosphingolipid metabolism etc have been studied. This study was aimed at getting a better insight into Metachromatic Leukodystrophy pathogenesis and identifying the important drug targets for this disorder. Another aim of study was to identify the functional, highly interconnected nodes and bottleneck nodes in the network, identifying the important sub networks in the system could be helpful in understanding the underlying molecular mechanism of Metachromatic Leukodystrophy. The study was also focused on identification of candidate genes with high centrality score and perform their disease and pathway enrichment analysis to find out the disease classes in which these genes are involved and the pathways they are affecting. An additional aspect of study was chromosome enrichment analysis of candidate genes to calculate the distribution of genes across different chromosomes. The objective of this study was to build the network using Cytoscape and analyze the statistical significance of the network, and using its plug-ins, the important nodes (genes) which could be functional, were found.

4.10 CONSTRUCTION AND ANALYSIS OF THE MLD RISK GENES NETWORK

Metachromatic leukodystrophy (MLD) is inherited in an autosomal recessive pattern and is caused by deficiency of enzyme arylsulfatase-A (ARSA), or, more rarely, of its activator protein saposin-B (PSAP). Mutations in the ARSA and PSAP genes result in a decreased ability of enzyme arylsulfatase-A (ARSA), and protein saposin-B to break down sulfatides, resulting in the accretion of these substances in cells. High quality interactions with a confidence score of 0.7 including all the types of associations between the ARSA and PSAP with other genes and proteins were considered. Some of the interacting partners were not given in STRING and hence other databases and literature was thoroughly searched for the possible interactions. Genes extracted from various databases are given in **Table 3** and which are taken from literature are given in **Table 4** with their Pubmed IDs. Other databases such as KEGG, PubMed, BIND, DIP, MINT and BioGRID were also searched. To avoid false

positive results which may have obtained while using text mining procedures, PolySearch [Dean Cheng *et al.*, 2008] and Phenopred [Radivojac P *et al.*, 2008] tools were also used. Thereafter MiMI plug-in in Cytoscape was used to build gene-gene interaction network taking these 70 genes as input. The direct or close interactions of these query genes with the nearest neighbors were found. The interaction data was processed with the help of Oracle 10g database software and were saved in excel sheet which was then imported in Cytoscape 2.8.3 for further analysis. The top 20-30 genes obtained from the scoring were selected and analyzed individually for their association with MLD related scientific evidences in literature. The interaction search was restricted to *homo sapiens* and included all the kinds of experimental as well as predictive interactions. The curation of final database was performed manually to remove the duplicate interactions and uniform isoform notations with unique genes. The final network thereafter obtained consisted of 1050 nodes and 8019 edges. **Figure 9** showing the dialogue box for MiMI plugin.

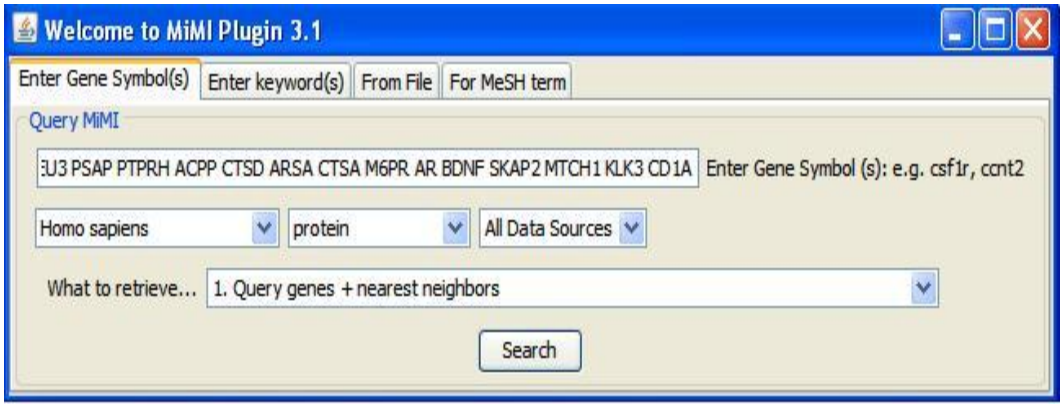


Figure 9: The dialogue box of MiMI plug-in with the query genes and type of information retrieval is shown.

Table 3: Genes known to interact with ARSA and PSAP taken from databases.

GENES	DATABASE
BMPR2, BRCA1, CD1A, CTSL1, SMAD2, SMAD9, SMAD9	HPRD
CETN2, CFTR, CLC4G, CLEC4G, COPS6, PPIC, SGCG	INTACT
GHR, UBC, UBE3A, VCP	PREDICTOME
ACPP, ALAS2, ARSD, ARSE, ARSF, ARSH, ARSI, ARSJ, ARSK, ASNA1, GAL3ST1, KLK3, MAFF, MTCH1, NEU1, NEU2, NEU3, NEU4, SGK223, SKAP2, SLURP1, UGT8, ZBED1	STRING

Table 4: Genes known to interact with ARSA and PSAP with their PMID's.

GENES	PMID
AR	PMID: 17044040
ARSA	PMID: 2320574
ARSB	PMID: 22216298,PMID: 4698279
ARSG	PMID: 18283100
ASAH1	PMID: 11104761
BDNF	PMID: 9748612
CD1B	PMID: 14716313
CD1D	PMID: 14716312
CTSA	PMID: 8206147 ;PMID: 18480170 ;PMID: 10484606
CTSD	PMID: 23108186
DEGS1	PMID: 19471889
GALC	PMID: 2717620, PMID: 21709,PMID: 19267410
GBA	PMID: 8206997
GBAP1	PMID: 17353235
GLA	PMID: 19267410
GLB1	PMID: 2320574, PMID:36575,PMID: 22216298
GM2A	PMID: 20510729
GUSB	PMID:6766092
M6PR	PMID: 17848585
PSAP	PMCID: PMC53725,PMID: 7866401 ,PMID: 2764035
PTPRH	PMID:8311994,PMID: 2320574,PMID: 1689485
SMPD1	PMID: 11856752

SORT1	PMID: 20571055, PMID: 19732768
STS	PMID:15561802,PMID: 4363968
SUMF1	PMID:17080200,PMID: 15962010
TFPI2	PMID: 21943334
TPX2	PMID: 15948238
TSG101	PMID: 22315426
VPS4A	PMID: 12915562

4.11 ANALYSIS OF THE INTERACTION NETWORK

Once the network was obtained for MLD it was analyzed for it topology and biological significance. To confirm the presence of molecular systems structures that may better explain missing linkage in the mechanism of MLD, systems biology approach was used to study the genetic risk gene networks as a whole rather4 than the risk genes individually. Till date only two candidate risk genes (ARSA and PSAP) are known in causing MLD. It was only imperative to study the system using the available bioinformatics resources and tools available. Cytoscape's base is in Systems Biology, where it is used for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. The software core is extensible through a plug-in architecture; align rapid development of additional computational analysis and features [Raman 2010]. In the current study we have used cytoscape version 2.8.3.

Network Analyzer, a java based freely available Cytoscape application for the analysis of various kinds of networks. It can be used for both directed and un-directed kinds of networks. It is a statistical tool to compute various topological network parameters. Parameters computed by this plug-in can be characterized in two groups given below in **Table 5**:

Table 5: List of parameters computed by Network Analyzer

Simple Network Parameters	Complex Network Parameters
Number of connected components	Betweenness centrality
Parameters related to shortest paths	Closeness centrality
Parameters related to neighborhood	Degree distribution
Clustering coefficient	Neighborhood connectivity
	Shortest path
	Stress centrality
	Topological coefficient

These parameters obtained are further discussed in results and discussion section and provided us comprehensively about the topology of the network obtained. The understanding of network topology of a biological network can directly give an insight into various network characteristics. Further cluster analysis was carried out which provided a better topology description of the network including the location of highly connected sub-graphs and/or overlapping modules that usually correspond with relevant biological information. Cluster analysis was done using ClusterONE**.** Cluster ONE strives to discover densely connected and possibly overlapping regions within the Cytoscape network. These dense regions usually correspond to protein complexes or fractions of them. Cluster ONE works by "growing" dense regions out of small seeds (typically one or two vertices), driven by a quality function called cohesiveness. Highly inter connected nodes in the network were obtained suggesting their role in the pathogenesis of the disease. Density threshold was set 0.5 for unweighted networks. The network was un-weighted and the global clustering coefficient was set smaller than 0.1. The node penalty value was set to 2, that excludes the weaker connections in the network and thus were not added to the cluster.

Proteins with special graph (network) characters in an interactome may play critical roles in controlling or regulating cellular responses to a special physiological stimulus. For the present study cyto-Hubba (Hub Objects Analyzer) plug-in was used to explore important nodes/hubs in an interactome based on previous framework, Hub object Analyzer. To identify essential nodes/hubs the protein networks is a way to decipher the critical key controllers inside biochemical pathways or complex networks. These essential nodes/hubs may serve as candidates of drug-targets for developing novel therapy of human diseases.

Several centrality indexes were evaluated using CytoHubba. A normalized centrality score was calculated using ten centrality indexes Betweenness, Bottleneck, DMNC, MNC, MCC, EPC, Eccentricity, Radiality and Stress. To obtain normalized centrality score (Score *I)* following formula was used:

$$\text{Score}I = \sum_{i=1}^{Nc} \left(1 - \frac{\max(Ici) - Ici}{\max(Ici) - \min(Ici)}\right).100$$

Where Ic_i is the values of centrality indexes and i=1...Nc, and is the number of calculated centrality indexes (N= 10). Through this analysis all 1050 genes were ranked according to their score I value.

4.12 STATISTICAL SIGNIFICANCE TESTING OF MLD RISK GENE PROTEIN NETWORK

Biological Network Gene Ontology (BiNGO) [Maere S *et al.*, 2005], a plugin in Cytoscape was used to identify ontologies in the gene protein network. The overrepresented processes were assessed using hypergeometric statistical test. A FDR multiple hypothesis test adjustment was carried out using the Benjamini Hochberg (BH) procedure Benjamin and Hochberg false discovery rate (FDR) correction with a significance level of 0.05. BiNGO analysis identified three ontologies (cellular component, molecular function and biological processes).

4.13 GENE FUNCTIONAL CLASSIFICATION AND FUNCTIONAL ENRICHMENT ANALYSIS USING DAVID

To analyze a biological network, it is essential to concentrate on the larger biological network rather than at the level of an individual gene. DAVID gene functional classification tool was used to explore the functionally related genes together on the basis of their annotation term co-occurrence. The Uniprot IDs of 126 candidate genes were given as input in DAVID. The DAVID gene ID Conversion Tool was used to convert the gene IDs to relevant gene names with biological annotation in the DAVID database. 120 out of 126 genes were recognized and converted to DAVID id from their Uniprot IDs. To control the behavior of DAVID fuzzy clustering, the classification stringency was set to medium for balanced

results [Da Wei Huang *et al.,* 2008]. The Kappa Similarity Term Overlap was set to 4 and Kappa Similarity Threshold was set to 0.35. Six clusters were obtained out of which four were analyzed further with an enrichment score > 1.3. The highly interconnected clusters were further explored by Gene ontology (GO) analysis.

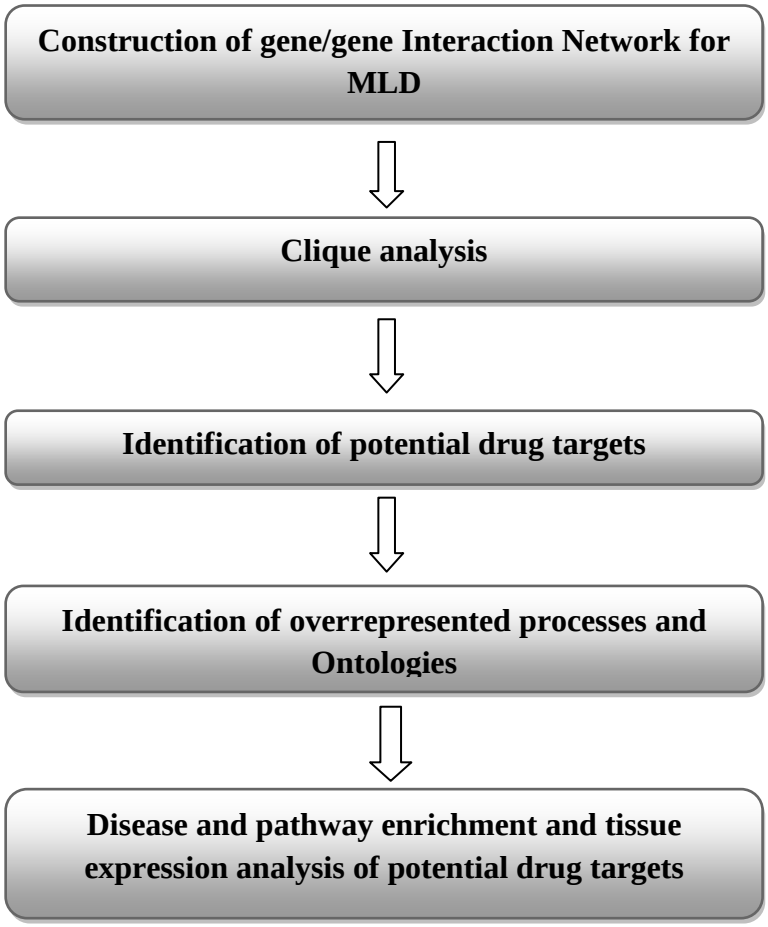
4.14 PATHWAY AND DISEASE ENRICHMENT ANALYSIS

To further analyze the results and explore the functional connections, 153 genes were mapped onto known gene sets. For this purpose DAVID was used to search for enriched KEGG pathways. To carry out the disease enrichment analysis DAVID extensively uses the genetic disease related databases such as Genetic Association Database (GAD) and OMIM. In addition a substantial statistical description in regards to the wide range of enrichment analysis has also been provided using DAVID. The analysis was carried out considering all the genes/proteins involved in the pathogenesis of MLD which were obtained from cluster and hub analysis using Cytoscape as well as proteins/genes which were found through literature.

4.15 TISSUE GENE EXPRESSION AND CHROMOSOME ENRICHMENT ANALYSIS

Gene activation patterns vary widely in complexity. While some genes are known to express in a straightforward and static manner, some are extraordinarily intricate and difficult to predict and model, with expression fluctuating wildly from minute to minute or from cell to cell [Shestopalov *et al.,* 2011]. Tissue gene expression profile and chromosome enrichment analysis have been widely used to study disease states. Tissue-specific gene expression profiles and chromosomal distribution of genes was studied through DAVID and GAD respectively. The analysis was carried out considering all the genes/proteins involved in the pathogenesis of MLD which were obtained from cluster and hub analysis using Cytoscape as well as proteins/genes which were found through literature.

The Overall work flow for gene interaction network and further analysis of network is shown below:



5. RESULTS

A ring like lattice network was obtained with 1050 nodes representing genes and 8019 edges representing the interactions between these genes. Edge Weighted Force Directed (BioLayout) network is shown in **Figure 10**. The Figure shows most of the nodes (genes/proteins) are present at the central core in the network while few nodes are lying at the periphery with lesser number of connections.

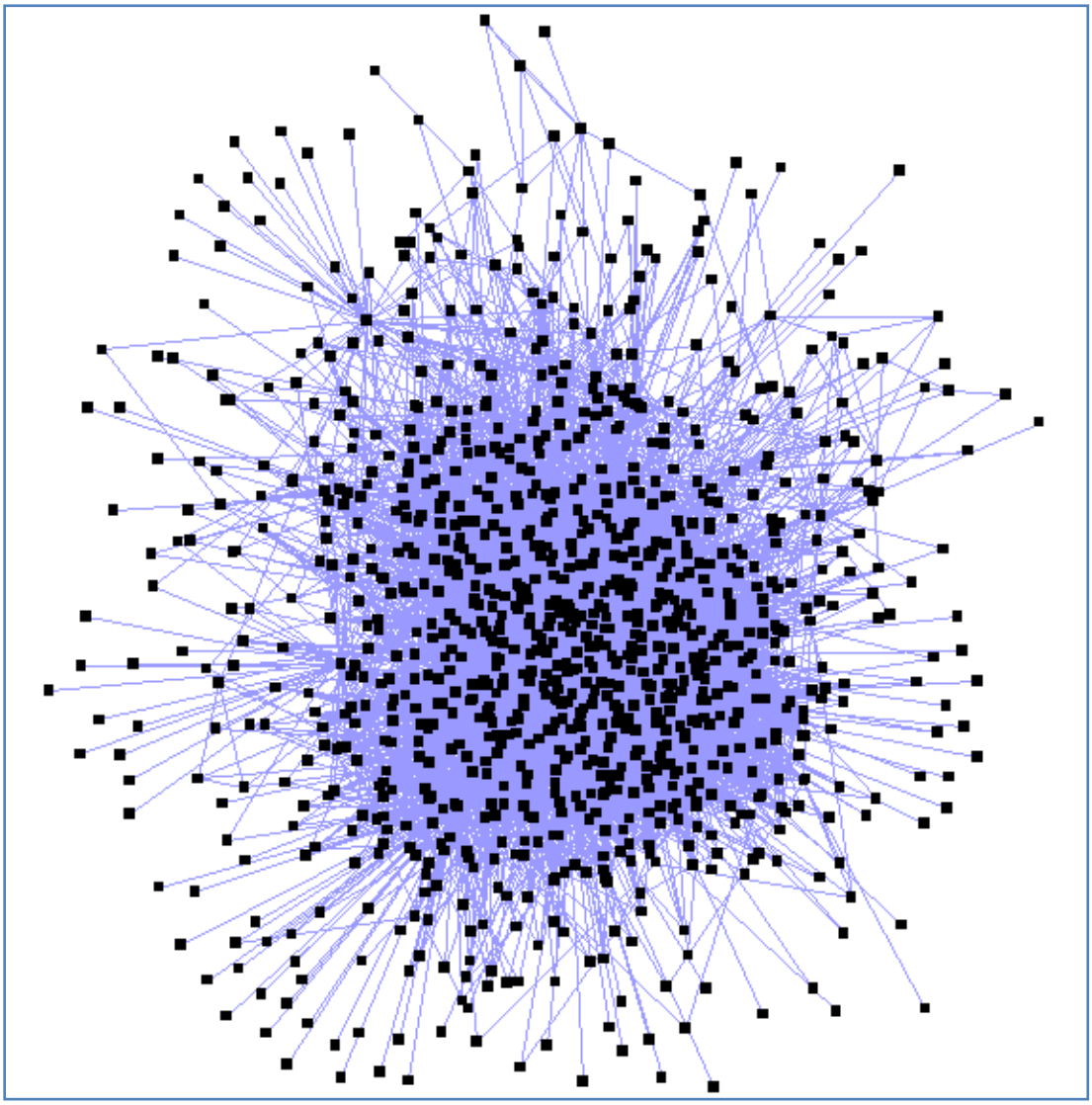


Figure 10: Protein interaction network having 1050 nodes and 8019 edges showing a ring-like lattice.

The candidate genes which are responsible for MLD were searched from published literature and their interactions were found from DIP, HPRD, BIND, STRING, MINT, IntAct, PUBMED, STRING, and Reactome databases. Journals with published literature and reports on MLD were also referred. There are around 70 genes responsible for this multiple system disease in human beings.

Literature survey and various databases search like HPRD, KEGG, etc was performed to get an insight about various disorders caused by mutations in genes interacting with ARSA and PSAP. The genes which play important role in causing disorders in different biological systems are shown in **Figure 11**. In MLD, Central Nervous System, Reproductive System, Respiratory System and Circulatory System with various other systems were found to be affected (**Figure 11**). MLD is known to be a neuro-degenerative disorder, the genes which are expressed in Nervous System were found to be highly affected. Genes associated with particular disorder are shown in corresponding color.

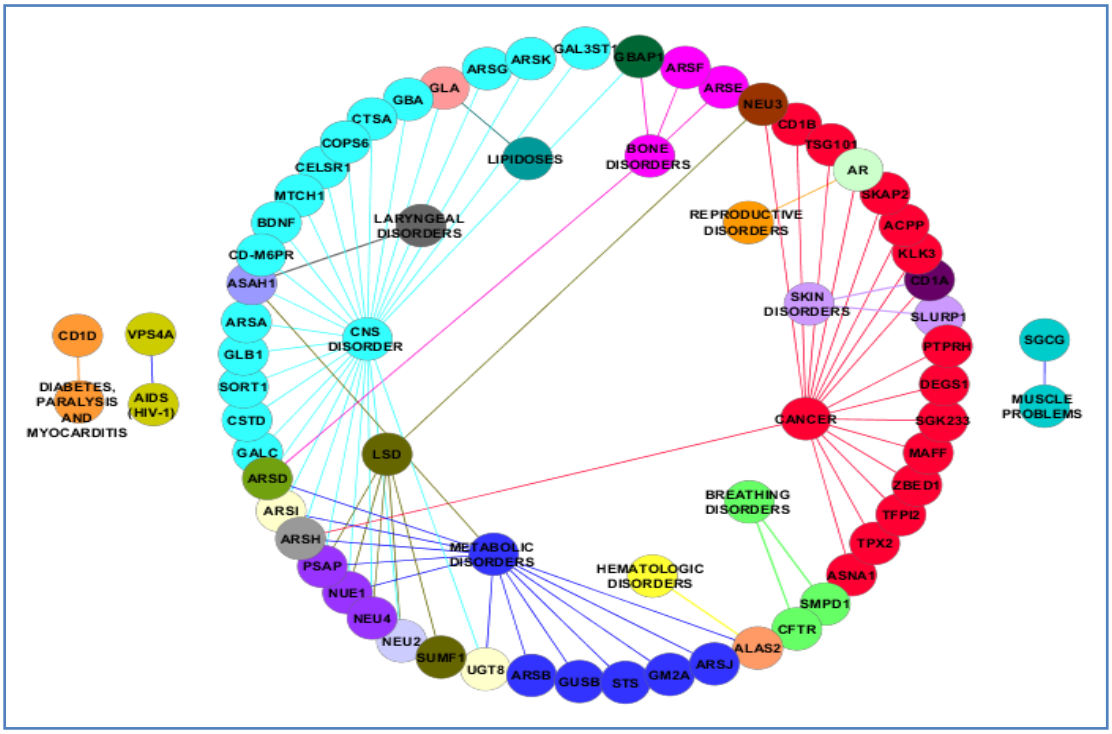


Figure 11: Disorders caused by genes interacting with ARSA and PSAP are shown. Out of 71 only 57 genes are found through literature and various databases search like HPRD, KEGG, etc.

The biological gene interaction network and its topological properties were studied and analyzed using Network analyzer. The analysis of network was done treating it as an undirected graph. For the network obtained, the overall calculated clustering coefficient is 0.288 indicating very limited number of connections in the neighborhood but nodes may be distantly connected. Centralization gives important information about how central the network's most central node is in relation to how central all the other nodes are. The value 0.253 of centralization indicates that the network topology have a few characteristic features of star topology. Network heterogeneity value 1.45 represents that network has tendency of having hubs (essential) nodes. The important hubs were identified in the network. In order to obtain these hubs the node degree value was taken into consideration. Nodes having degree more than 150 were considered and the genes which were found with larger node degree are TAF1(276), HNF4A(210), SMAD2(195), BRCA1(178), AR(163). Through literature study it was found that PSAP increases the activity of few of these genes [Koochekpour S *et al.*, 2007]. Any mutation in PSAP gene may cause the dysfunction of these genes. We were able to identify the significant genes which could play important roles in the underlying mechanism of MLD. The decreasing trend of neighborhood connectivity indicates that network is dominated by edges between low and highly connected nodes. When the power law curve function was fitted ($y=ax^b$) to closeness centrality and degree distribution, it is transformed into linear model ln y= ln b + a * ln x. A correlation value of 0.806 and R-square value 0.626 was obtained for closeness centrality. Similarly, for degree distribution the correlation value obtained was 0.779 and R-square was 0.852. The significant value of correlation and R-square shows that these two variables are negatively correlated and power law function is best fitted to this node degree distribution data which is the characteristic feature of Small world network model and can be seen in **Figure 12 and Figure 13**.These results signify the accuracy of the obtained network and imply that the network is highly significant and can be considered for further analysis.

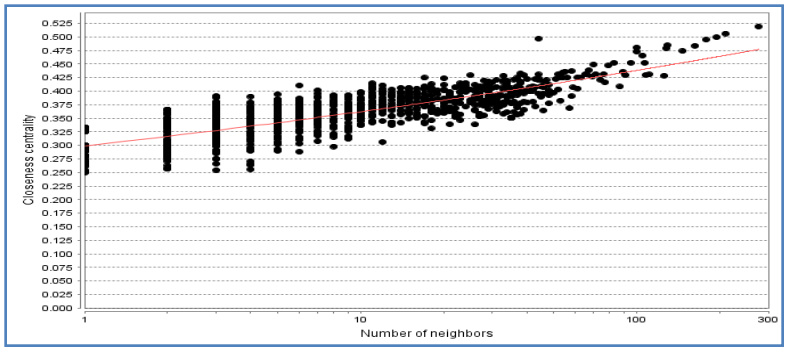


Figure 12: Linear model of power law function between closeness centrality and number of neighbors

The stress centrality parameter was visualized to observe the nodes which have greater stress implying larger number of edges passing through those nodes. The nodes having high stress were found: TAF1 (1772108.0), HNF4A (1202948.0), SMAD9 (107708), BRCA1 (989502.0), indicating that these nodes travel by high number of shortest path. It was found that PSAP node is connected through TAF1, SMAD9 and BRCA1 but ARSA is not connected to these stressed nodes.

We had calculated following parameters to determine the overall topology of network:

- Power law node degree distribution

$$p(k)=k^{\alpha}$$

- Characteristic path length

$$L=\log N$$

When the power law curve function was fitted ($y=ax^b$) to degree distribution, it is transformed into linear model ln y= ln b + a * ln x with a correlation value of 0.779 and R-square of 0.852 (**Figure 13**).

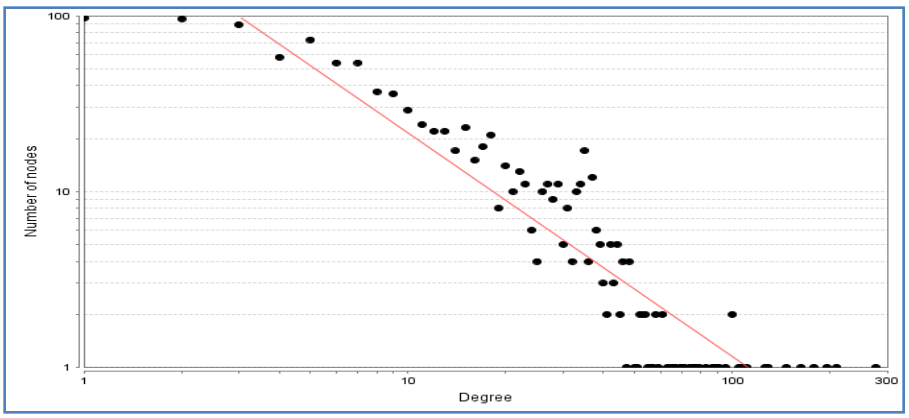


Figure 13: Linear model of power law function between degree and number of nodes**.**

The significant value of correlation and R-square shows that these two variables are negatively correlated and power law function is best fitted to this node degree distribution data which is the characteristic feature of Small world network model.

5.1 CHARACTERISTIC PATH LENGTH

We have calculated characteristic path length using the formula "Log N" which is a feature of small world network model [Danielle Smith Bassett and Ed Bullmore, 2006].

We have also compared this value with the value obtained from Network analyzer.

- Characteristic path length from network analyzer = 2.840
- Characteristic path length of Small World network model = log N
  = log (1050)
  = 3.02

The values obtained from network analyzer and the computed values are almost comparable.

Our network clearly resembles the Small world network model which was determined from power law degree distribution, characteristic path length, ring lattice like network. Additionally, it was also observed that the individual nodes were found to have few neighbors and most nodes can be reached from one another through few steps.

Cluster analysis for an interaction network is a crucial step in network analysis. Densely connected regions in graphs are most frequently identified by some unsupervised clustering method. However, standard clustering is not ideal for PPI networks: proteins may have multiple functions, and therefore the corresponding nodes may belong to more than one

cluster [Tamás Nepusz *et al.,* 2012]. Thus ClusterONE [Fan S *et al*., 2012] was used to identify the clusters with highly interconnected nodes. As a result, 145 clusters were obtained from which the cluster with density 34.436 and 126 nodes with 3.208E-9 p-value was selected for further analysis. Further study of the cluster showed that SMAD9, PSAP, BMPR2, ARSA, UBE3A genes were highly interconnected in the cluster and shown in **Figure 14**.

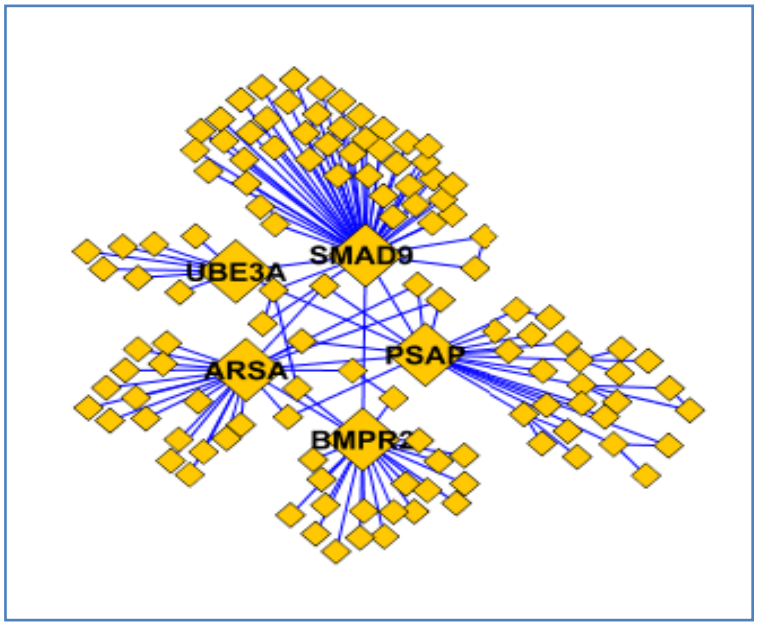


Figure 14: Cluster of 126 nodes representing highly interconnected genes

Disease-causing aberrations in the normal function of a gene define that gene as a disease gene. Proving a causal link between a gene and a disease experimentally is expensive and time-consuming. Comprehensive prioritization of candidate genes prior to experimental testing drastically reduces the associated costs. Computational gene prioritization is based on various pieces of correlative evidence that associate each gene with the given disease and suggest possible causal links. Candidate gene identification and prioritization of genes was done using Cytohubba plug-in in Cytoscape to find out essential nodes in the network which are also known as hub nodes or essential nodes. There are a number of methodologies available for hubs and essential genes identification, and all of them are available with the respective advantage and limitations [Chung-Yen L *et al.,* 2008, Kong W *et al.,* 2011, Mishra H *et al.,* 2011, Rio G *et al.,* 2009]. Some strategies are the use of genetic algorithm or machine learning procedures, however, the centrality approaches are by far the most applied

procedures even by simplicity and because several studies had being pointed out its applicability. Therefore, several centrality indexes were evaluated to obtain the potential genes: Betweenness, bottleneck, density of maximum neighborhood component (DMNC), eccentricity, maximal clique centrality (MCC),maximum neighborhood component (MNC), node degree, edge percolation component (EPC), radiality and stress [Chung-Yen L *et al.*, 2008]. On the basis of this normalized centrality index score, the top 50 genes with highest score I are considered as potential target genes **(Table 6).**

Normalized centrality index score is the combined percentage value of all ten centrality indexes. In this analysis we found MLD related genes like PSAP and ARSA in addition with other genes with high score I value like TAF1, SMAD2, BRCA1, HNF4A, AR, SMAD9 are also found which may be proved as potential drug targets for MLD (**Table 6**). In this analysis score I obtained for ARSA gene is found to be low as compared to other genes because it does not comprise any bigger hub center in protein network but it plays a critical role in passing information between hub centers.

Two hypothetical genes LOC729948 and LOC730314 were found to be among important genes. It is found that these hypothetical genes are similar to CDC26 subunit of anaphase promoting complex, so it is possible that they may further interact with CDC27, PTTG1, CCNB1, CDC16, ANAPC2, ANAPC16, ANAPC13, ANAPC5, CDC23, ANAPC7. Emphasize should be given to these hypothetical genes because these are found to be among top 50 candidate genes (**Table 6**). These genes should be further explored through Micro array data analysis to check their expression profile with respect to MLD.

Table 6: Top 50 genes obtained according to normalized centrality index.

Gene Symbol	Chromosomal location	Score I	Gene Symbol	Chromosomal location	Score I
TAF1	X	747.5212087	CDC26	9	367.7928177
SMAD2	18	646.9450308	FZR1	19	367.0962171
BRCA1	17	562.4871301	SMAD3	15	366.886843
HNF4A	20	553.005314	LOC729948	7	366.8359585
AR	X	484.9451925	ANAPC4	4	366.2309342
SMAD9	13	419.2151898	LOC730314	7	365.6949616
CDC2	10	416.0214789	CDKN1B	12	365.2963005
RB1	13	415.9210528	ANAPC2	9	364.6175973
UBC	12	415.0906131	ANAPC1	2	364.3057073
CDK2	12	413.5876164	ANAPC7	12	364.2794019
UBB	17	413.5730965	CCNB1	5	361.1748873
PSAP	10	397.6550792	ANAPC11	17	361.0961297
ANAPC10	4	392.6502057	ANAPC5	12	360.2940108
CDC23	5	391.3418829	CDC6	9	356.515173
MYC	8	387.2418442	CDC20	1	356.1679529
MNAT1	14	383.831744	UBE2E1	3	355.9524951
CCNH	5	382.3370698	BUB3	10	355.2756671
CDK7	5	381.3404889	ORC1L	1	355.1152718
CDC27	17	381.2638475	BUB1B	15	354.3007697
TP53	17	379.242666	MAD2L1	4	353.4248162
CDC16	13	377.5941342	PTTG1	5	353.3694199
CDKN1A	6	374.3586452	CDC25B	20	351.9598547
CDC25A	3	371.6090863	ORC5L	7	350.3714013
UBE2D1	10	368.7256054	ORC3L	6	349.6877287
UBE2C	20	368.1183305	PLK1	16	349.5862661

5.2 GENE ONTOLOGY ANALYSIS

GO describes gene and gene product attributes in term of their associated biological processes, cellular components and molecular functions. The size (area) of the nodes is proportional to the number of genes in the test set which are annotated to that node. The color of node represents the (corrected) p-values which are more than 5 order of magnitude smaller

than the chosen significance level. We performed molecular network analysis for the entire network using BiNGO application. From this analysis we found gene distribution across different molecular functional categories like protein binding, catalytic activity, etc (**Figure 15**). Binding node was found to be statistically significant with p-value 7.5649E-42 and 95.0% frequency (**Table 7)**. The binding node encompasses multiple types of molecular interactions, including protein binding, nucleic acid binding and nucleoside binding. The binding node is the most significant node because it encompasses 997 genes of 1050 genes. Out of 1050 genes in the MLD molecular function network, 892 genes are involved in protein binding. The binding node is further linked to nucleic acid binding node (304 genes), and nucleoside binding node (189 genes). Genes responsible for MLD i.e ARSA and PSAP were found to be involved in statistically significant node Binding. The secong ost important node was found to be the catalytic activity node. It was also found that ARSA and PSAP both are also involved in catalytic activity node encompassing 420 genes. However, when molecular function analysis was performed on the cluster with density 34.436 and 126 nodes, catalytic activity was found to be most statistically significant encompassing 63 genes out of 126 genes. Catalytic activity is further linked to statistically significant hydrolase activity node. Hydrolase activity node is further shown to be linked with arylsulfatase activity node. Though the size of the arylsulfatase node is comparatively small but it is statistically significant and is responsible for causing MLD. Further, the node for binding has size larger than catalytic activity but it is not found to be statistically significant but the binding node is further linked to HECT domain binding node which is found to be statistically significant **(Figure 16).**

Table 7: Molecular function analysis of MLD network

GO Description	Frequency	P-value
Binding	95.0%	7.5649E-42
Protein Binding	84.4%	0.0000E-100
Nucleic acid Binding	29.3%	2.0448E-9
Nucleoside Binding	17.9%	4.5978E-12
Catalytic activity	40.6%	1.7005E-6
Transcription regulatory activity	23.2%	1.09333E-36
Enzyme Regulatory activity	8.4%	6.2882E-4

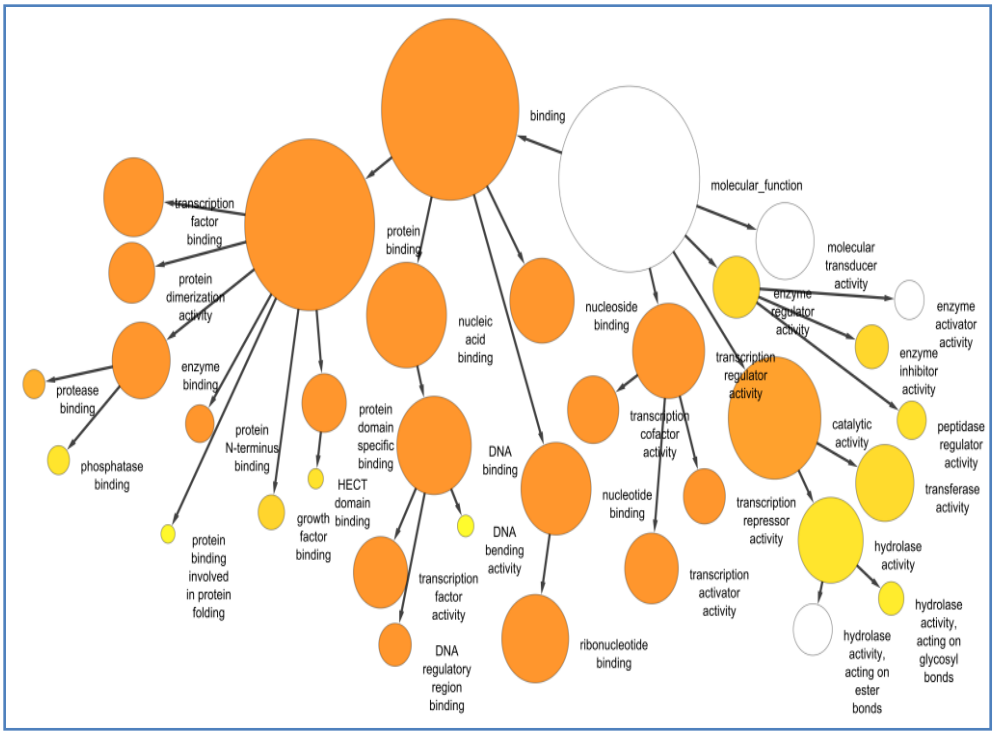


Figure 15: Showing major molecular function categories for MLD network**.**

Most important genes which are responsible for MLD like ARSA and PSAP belong to binding molecular function category. More specifically ARSA is responsible for Arylsulfatase activity, catalytic activity and hydrolase activity and PSAP is responsible for protein binding and enzyme regulatory activity.

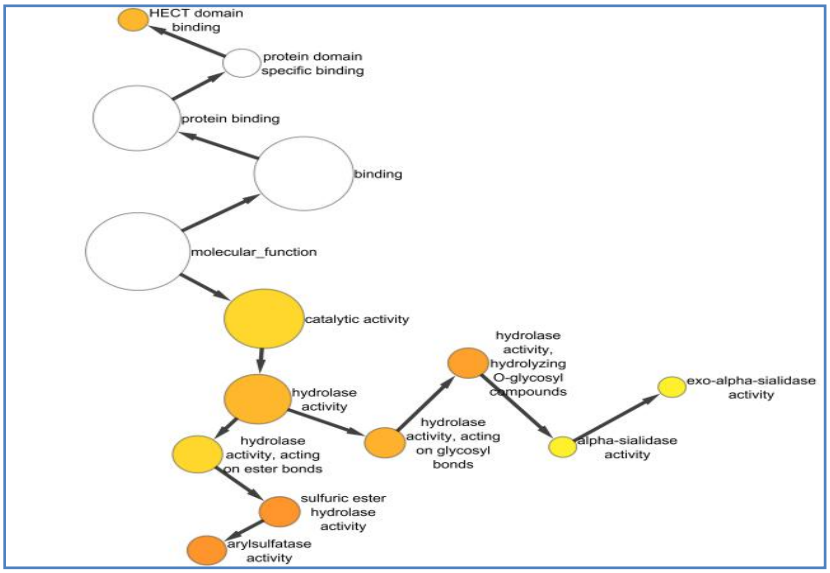


Figure 16: Showing major molecular function categories for ClusterONE result.

It is important to consider that several genes in the network do not present a known relation with specific disease. So in order to identify the disease classes to which these potential genes belong, disease enrichment analysis was performed. Several types of disease classes were found statistically significant in the enrichment analysis. In the GAD database, out of 126 genes, only 88 genes were found and were classified against their major disease classes. Various disease classes were obtained from DAVID, four were considered important for further analysis on the basis of number of genes in each disease class. As expected neurological and metabolic diseases classes were found to be significant in MLD. Surprisingly our results revealed the presence of significant number of genes which may be involved in Cancer pathways. Few of them are TAF1, BMPR2, BRCA1, PSAP, SMAD2. Immune disease class was also found to be significant. Genes such as HNF4A, SMAD2, BMPR2, BRCA1, CTSD were found under this class. A summarized result with the corresponding p-values for each disease subclass is shown in **Table 8.** The genes which are involved in major disease classes are shown in **Figure 17**.

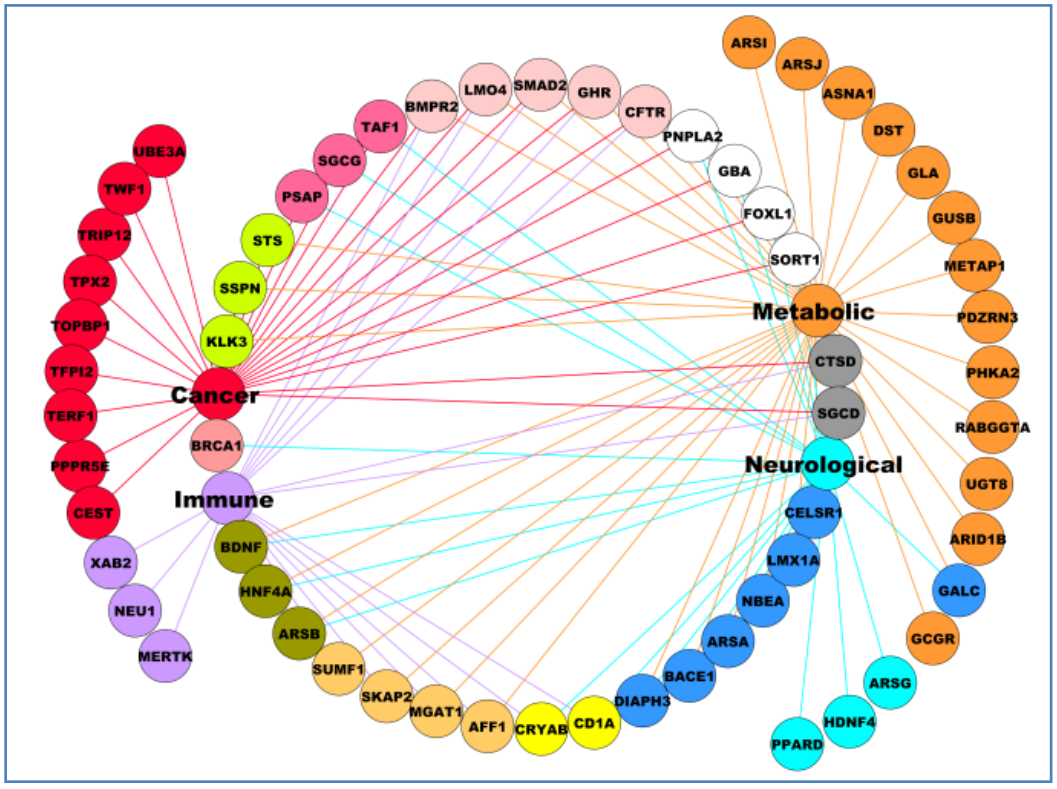


Figure 17: Showing Disease Enrichment Analysis for genes interacting with ARSA and PSAP genes.

Table 8: The disease enrichment analysis

Major Disease classes obtained from DAVID	P-Value	Disease
Metabolic	5.9e-4	Cholestrol, Hdl Body Weight Changes Diabetes Type 2 Lipoprotein,Ldl Triglycerides Bone Mineral Density
Neurological	2.9e-3	Alzheimer's Disease Stroke Parkinson's Disease
Cancer	8.2e-3	Breast Cancer Lung Cancer Prostate Cancer Ovarian Cancer
Immune	1.8e-2	Multiple Sclerosis Human Immuno Monocytes Diabetes Type 1

For better understanding and fine representation of disease enrichment analysis results, Venn and Euler diagram plug-in [Wilkinson L., 2012] in Cytoscape was used. The graphical interface shows the gene distribution among four major disease classes. From this analysis we found that out of 88 genes, 27 are involved in Cancer, 41 are involved in Metabolic, 23 are involved in Neurological and 20 are involved in Immune class (**Figure 18**). The gene distribution shared by various classes is also shown in tabular format (**Table 9**).

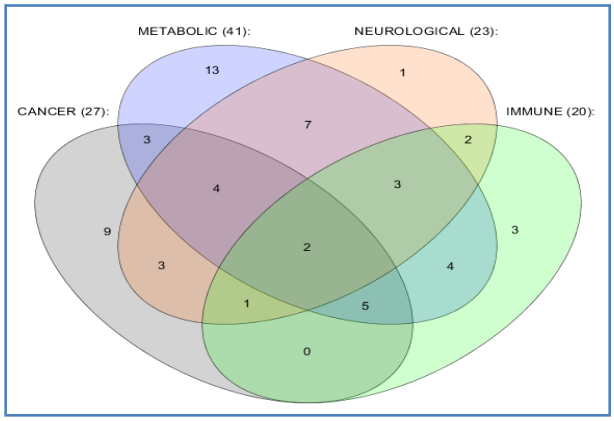


Figure 18: Venn Diagram representing number of common genes involved in various disease classes.

Table 9: Genes shared by various disease classes.

Disease Class	Number of Gene	Gene Name
Cancer + Metabolic	3	KLK3, STS, SSPN
Cancer + Neurological	3	PSAP, TAF1, SGCG
Metabolic + Neurological	7	DIAPH3, LMX1A, GALC, BACE1, CELSR1, ARSA, NBEA
Metabolic + Immune	4	SUMF1, SKAP2, MGAT1, AFF1
Neurological + Immune	2	CRYAB, CD1A
Cancer + Metabolic + Neurological	4	FOXL1, SORT1, GBA, PNPLA2
Cancer + Metabolic + Immune	5	GHR, CFTR, BMPR2, LMO4, SMAD2
Cancer + Neurological + Immune	1	BRCA1
Metabolic + Neurological + Immune	3	BDNF, HNF4A, ARSB
Cancer + Metabolic + Neurological + Immune	2	CTSD, SGCD

From **Table 9** it is evident that one of the major candidate gene, PSAP which is a well known gene in MLD is also involved in cancer disease class.

To calculate the distribution of potential genes across different chromosomes, we performed chromosomal enrichment analysis of top 50 potential genes obtained from normalized centrality score analysis (Score *I*) (**Table 6).** We observed that potential genes are distributed on different chromosomes with chromosome 17 having maximum number of candidate genes implying its importance in the pathogenesis of MLD. There are six genes: ANAPC11, BRCA1, UBB, CDC27, TP53, CDC6 found on chromosome 17 and all are related to cell cycle and ubiquitin mediated proteolysis pathways. 5 genes were found on chromosomes 12 and 5. CDKN1B, ANAPC7, ANAPC5, UBC, CDK2 genes are present on chromosome 12. PTTG1, CCNB1, CDC23, CCNH, CDK7 genes are found on chromosome 5. Interestingly, these genes on both the chromosomes are known to play role in cell cycle and cancer pathways. PSAP, BUB3, CDC2, UBE2D1 genes are found to be present on chromosome 10 and are involved in cell cycle, metabolic pathways and proteolysis (**Table 6)**.

To find out the pathways which are getting affected with the genes involved in MLD network, pathway enrichment analysis was performed through DAVID bioinformatics resource. The analysis was carried out on 126 genes present in the cluster obtained from ClusterONE. Out of 126 potential genes, 120 genes were found in DAVID, which shows DAVID database is the most representative of our gene space and query coverage was around 96%. It was found that Sphingolipid metabolic, lysosomal, proteolytic and ARVC pathways get mostly affected with the genes involved in MLD network (**Table 10**). Genes related to their corresponding pathways are shown in **Figure 2.**

Table 10: Pathway enrichment analysis of potential genes found through DAVID.

Gene Ontology term	Genes involved	Gene Names	%	p-value
Sphingolipid metabolism	14	ASAH1,UGT8,ARSA,DEGS1,GAL3ST1, GLA,GLB1,GALC,GBA,NEU1,NEU2, NEU3,NEU4,SMPD1	9.5	1.2E-15
Lysosome	19	GM2A,ASAH1,ARSA,ARSB,ARSG,CTS A,CTSD,CTSL1,GLA,GLB1,GALC,GBA, GUSB,M6PR,MAN2B1,PSAP,NEU1, SORT1,SMPD1	12.9	7.5E-15
Other glycan degradation	7	GLB1,GBA,MAN2B1,NEU1,NEU2,NEU3 ,NEU4	4.8	3.9E-8
Ubiquitin mediated Proteolysis	8	BRCA1,CUL7,TRIP12,UBE3A,UBE2Q1, UBE2L6,UBE2M,UBE2MP1,UBA6	5.4	2.6E-3
Glycosaminoglycan Degradation	3	ARSB,GLB1,GUSB	2.0	3.3E-2
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	4	PKP2,SGCB,SGCD,SGCG0	2.7	8.6E-2

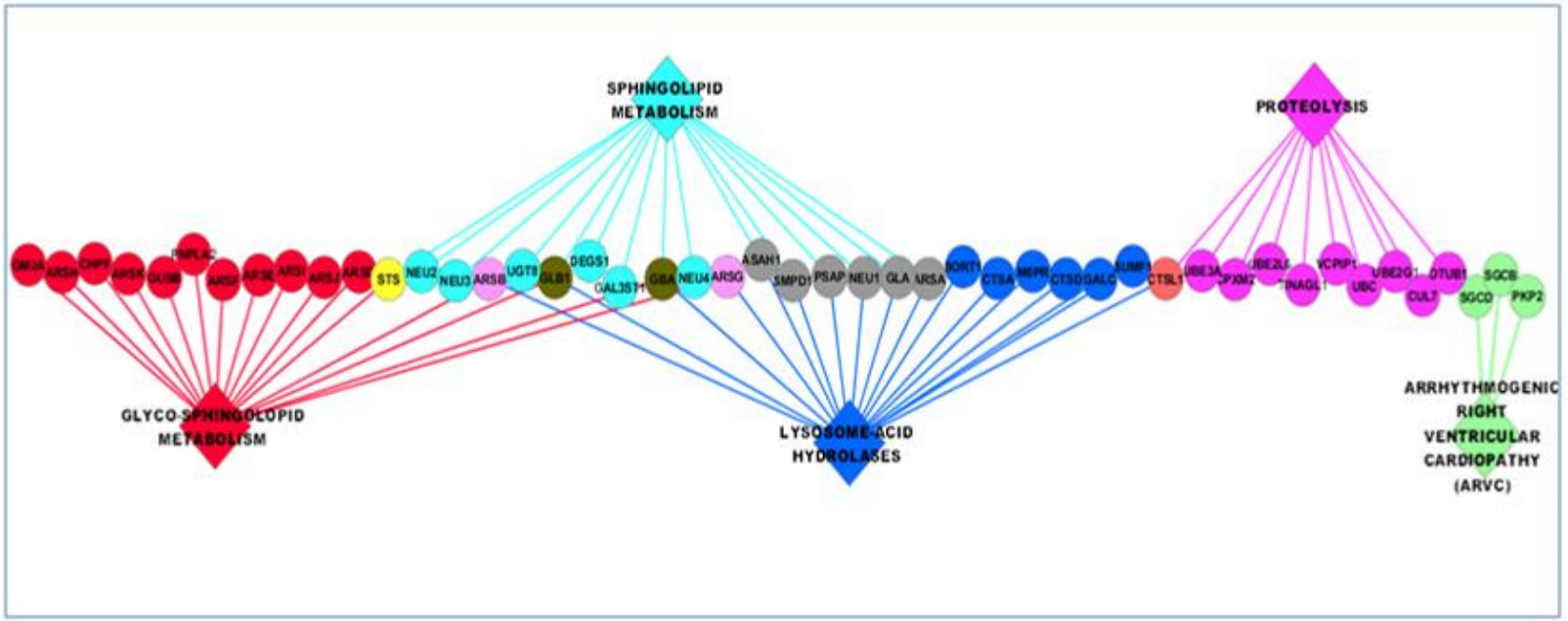


Fig 2 : Figure shows genes involved in major pathways obtained from Pathway Enrichment Analysis using Cytoscape

Tissue gene expression analysis was performed for better understanding of important genes which may express in various organs in MLD. Genes are distributed and are getting expressed mainly in Liver, Placenta, Fetal Brain, Small Intestine, Urine and Nervous System (**Table 11**). Surprisingly, in addition to Nervous system and excretory system, major number of genes were getting expressed in Placenta and Liver.

Table 11: Tissue Gene Expression Analysis Using DAVID.

Term	%	Gene	Gene Name	p-value
LIVER	22.4	33	CLEC4G,ASAH1,SIL1,TPX2,ARSA,ARSB,ARSD,ARSE,ALAS2,CES1,CES3,CTSA,CTSD,CTSL1,GAL,GLB1,GCGR,GBA,GUSB,QARS,GHR,HNF4A,M6PR,MAN2B1,PSAP,PHKA2,SGCB,NEU1,TERF1,SLC9A6,SKAP2,SUMF1,TINAGL1	4.2E-5
PLACENTA	30.6	45	AFF1,COPS6,GM2A,SIL1,SMAD2,ACPP,ALAS2,ARS,BRCA1,CES1,CTSA,CEP135,DEGS1,DST,EXPH5,GLB1,GALC,GCGR,GBA,GUSB,QARS,GHR,HERVV1,MAN2B1,METAP1,MTMR10,PPARD,PKP2,POMGNT1,PPP2R5E,SGCB,SGCD,SLURP1,NEU1,PLEC,UBASH3A,SUMF1,TFPI2,TINAGL1,TSG101,UBE2M,MAFF,MOS,VPS8,ZSCAN4	5.3E-5
FIBROBLAST	4.8	7	ASAH1,GLA,GUSB,NEU1,NEU4,SMPD1,TFPI2	4.2E-4
SMALL INTESTINE	5.4	8	ACTR8,ARHGAP9,ARSA,CHPF,KDM6A,MTMR11,NEU4,MAFF	1.2E-2
FETAL BRAIN	8.8	13	GPSM1,RABGGTA,SMAD9,TAF1,ARSF,CETN2,C10ORF2,DYNLT1,DST,SGK223,PPP2R5E,UBE3A,UBE2G1	1.4E-2
NERVOUS SYSTEM	2.0	3	BRCA1,DST,PSAP	6.7E-2
EXCRETORY SYSTEM	2.0	3	CRYAB,GAL3ST1,UBE3A	7.8E-2
LEUKOCYTE	2.7	4	BDNF,KLK3,M6PR,SKAP2	9.6E-2

6. DISCUSSION

The aim of the study was to built and analyzes the protein-protein interaction network for MLD. With the help of MiMI application in Cytoscape 2.8.3, protein-protein interaction network was built for MLD. The obtained network has Small World Network Model characteristics. Five genes were found to be highly inter-connected, namely SMAD9, PSAP, BMPR2, ARSA, UBE3A from ClusterONE analysis, which may play a major role in pathogenesis for this disease. The main objective behind this study was to identify probable drug targets for MLD. These were found using Cytohubba. The proteins encoded by the identified genes: TAF1, SMAD2, BRCA1, HNF4A, AR, SMAD9, CDC2, RB1, UBC, CDK2, UBB, PSAP, CDC23, MYC, MNAT1, CCNH, CDK7 can be potential drug targets for this disorder. Normalized centrality index score also supports potential gene targets obtained from Cytohubba algorithms.

Molecular function of these genes was studied using BiNGO. It was found that the genes causing MLD (ARSA and PSAP) and other important candidate genes are mostly involved in Binding and Catalytic activity. Major pathways which may be affected in MLD are Sphingolipid Metabolism, Lysosome, Proteolysis, Arrhythmogenic right Ventricular cardiomyopathy (ARVC). Organs like Liver, Small Intestine, Brain and Nervous System may get affected in MLD. Disease enrichment analysis shows that MLD has a possible link with metabolism, immune disorders and cancer.

7. CONCLUSION AND FUTURE PERSPECTIVE

The molecular interaction study of MLD was done using Cytoscape tool and its plug-ins. This study focuses on building and analyzing the protein-protein interaction network for MLD. MiMI was used to retrieve interactions of genes which were involved in causing MLD. Network was analyzed by Network Analyzer and we found that the query protein-protein interaction network have a Small World Network Model characteristics. From the list of clusters obtained through ClusterONE, the cluster with 34.436 density was selected. This highest ranking sub-network was found to have 126 genes, from which five genes were found to be highly inter-connected, namely SMAD9, PSAP, BMPR2, ARSA, UBE3A which may play a major role in pathogenesis for this disease. The main idea behind this study was to identify probable drug targets for MLD. These were found using Cytohubba. The proteins encoded by the identified genes: TAF1, SMAD2, BRCA1, HNF4A, AR, SMAD9, CDC2, RB1, UBC, CDK2, UBB, PSAP, CDC23, MYC, MNAT1, CCNH, CDK7 can be potential drug targets for this disorder. These genes have been poorly explored or unknown in the current state of the art of MLD physiopathology and should be further validated using experiments such as microarray, or genome wide association studies. The lack of gene expression data or microarray studies make it all the more important to study the MLD system in different tissues and cell lines. Structure of SMAD9 and MNAT1 proteins are still unknown, they need to be further explored in near future. Normalized centrality index score also supports potential gene targets obtained from Cytohubba algorithms. In this analysis score obtained for ARSA gene is found to be low as compared to other genes because it does not comprise any bigger hub center in protein network but it plays a critical role in passing information between hub centers which is also confirmed by Bottleneck algorithm. Molecular function of these genes was studied using BiNGO. It was found that the genes causing MLD (ARSA and PSAP) and other important candidate genes are mostly involved in Binding and Catalytic activity.

Pathway enrichment analysis was done to get deeper insight into the molecular mechanism of MLD. From this analysis we conclude that major pathways which may be affected in MLD are Sphingolipid Metabolism, Lysosome, Proteolysis, Arrhythmogenic right Ventricular cardiomyopathy (ARVC). Through tissue gene expression analysis we found that Organs like

Liver, Small Intestine, Brain and Nervous System may get affected in MLD. The comprehensive study was also done to find out other disease linked to MLD. Disease enrichment analysis was carried out to find to which disease class these potential genes belong. Through disease enrichment analysis we corroborated that MLD is a well known neurological disorder but we also found a possible link of MLD with metabolism, immune and cancer.

8. REFERENCES

Aicardı` J (1998).Disease of the Nervous System in Childhood.Cambridge University Press: Cambridge, 2nd edn, **907**, 122-27.

Anderson, T.J.; A. Schneider; J.A. Barrie; M.Klugmann, M.C.; McCulloch, D.; Kirkham, E.; Kyriakides, K.A.; Nave, and I.R. Griffiths (1998).Late-onset neurodegeneration in mice with increased dosage of the proteolipid protein gene. J Comp Neurol .**394**, 506-19.

Andrew Chatr-aryamontri; Arnaud Ceol; Luisa Montecchi Palazzi; Giuliano Nardelli; Maria Victoria Schneider; Luisa Castagnoli; and Gianni Cesareni (2007).MINT: the Molecular INTeraction database.Nucleic Acids Res.**35**. D572–D574.

Arvan L Fluharty (1993).Arylsulfatase A Deficiency. GeneReviews. Seattle (WA): University of Washington, Seattle.**47**, 76-82.

Arvanitis, D.; M. Dumas ;and S. Szuchet (1992). Myelin palingenesis. 2. Immunocytochemical localization of myelin/oligodendrocyte glycolipids in multilamellar structures.Dev Neurosci. **14**:328-35.

Barboura I; Ferchichi S; Dandana A; Jaidane Z; Ben Khelifa S; Chahed H; Ben Mansour R; Chebel S; Maire I; Miled A (2010). Metachromatic leucodystrophy. Clinical, biological, and therapeutic aspects . Ann Biol Clin (Paris).**68**(4), 385-91.

Benjamini,Y. and Yekutieli,D (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Statist. **29**, 1165–1188.

Bolin, L.M.; T.P. Iismaa; and E.M. Shooter (1992). Isolation of activated adult Schwann cells and a spontaneously immortal Schwann cell clone. J Neurosci Res. **33**, 231-8.

Bosio, A.; E. Binczek; and W. Stoffel (1996). Functional breakdown of the lipid bilayer of the myelin membrane in central and peripheral nervous system by disrupted galactocerebroside synthesis. Proc Natl Acad Sci U S A. **93**, 13280-5.

Boutry, J.M.; J.J. Hauw, A.; Gansmuller, N.; Di-Bert, M.; Pouchelet; and A. Baron-Van Evercooren (1992). Establishment and characterization of a mouse Schwann cell line which produces myelin in vivo. J Neurosci Res. **32**, 15-26.

Chen, G.L.; N.L. Halligan; N.F. Lue; and W.W. Chen (1987). Biosynthesis of myelinassociated proteins in simian virus 40 (SV40)-transformed rat Schwann cell lines. Brain Res. **414**, 35-48.

Chin,C.-S; and Manoj,P.S (2003). Global snapshot of a protein interaction network—a percolation based approach. Bioinformatics, **19**, 2413–2419.

Chung-Yen Lin; Chia-Hao Chin; Hsin-Hung Wu; Shu-Hwa Chen; Chin-Wen Ho2 and Ming-Tat Ko (2008). Hubba: hub objects analyzer—a framework of interactome hubs identification for network biology, Nucleic Acids Research, **36**, 717-26.

Clouter-Mackie; M.B. and Gagnier L. (2003). Spectrum of mutations in the arylsulfatase A gene in a Canadian collection including two novel frameshift mutations, a ion (C488R) and an MLD mutation (R84) in cis with a pseudodeficiency allele. Mol Genet Metab. **79**, 91-108.

Coetzee, T.; N. Fujita, J.;Dupree, R. Shi; A. Blight; K. Suzuki and B. Popko (1996). Myelination in the absence of galactocerebroside and sulfatide: normal structure with abnormal function and regional instability. Cell. **86**, 209-19.

Coulter-Mackie, M.B.; D.A. Applegarth; J.R.Toone; L. Gagnier; A.R. Anzarut and G. Hendson (2002). Isolated peripheral neuropathy inatypical metachromatic leukodystrophy: a recurrent mutation. Can J Neurol Sci. **29**,159-63.

Dennis G Jr; Sherman BT; Hosack DA; Yang J; Gao W; Lane HC; Lempicki RA (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery, Genome Biol. **45,** 143-49.

Farrell, D.F (1974). Enzymatic sulphation of some galactose- containing sphingolipids in developing rat brain. J Neurochem. **23**, 219-25.

Fressinaud, C.; J.M. Vallat; M. Masson; M.O. Jauberteau; N. Baumann and J. Hugon(1992) Adult-onset metachromatic leukodystrophy presenting as isolated peripheral neuropathy. Neurology. **42**, 1396-8.

Gary D. Bader; Doron Betel and Christopher W. V. Hogue (2003).BIND: the Biomolecular Interaction Network Database, Nucleic Acids Res. **31**(1), 248–250.

Ge,Y.; Dudoit,S. and Speed,T.P (2003). Resampling-based multiple testing for microarray data analysis. Technical Report, Dept. of Statistics, UC Berkeley. **633**, 433-46.

Gieselmann, V. (2003). Metachromatic leukodystrophy: recent research developments. J Child Neurol. **18**, 591-4.

Gieselmann V; Polten A; Kreysing J; Kappler J; Fluharty AL; Bohne W; von Figura K (1991). Mutations in arylsulfatase A alleles causing metachromatic leukodystrophy. Brain Dysfunction **4**,235-243.

Goda, S.; J. Hammer; D. Kobiler and R.H.Quarles(1991). Expression of the myelinassociated glycoprotein in cultures of immortalized Schwann cells. J Neurochem. **56,** 1354-61.

Guenard, V.; L.A. Gwynn and P.M. Wood (1994). Astrocytes inhibit Schwann cell proliferation and myelination of dorsal root ganglion neurons in vitro. J Neurosci. **14**, 2980-92.

Hai, M.; N. Muja; G.H. DeVries; R.H. Quarles and P.I. Patel (2002). Comparative analysis of Schwann cell lines as model systems for myelin gene transcription studies. J Neurosci Res. **69**, 497-508.

Hansen, L.M.; O. Kristensen and M.L. Friis (1994).Neuropathy in adult metachromatic leukodystrophy. Ugeskr Laeger. **156**, 2252-3.

Holtschmidt H; Sandhoff K; Kwon HJ; Harzer K; Nakano T; Suzuki K (1991). Sulfatide activator protein. Alternative splicing that generates three mRNAs and a newly found mutation responsible for a clinical disease. J Biol Chem **266**, 7556–7560.

Honke, K.; Y. Hirahara; J. Dupree; K. Suzuki; B. Popko; K. Fukushima; J. Fukushima; T. Nagasawa; N. Yoshida; Y. Wada and N. Taniguchi (2002). Paranodal junction formation and spermatogenesis require sulfoglycolipids. Proc Natl Acad Sci U S A. **99**, 4227-32.

Hyde TM; Ziegler JC; Weinberger DR. Psychiatric disturbances in metachromatic leukodystrophy. Insights into the neurobiology of psychosis. Arch Neuro l1992.**49**, 401–406.

Ilhem Barboura; Irène Maire; Salima Ferchichi and Abdelhedi Miled(2010). Laboratory of Biochemistry, CHU Farhat Hached, Sousse, Laboratory of Biochemistry Pediatric Debrousse Hospital, Lyon, Tunisia France, Metachromatic Leukodystrophy Clinical, Biological and Therapeutic Aspects, Ann Biol Clin (Paris). **68**(4), 385-91.

Jeong,H., Mason,S.P., Baraba´si,A.L. and Oltvai,Z.N (2001). Lethality and centrality in protein networks. Nature. **411**, 41–42.

Jing Gao; Alex S. Ade; V. Glenn Tarcea; Terry E. Weymouth; Barbara R. Mirel; H.V. Jagadish and David J. States (2009). Integrating and annotating the interactome using the MiMI plugin for cytoscape. Bioinformatics. **25**(1), 137–138

Jirsova, K.; P. Sodaar; V. Mandys and P.R. Bar (1997). Cold jet: a method to obtain pure Schwann cell cultures without the need for cytotoxic, apoptosis-inducing drug treatment. J Neurosci Methods. **78**, 133-7.

Joachim Kreysing; Wolfgang Bohne; Claudia B6senberg; Sergio Marchesini; Jean Claude Turpin; Nicole Baumann; Kurt von Figura and Volkmar Gieselmann (1993), High Residual Arylsulfatase A (ARSA) Activity in a Patient with Late-infantile Metachromatic Leukodystrophy. Am. J. Hum. Genet. **53**, 339-346.

Joel Zlotogora; Yael Furman-Shaharabani; Ann Harris; Maria Luiza Barth; Kurt von Figura; Volkmar Gieselmann(1994). A single origin for the most frequent mutation causing late infantile metachromatic leucodystrophy. JMed Genet **31**,672-674.

J. W. Black and J. N. Cumings (1961). Infantile Metachromatic Leucodystrophy J. Neurol. Neurosurg. Psychiat. **24**, 233-37.

Kappler J; Leinekugel P; Conzelmann E; Kleijer WJ; Kohlschutter A; Tonnessen T; Rochel Mac Kevin; *et al*. (1991). Genotype phenotype relationship in various degrees of arylsulfatase A deficiency. Hum Genet **86**, 463-470

Karthik Raman (2010). Construction and analysis of protein–protein interaction networks, Automated Experimentation. **2**, 21-27.

Kelly J Perkins; William F Careyand and C Phillip Morris (2005). An arylsulphatase A (ARSA) frameshift mutation (289insG) in metachromatic leukodystrophy (MLD). Journal of Molecular and Genetic Medicine. **1**(1), 3-4.

Kevin G Becker; Kathleen C Barnes; Tiffani J Bright and S Alex Wang (2004). The Genetic Association Database, Nature Genetics **36**, 431 – 432.

Kishimoto Y; Hiraiwa M; O'Brien JS (1992). Saposins: structure, function, distribution, and molecular genetics. J. Lipid Res. **33** (9), 1255–67.

Kong W; Mou X; Hu X(2011). Exploring matrix factorization techniques for significant genes identification of Alzheimer's disease microarray gene expression data. BMC Bioinformatics **12**(5):S7.

Krivit W (2004). Allogeneic stem cell transplantation for the treatment of lysosomal and peroxisomal metabolic diseases. *Springer Semin Immun*. **26**,119-132

Ługowska A; Ponin´ska J; Krajewski P; Broda G; Płoski R (2011). Population Carrier Rates of Pathogenic ARSA Gene Mutations: Is Metachromatic Leukodystrophy Underdiagnosed? PLoS ONE **6**(6), 163-69.

Martin PL; Carter SL; Kernan NA (2006). Results of the cord blood transplantation study (COBLT): outcomes of unrelated donor umbilical cord blood transplantation in pediatric patients with lysosomal and peroxisomal storage diseases. *Biol Blood Marrow Transplant*. **12**(2), 184-94.

Martinez, A.C.; M.T. Ferrer; E. Fueyo and L. Galdos (1975). Peripheral neuropathy detected on electrophysiological study as first manifestation of metachromatic leucodystrophy in infancy. J Neurol Neurosurg Psychiatry. **38**, 169-74.

Mehl, E. and H. Jatzkewitz (1965). Evidence for the Genetic Block in Metachromatic Leucodystrophy (Ml). Biochem Biophys Res Commun. **19**, 407-11.

Mishra H; Singh N; Misra K; Lahiri T (2011). An ANN-GA model based promoter prediction in Arabidopsis thaliana using tilling microarray data. Bioinformation. **6**(6), 240–243.

Morimoto S; Martin BM; Yamamoto Y; Kretz KA; O'Brien JS; Kishimoto Y (1989). Saposin A: second cerebrosidase activator protein". Proc. Natl. Acad. Sci. U.S.A. **86** (9), 3389–93.

Morimoto S; Yamamoto Y; O'Brien JS; Kishimoto Y (1990). Distribution of saposin proteins (sphingolipid activator proteins) in lysosomal storage and other diseases. Proc. Natl. Acad. Sci. U.S.A. **87** (9), 3493–7.

Nave, K.A.(1994). Neurological mouse mutants and the genes of myelin. J Neurosci Res. **38**, 607-12.

Neufeld, E.F. (1991). Lysosomal storage diseases. Annu Rev Biochem. **60**, 257-80.

Norton, W.T. and S.E. Poduslo (1982). Biochemical studies of metachromatic leukodystrophy
in three siblings. Acta Neuropathol (Berl). **57**, 188-96.

Norton, W.T., and S.E. Poduslo (1973). Myelination in rat brain: changes in myelin composition during brain maturation. J Neurochem. **21**, 759-73.

O'Brien JS, Kishimoto Y (1991). Saposin proteins: structure, function, and role in human lysosomal storage disorders. FASEB J. **5** (3), 301–8.

Paul Shannon; Andrew Markiel; Owen Ozier; Nitin S.; Baliga,Jonathan;; T. Wang; Daniel Ramage; Nada Amin; Benno Schwikowski and Trey Ideker (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks, Genome Res. **13**(11): 2498–2504.

Polten A; Fluharty AL; Fluharty CB; KapplerJ; von Figura K; Gieselmann V (1991). Molecular basis of different forms of metachromatic leukodystrophy. N Engl J Med, **324**:18-22.

Porter, S.; L. Glaser and R.P. Bunge (1987). Release of autocrine growth factor by primary and immortalized Schwann cells. Proc Natl Acad Sci U S A. **84**:7768-72.

Przulj,N.; Wigle,D.A. and Jurisica,I. (2004). Functional topology in a network of protein interactions. Bioinformatics, **20**, 340–348.

R Macfaul; N Cavanagh; B D Lake; R Stephens and A E Whitfield (1982) Metachromatic leucodystrophy: review of 38 cases, Archives of Disease in Childhood.**57**, 168-175.

Rice, John A. (2007). Mathematical Statistics and Data Analysis.Duxbury Press. **42,** 67-73.

Ridley, A.J.; J.B. Davis; P. Stroobant and H. Land (1989). Transforming growth factors-beta 1 and beta 2 are mitogens for rat Schwann cells. J Cell Biol. **109**:3419-24.

Rintoul, D.A. and R. Welti. (1989). Thermotropic behavior of mixtures of glycosphingolipids and phosphatidylcholine: effect of monovalent cations on sulfatide and galactosylceramide. Biochemistry. **28**:26-31.

Rio G; Koschützki D; Coello G (2009).How to identify essential genes from molecular networks? BMC Syst Biol. **3**:102-09.

Rorman EG; Scheinker V; Grabowski GA. (1992) Structure and evolution of the human prosaposin chromosomal gene, Genomics. **13**(2):312-8.

Schluff, P.; B. Flott-Rahmel; V. Gieselmann; P. Zimmer; A. Das and K. Ullrich (1998). Localization of receptors for endocytosis of lysosomal enzymes on different brain cells. J Inherit Metab Dis. **21**:313-7.

Stein C; Gieselmann V; Kreysing J *et al*. (1989). Cloning and expression of human arylsulfatase A. J Biol Chem. **264**, 1252-1259.

Stewart, K.; O.A. Brown; A.E. Morelli; L.J. Fairbairn; L.S. Lashford; A. Cooper; C.E. Hatton; T.M. Dexter; M.G. Castro and P.R. Lowenstein (1997). Uptake of alpha-(L)-iduronidase produced by retrovirally transduced fibroblasts into neuronal and glial cells in vitro. Gene Ther. **4**:63-75.

Stoffyn, P. and A. Stoffyn (1963). Structure ofsulfatides. Biochim Biophys Acta. **70**:218-20.

Stoffyn, P.J. (1966). The structure and chemistry of sulfatides. J Am Oil Chem Soc. **43**:69-74.

Suzuki, K. (2003). Globoid cell leukodystrophy (Krabbe's disease): update. J Child Neurol. **18**:595-603.

Tennekoon, G.I.; J. Yoshino; K.W. Peden; J. Bigbee; J.L. Rutkowski; Y. Kishimoto; G.H. DeVries, and G.M. McKhann (1987). Transfection ofneonatal rat Schwann cells with SV-40 large T antigen gene under control of the metallothionein promoter. J Cell Biol. **105**:2315-25.

Thi, A.D.; C. Evrard and P. Rouget (1998). Proliferation and differentiation properties of permanent Schwann cell lines immortalized with a temperature-sensitive oncogene. J Exp Biol. **201**:851-60.

Toda, K.; J.A. Small; S. Goda and R.H. Quarles (1994). Biochemical and cellular properties of three immortalized Schwann cell lines expressing different levels of the myelin-associated glycoprotein. J Neurochem. **63**:1646-57.

Vacher, M.; M. Waks and C. Nicot. (1989). Myelinproteins in reverse micelles: tight lipid association required for insertion of the Folch-Pi proteolipid into a membrane-mimetic system. J Neurochem. **52**:117-23.

Watabe, K.;T. Fukuda; J. Tanaka; H. Honda; K. Toyohara and O. Sakai. (1995). Spontaneously immortalized adult mouse Schwann cells secrete autocrine and paracrine growth-promoting activities. J Neurosci Res. **41**:279-90.

Watabe, K.; M. Yamada; T. Kawamura and S.U. Kim. (1990). Transfection and stable transformation of adult mouse Schwann cells with SV-40 large T antigen gene. J Neuropathol Exp Neurol. **49**:455-67.

Werner, H.; M. Jung; M. Klugmann; M. Sereda; I.R. Griffiths and K.A. Nave. (1998). Mouse models of myelin diseases. Brain Pathol. **8**:771-93.

Yaghootfam, A.; F. Schestag; T. Dierks and V. Gieselmann. (2003). Recognition of arylsulfatase A and B by the UDP-N-acetylglucosamine:lysosomal enzyme Nacetylglucosamine phosphotransferase. J Biol Chem. **278**:32653-61.

Yu,H.; Kim,P.M.; Sprecher,E.; Trifonov,V. and Gerstein,M. (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput. Biol.**73**,59-64.

9. APPENDIX

Appendix I: SQL queries for processing protein–protein interaction data obtained from MiMI plugin, in Oracle 10g Database Management System.

```
CREATE TABLE  "AIIMS1"
   (   "ID" NUMBER,
       "GENE" VARCHAR2(30),
        CONSTRAINT "AIIMS1_PK" PRIMARY KEY ("ID") ENABLE
   )
/
CREATE TABLE  "AIIMS2"
   (   "ID" NUMBER,
       "ID1" NUMBER,
       "ID2" NUMBER,
       "GENE_NAME1" VARCHAR2(4000),
       "GENE_NAME2" VARCHAR2(4000),
        CONSTRAINT "AIIMS2_PK" PRIMARY KEY ("ID") ENABLE
   )
/
CREATE OR REPLACE TRIGGER  "bi_AIIMS2"
  before insert on "AIIMS2"
  for each row
begin
  for c1 in (
    select "AIIMS2_SEQ".nextval next_val
    from dual
  ) loop
    :new."ID" :=  c1.next_val;
  end loop;
end;

/
ALTER TRIGGER  "bi_AIIMS2" ENABLE
/

SELECT GENE
FROM AIIMS1,AIIMS2
WHERE AIIMS1.ID = AIIMS2.ID1;

SELECT GENE
FROM AIIMS1,AIIMS2
WHERE AIIMS1.ID = AIIMS2.ID2;
```

Appendix II: **Normalized centrality index (SCORE *I*) for all genes involved in MLD protein- protein interaction network.**

GENE NAME	***SCORE I***
TAF1	747.5212087
SMAD2	646.9450308
BRCA1	562.4871301
HNF4A	553.005314
AR	484.9451925
SMAD9	419.2151898
CDC2	416.0214789
RB1	415.9210528
UBC	415.0906131
CDK2	413.5876164
UBB	413.5730965
PSAP	397.6550792
ANAPC10	392.6502057
CDC23	391.3418829
MYC	387.2418442
MNAT1	383.831744
CCNH	382.3370698
CDK7	381.3404889
CDC27	381.2638475
TP53	379.242666
CDC16	377.5941342
CDKN1A	374.3586452
CDC25A	371.6090863
UBE2D1	368.7256054
UBE2C	368.1183305
CDC26	367.7928177
FZR1	367.0962171

SMAD3	366.886843
LOC729948	366.8359585
ANAPC4	366.2309342
LOC730314	365.6949616
CDKN1B	365.2963005
ANAPC2	364.6175973
ANAPC1	364.3057073
ANAPC7	364.2794019
CCNB1	361.1748873
ANAPC11	361.0961297
ANAPC5	360.2940108
CDC6	356.515173
CDC20	356.1679529
UBE2E1	355.9524951
BUB3	355.2756671
ORC1L	355.1152718
BUB1B	354.3007697
MAD2L1	353.4248162
PTTG1	353.3694199
CDC25B	351.9598547
ORC5L	350.3714013
ORC3L	349.6877287
PLK1	349.5862661
MCM3	349.3653832
MCM8	349.2763361
ORC6L	348.5686112
MCM5	347.085913
AURKB	343.0523684
ORC2L	339.3579199
ORC4L	337.8763026
MCM7	335.4237801

NEK2	334.5537337
CDT1	334.3503563
MCM6	334.3471853
MCM4	334.097321
E2F4	333.8887679
AURKA	333.3824952
ESR1	331.0040489
RPA2	330.1230435
JUN	328.6891258
SMAD4	325.5501426
MCM2	322.1628447
E2F1	314.5712598
ATR	310.97571
MAX	294.5083353
EP300	290.7062687
POLR2A	281.4022193
GMNN	274.8191294
BMPR2	274.544999
CREBBP	274.225998
CLSPN	273.9703075
COPS6	272.9752071
RBL2	271.466884
POLR2K	261.4591552
RELA	259.1309668
GTF2F1	251.0905902
NR3C1	248.9397057
POLR2B	247.8752513
POLR2C	246.9175718
CTNNB1	244.7559184
GTF2F2	244.0809087
ARSA	243.1730769

ATF2	242.7004655
MAPK1	242.6505876
SMAD1	240.7029348
SMC1A	240.4197293
CCND1	239.7241032
SP1	238.8570997
SNRP70	237.857716
ATM	237.7708895
PAPOLA	237.7378542
HDAC1	237.6715911
CSTF1	237.2986349
TSC2	237.0411043
GTF2H1	233.0014393
STAT3	231.0216735
GTF2H4	229.0644495
PRPF6	228.5099408
MDM2	228.3589142
SRC	227.9570241
TGFBR1	227.5865843
TBP	226.3222142
SF3B1	225.9608491
CSTF2	224.5812223
CDC40	224.2055978
RBL1	223.4022922
CTSD	220.7391566
DHX9	220.6971249
CCNA2	220.5711752
GRB2	218.259358
PRKDC	217.4698558
ABL1	216.8352979
XRCC6	216.26725

CSNK2A1	215.8053656
CDK4	215.4054913
BTRC	214.1324418
SNRPG	213.4664103
TSG101	213.3615749
EGFR	212.6245028
MLH1	212.3249275
NFKB1	211.3269214
DEGS1	210.9040197
GTF2E1	210.6723105
UBE2I	209.6231212
STAT1	208.9296194
NCOA6	207.8550041
PML	206.7021002
NCOA3	206.5104355
RPS27A	206.2538824
PTPN11	206.2347175
PIK3R1	206.1458807
COBRA1	205.2771078
RPL31	205.2524775
TRAF6	205.0892714
BARD1	205.0638008
HSP90AA1	205.02548
AKT1	204.9615451
MSH2	204.8699056
PTPRH	204.560985
HSPA8	204.4221704
RAD50	204.2454889
SMURF1	202.8159178
GALC	202.376812
CAV1	202.1974109

SMAD7	201.7580308
VCP	200.8906449
UBE3A	200.8274177
EEF1A1	200.6978361
RAF1	200.6740292
GSK3B	200.6129267
XAB2	200.3576295
PIAS1	200.2775599
CDK9	199.9365895
SHC1	199.4914565
FBXO5	198.4614916
SMURF2	198.3543494
MAPK3	197.7624107
SKP1	197.5773279
ERBB2	196.9030155
TCEB2	196.7457737
CUL1	196.3220871
ZFYVE16	196.042801
SMAD6	195.0888223
BMPR1A	195.0311772
MRE11A	194.8344366
HNRNPR	194.5244825
RPS2	194.2733558
HTATIP	193.0372685
UCP2	192.5813543
RPS15A	191.7993047
PXN	191.7102885
MAPK8	191.6425827
POU2F1	191.480395
COPS5	191.3313032
SMARCA4	191.0602105

BMPR1B	191.0574068
RPS14	190.8253697
BLM	190.8197299
PIAS3	190.4898371
YWHAG	190.4417747
DAXX	190.2464216
RPL15	190.061881
HDAC3	189.8861249
BRCA2	189.659149
BMP2	189.4847762
RAD51	189.2889665
CCNE1	189.2510209
EIF3E	189.1230274
RXRA	188.7124876
TP73	188.5137413
CCNA1	187.8114134
STAT5A	187.5847605
NCOA2	187.5846505
PTPN6	187.5317896
INSR	187.331353
ACPP	187.1741885
FAU	187.0824691
RFC1	187.0120865
PCAF	187.0028281
CHEK2	186.6101755
DCLRE1C	186.2416429
TRIM24	186.1652634
CUL5	186.0592392
JAK2	185.9509894
GLB1	185.9315964
ACVR2A	185.3071954

JUNB	184.8868288
NCOA1	184.7245819
TRIP4	184.3520334
TOP2A	184.3517549
JAK1	184.1059764
NFKBIA	183.7550481
STUB1	183.7411652
ARHGEF7	183.7308799
YWHAB	183.5062725
STRAP	183.1743394
HMGB1	183.1736419
PTK2	183.0238616
KPNB1	182.9731858
MSH6	182.6986446
HDAC2	182.6752761
H2AFX	182.4472023
ACVR2B	182.2087736
NCOR2	181.9487077
CASP3	181.7218254
EIF3F	181.5971076
SQSTM1	181.1798642
SUMO1	180.9915623
CPSF4	180.9447112
MED1	180.9210401
XRCC5	180.8079983
RBBP8	180.7177704
TGFBR2	180.2821431
EIF3C	179.9837522
KHDRBS1	179.842994
FOXO3	179.8426779
FOXO1	179.7192235

TDG	179.622586
TOPBP1	179.5892953
GBA	179.3804182
RARA	179.3244774
UBA52	179.178597
UXT	179.1475495
SMAD5	178.692929
YWHAZ	178.6640941
MAP3K3	178.4796029
CREB1	178.4644572
GNB2L1	178.3497371
JUND	178.294176
KLK3	178.2657332
NUP153	178.1576486
IRS1	177.5074412
BRCC3	177.2053228
TCEB1	177.0877557
LCK	177.0319725
AXIN1	176.4659213
NBN	176.3230223
UBE2D3	176.0064752
MDC1	175.9459832
COPS3	175.8925119
TP53BP1	175.7593893
VIM	175.5941408
FYN	175.5747467
PRKCA	175.4271204
A2M	175.1578648
RFC5	174.767059
FHL2	174.6809765
FLNA	174.5290353

ETS1	174.3274216
ACVR1	174.3147702
NFKBIB	173.972924
HIPK2	173.8501692
SKI	173.694338
PIAS4	173.6421025
BCAR1	173.6207384
COPS2	173.3320132
RAD23B	173.2271796
RHOBTB1	173.1861754
PAK1	172.8022733
AATF	172.6312655
CSNK2B	172.539201
CKS1B	172.2789823
NFYA	172.0457828
CDK6	172.0231369
NEDD9	171.8573129
APC	171.7133137
RUNX2	171.5398234
FBXO3	171.3112791
DLEU1	170.8950855
RASL12	170.8371161
DNAJA3	170.68999
GHR	170.6536088
CSNK1A1	170.6132347
PSMD11	170.2844605
TOB1	170.2248344
BRE	170.216888
SETDB1	170.0964322
KPNA2	169.9777301
SKIL	169.5168229

DDB2	169.417876
PSMD8	169.4007005
YWHAQ	169.2632851
HGS	168.9864665
STAT5B	168.75267
SMARCB1	168.730303
VCL	168.6935273
PRMT1	168.5305227
NCOR1	168.1679541
ESR2	168.0800821
RASD2	168.0395408
ERBB2IP	168.0360685
NPM1	167.9867047
COPS8	167.766895
PRKACA	167.3175658
LYN	167.1144874
NRIP1	166.8912717
TGFB1	166.8653363
HOXA9	166.7708072
BRIP1	166.5774455
CTBP1	166.4577519
RAC3	166.4456902
PTPN2	166.1995221
ARHGEF6	165.1121409
E4F1	165.0238692
PELP1	164.2187775
HIST1H2BM	164.1491988
RHOD	164.0812908
TYK2	163.7475436
GAPDH	163.7063025
PSMD4	163.1458535

TGIF1	163.1143469
SNIP1	163.0754273
PPP1CA	162.970187
RHOJ	162.8110185
MED17	162.5753946
RNF11	162.2655914
TRRAP	162.1041859
GSTO1	161.7236668
LEF1	161.7115357
ETS2	161.5965182
RBBP4	161.5688186
DAB2	161.3974236
COPS4	161.0968061
SNW1	160.8773884
ZFYVE9	160.8716674
ASH2L	160.5559716
HSPA1A	160.3809352
EPOR	160.3788292
YY1	160.0187881
SOCS1	159.8670318
RAN	159.7732208
SFPQ	159.6972459
ONECUT1	159.685963
MAP2K3	159.5405348
RANBP9	159.4417763
APOBEC3G	158.9627449
MBD1	158.9265634
CSH2	158.8271327
RBX1	158.546085
TK1	158.1367569
GLI3	158.0432116

NUP50	157.8588741
ARL4D	157.6557658
MAP3K8	157.6119647
TUBB	157.3718306
NAGK	157.0585633
PRKRA	157.0139259
NONO	156.9703409
ZNF8	156.9638293
EID2	156.7451357
NUP214	156.6202775
SMARCA2	156.5140572
MED24	156.3632028
TERF1	156.2512742
SMARCC1	156.1918874
CALM1	156.0284252
HIST2H2AC	155.5193739
ARHGEF1	155.4727472
RBBP7	155.4504594
WWP2	155.4216947
UBQLN1	155.3498157
SUCLA2	155.2455667
ARNT	154.8806744
PARD3	154.8011165
RAD9A	154.6870686
DMAP1	154.6605901
RAB38	154.5363598
PSMC3IP	154.4847455
MAPK9	154.1157898
MED15	154.0065498
GTF2I	153.9881257
MMP2	153.9039551

HDAC6	153.7405285
BACE1	153.7073035
UCHL1	153.6784356
NCOA4	153.6643238
TUBG1	153.6176209
SMPD1	153.3450082
GPS1	153.2569032
BDNF	153.0335793
ATP6V1D	152.8979176
ACVR1B	152.7953198
CASP7	152.7017658
ZEB2	152.5142234
CALR	152.4737495
ITSN1	152.4361134
DNAJA1	152.4319129
ROCK1	152.400003
PA2G4	152.3791785
CISH	152.3022691
GAK	152.2985378
ARHGAP9	152.2064757
GCN5L2	152.1024588
CDKN2C	151.9823281
MED13	151.9410433
PTEN	151.8926776
SPTBN1	151.7314744
PRKAR1A	151.7010961
DVL1	151.6189155
CBL	151.5915167
INS	151.3044567
PCTK1	151.1866262
UHMK1	151.1259633

IFI16	150.5972374
DOK1	150.5461752
ACTB	150.4516319
PPP2R1A	150.3083241
LMO4	150.237295
TGFB1I1	150.1992749
EMD	149.9058588
PPP2R2B	149.889661
HDAC4	149.803755
ZEB1	149.5574122
LNPEP	149.1243723
NFYC	149.1135313
TUBA4A	149.0890903
WRN	148.9629618
MED23	148.8205992
PIAS2	148.7992079
CDC37	148.7018038
SOCS3	148.6284487
SMARCC2	148.5605103
FOXH1	148.3863872
DNAJC7	148.344394
PLSCR1	148.0549751
FOXG1	147.8077748
RUNX1	147.4323183
MYCBP	147.3948722
HIST2H3C	147.3900002
NR2C2	147.3773756
STAMBP	147.1830548
HIPK3	146.9781283
CARM1	146.8259195
ATF1	146.7352009

CASP8	146.6504739
NUAK2	146.6492254
PAPPA	146.5183719
PPARD	146.5170678
TRIM35	146.3277426
RNF14	146.1284281
SLPI	145.5044001
HLA-B	145.5019418
NMI	145.4639301
ELK1	145.2793262
TMF1	145.2510124
RFX1	145.2175173
PPP2R1B	145.095661
TUBB2A	144.9912076
RBAK	144.859061
WDR8	144.7869884
RGS3	144.7786545
ASAH1	144.6830166
TNNT1	144.5408835
APP	144.5107894
CDK8	144.4659421
GFI1B	144.2600586
MYBBP1A	144.1587125
GOPC	143.943011
GSK3A	143.6700416
POU5F1	143.6346742
CFTR	143.4371497
CTSL1	143.0751357
RHEBL1	143.0357204
SGK1	143.0117331
IL6ST	142.9475956

TGFBRAP1	142.9378095
C1orf181	142.9136009
CHMP1B	142.750061
AP1B1	142.6678798
PTPRC	142.6215492
BAG1	142.5172937
PSEN1	142.5052253
NFE2L1	142.4663605
CSH1	142.4592896
OTUD7B	142.4513385
GGA1	142.3377539
CYP11A1	142.3104151
RNF4	142.2853938
SRI	142.2706437
SUMO4	142.1115148
RNF5	141.9634132
CSNK1D	141.8619426
NR5A1	141.846937
UBE2L3	141.7820791
DAP3	141.7138192
EZR	141.6806935
CD4	141.6583999
NOTCH4	141.6477122
UBQLN4	141.6466527
SLC2A4	141.1938039
SNURF	141.1679376
LRP1	141.0959114
DKK1	141.0126048
HOXA13	141.0110849
TCF4	140.9461164
FTL	140.8015563

MED14	140.783216
DNAJB1	140.7334449
CETN2	140.6469343
ST13	140.4483325
UBE2G2	140.4376143
PRKCE	140.3903489
GLA	140.3476479
SAT1	140.3424189
NGF	140.3413792
TSC22D1	140.162575
MEF2A	139.6184085
AES	139.346911
PSMA2	139.3362436
NR0B2	139.3265134
HMMR	139.3079352
ZNF24	139.2349237
FN1	139.2314713
SMG1	139.1849597
SLC2A1	139.1251436
B2M	139.0726064
RAB34	139.0644404
PSG9	139.0425654
STX5	138.9022121
UBA1	138.4685419
BAZ1A	138.4247831
PABPC4	138.4118228
PNRC2	138.2583985
BLK	138.2045041
WIPI1	137.9698338
TGM2	137.9049495
MEF2C	137.820466

BTBD2	137.6569948
XRCC1	137.6257669
MED21	137.6103712
IDE	137.5249969
HEY1	137.380408
PPP2R2D	137.334572
DAZAP2	137.2370438
AP2A1	137.1778001
STAG1	136.7809546
MAFF	136.763799
PLEC1	136.719186
GGA3	136.6533482
VPS4B	136.6267661
LSP1	136.6198199
STAMBPL1	136.571443
KIF1A	136.5689868
GRN	136.3987154
KPNA6	136.3345442
SERPINA1	136.29977
MYST2	136.1766044
NFKBIE	135.9509746
DNMT1	135.9159412
IQWD1	135.765061
LRPAP1	135.7411562
DCC	135.7321424
PGR	135.4305746
HNF1A	135.1780468
KPNA1	135.16294
CHMP4A	135.1093739
CD74	135.0411269
FLI1	135.0154674

CD1D	134.8417879
GRB10	134.825349
HUWE1	134.6294395
SELENBP1	134.6184777
SVIL	134.3359368
RIT1	134.2859135
MIF	134.2138247
PAEP	134.1265532
XPC	134.0114562
GM2A	133.8909047
CTSC	133.8760573
NUFIP1	133.8236965
SFI1	133.7932218
GSN	133.5928869
PRKG1	133.1145894
P4HB	133.0962749
UBQLN2	133.0087261
NSD1	132.9646935
PLD2	132.9164945
FANCA	132.8896519
FURIN	132.8517004
DDX39	132.8131169
LSM10	132.7990875
CXXC5	132.7834107
SLC9A3R1	132.6830717
CTBP2	132.520139
KIF1B	132.5143834
RARB	132.4232439
LRP5	132.3075576
BGLAP	132.1496707
CAMK2A	132.1229474

CTR9	132.0859768
SERPINA5	132.0510607
SGCG	131.4783868
LAMA4	131.477236
NUAK1	131.2641523
HMOX2	131.2505511
HEYL	131.179888
PSMC1	131.1534637
SMN1	131.0444458
RUVBL1	130.9821729
PNRC1	130.875471
HCK	130.7195682
IGF2R	130.4911713
SULT1E1	130.4710181
TRDMT1	130.4084887
SERPINB9	130.3018238
ZNF350	130.1524933
LMO2	129.8868401
PAFAH1B3	129.7826439
GRIP1	129.7300939
SNAP23	129.354205
CTSA	129.0681542
FANCG	129.0375741
SORT1	128.7715447
UBR5	128.1756604
KHSRP	128.1333158
SART3	128.1236597
PRKAR2A	128.0976129
CTA-216E10.6	127.9291865
BRAP	127.79287
APLP2	127.1867127

SRY	127.1436291
IGFBP3	127.1197874
LRSAM1	127.0874709
COX5A	126.8119517
NSFL1C	126.7057522
PEX19	126.6850028
NGFR	126.6452408
XAB1	126.6277617
ZMYND11	126.6163652
TBC1D1	126.5585239
KNG1	126.4105953
PDZK1	125.9591637
PZP	125.9216975
SKAP2	125.8947799
QTRT1	125.8652638
KLK2	125.691774
H2AFY	125.5629373
COX5B	125.3584984
CASP1	125.2336786
SHBG	125.1672684
FSHR	125.1278803
DERL1	125.0753802
REPS2	125.0557429
LEMD3	124.9061225
ERH	124.9038739
USF2	124.841948
PLG	124.8077461
GSC	124.5635631
SLC9A2	124.2138958
ANK3	124.1313602
RNF123	124.0902531

VPS4A	123.9662766
RP5-886K2.1	123.9157925
MLL2	123.8069202
RCHY1	123.3954724
SCAMP1	123.3747477
IKBKE	123.2990599
GRIPAP1	123.2323294
IRAK2	123.1943554
SELS	123.1733609
TFPI2	123.1093244
MCM3AP	122.7507257
ABTB1	122.6039977
SVEP1	122.4237443
TSSC1	122.3417846
PAK6	122.3043491
CCDC106	122.294952
MED6	121.7326454
ELK4	121.5790966
SLC16A1	121.5296052
NFE2L2	121.5218908
COX17	121.4952534
NTRK2	121.2904835
MLLT4	121.1543551
FANCD2	121.0542679
PAXIP1	120.7646944
FUSSEL18	120.4272305
BTG2	120.4184775
SLC25A4	120.3837221
HOXD13	120.2729519
SULT2B1	120.2397998
GDF9	120.2122927

CEP55	120.103385
BACH1	119.5566265
WNT2B	119.532431
ZNF83	119.4919886
NEU1	119.2648525
ACACA	119.2646072
LHX9	119.0961129
MCC	118.8099114
PDIA3	118.7686376
GATAD2B	118.0484821
TRIM62	117.941236
NCSTN	117.7135032
CD1A	117.1144363
FBN2	116.4954911
ZNF557	116.3127798
DOCK8	116.2576573
CANX	115.8578404
NOS3	115.8286882
TOM1L1	115.5863719
DDC	115.5762018
ALAS2	114.7873011
CD1B	114.679849
LPL	114.2465757
UNC45A	112.15715
PFKL	111.7991297
ST5	111.2726253
NFE2	110.5864411
NEU4	110.047706
KLKB1	109.7182792
SEMG2	109.4523772
CTNNBIP1	109.2903575

DIS3L2	108.9494207
NIP7	108.3489818
SEDLP	107.7751951
TUBB4	107.5663139
SERPINA3	107.4138751
JMJD1C	107.3521789
PRKCB1	107.301781
APPBP2	107.2593503
ZBED1	106.5954977
BMP6	106.3276142
CTSB	105.7218237
TPX2	105.6287427
NR2C1	105.5615017
ZNF592	105.1873246
SORBS3	105.0228438
MAN1A2	104.9555082
SEMG1	104.9025771
DSTN	104.9004836
BMP7	104.2355919
ZNF587	103.9822082
PIR	103.9106389
NR0B1	103.4747061
KIAA0372	103.1785898
F11	102.9813553
NDN	102.8615669
NOL5A	101.8771286
CLPB	101.7992001
NGFRAP1	101.5435188
GUSB	101.5400465
GDF6	100.9538829
SECISBP2	100.9426303

NEDD4L	100.0631088
SNAPIN	99.70664831
DLST	99.2890467
TRIB3	98.59005519
ASNA1	98.15335562
RABGGTB	97.44160026
FGF8	97.06851576
GDF5	96.86030925
LDB2	96.41246023
SERPINA4	96.33753145
DLD	96.27336296
C19orf6	95.96332063
BMP4	95.66623493
CMPK1	95.3946032
CELSR1	95.27976335
IARS	94.60389886
DYNC1H1	94.47190895
IL8	94.39404603
PKP2	94.16804963
DCUN1D1	93.43558253
SGCE	93.42275187
MSH3	93.22070518
AOF2	92.90748094
TLR6	92.74400977
EIF3EIP	92.72717957
HEXA	92.69737833
SPEF1	92.29002157
MAP4K4	92.27393373
NTF4	92.11197789
M6PRBP1	91.88990826
VCPIP1	91.23236335

QARS	91.19702018
C4BPA	91.00004371
FRS3	90.64294541
OGDH	90.39684598
RBPMS	90.0025446
C14orf44	89.95007978
OS9	89.8251291
ANTXR2	89.45356524
RANBP6	89.28258011
LAMA5	88.97677976
ZNF41	88.94718618
ARID1B	88.92270664
MTCH1	88.48933995
ARFGAP3	88.43258668
KIAA1033	88.42929851
LITAF	87.88181725
CUX1	87.5556121
CD-M6PR	87.51406775
RAB27A	87.07040927
PHYHIP	87.03640118
PRAGMIN	86.86300896
SGK223	86.69366791
GBAP1	86.64270119
EAPP	86.25083352
GAL3ST1	86.22678917
PYGM	85.97760828
NIF3L1	85.36790901
LIMK1	84.20305918
CSTB	83.59912358
CST3	83.59090314
SGTA	83.53740517

SLC35A2	83.42487785
PBX2	82.57907568
S100A10	82.08940181
NEU3	82.03481992
VPS37C	81.25928717
PDCD6IP	80.71537613
VPS37A	80.60260009
UBE2G1	80.5303657
EFCAB6	80.2439171
FYB	79.74765774
VPS28	79.22188733
ANTXR1	79.05489397
XIST	79.0296737
ATXN3	79.0275717
LGALS3BP	78.72043188
RAD51L1	78.63544022
ADAM19	78.08604535
MAGEA11	78.08187583
CHMP4B	77.74667745
EDN1	77.09278791
STX1A	77.09188826
ARSF	76.32948003
TRAPPC2	76.27930841
FOXA1	75.77112952
CHMP1A	74.99950739
UBOX5	74.7445431
ARSE	74.44479668
ASB2	74.05141334
MERTK	73.90366592
RRBP1	72.55644846
SGCA	72.30425664

TNK2	71.54903401
CD8A	71.14316951
TG	70.53631486
TOMM34	70.43515845
ZSCAN4	70.16763478
CUL7	69.90217425
TINAGL1	69.62337357
ALAS1	68.48703598
SOCS2	68.14695615
PPIC	67.66542234
STS	67.47813229
SLC9A3R2	66.1160829
VPS37D	65.72139931
SHMT1	65.53508651
KIF5A	65.47987759
CCL21	65.24975154
IGFBP4	65.23169355
UBE4B	64.09683076
NCAM1	63.98108311
ENAM	61.65487344
MTMR10	61.3675476
MBTPS1	61.0057735
TRIM29	60.72760659
SNF8	59.83415737
SURF4	59.66306349
UGT8	58.78229694
SERPINB3	58.63650107
VPS36	58.61968767
VPS37B	58.06028206
TRB@	57.92974965
CHMP4C	57.53593065

BCAT1	56.89297976
FLNC	56.24210787
COQ6	55.44734002
PPM1B	55.28202507
KCNJ1	54.91093112
TBCD	54.51796779
RHOXF1	54.1408508
UFD1L	54.11080784
RHOXF2	53.49801244
MASP1	53.22632958
HLA-DRB1	52.33882022
PEG3	51.72906713
CTCFL	51.59918419
CDS1	51.49889483
PLAU	51.2651784
BAP1	51.06485564
MRPL36	50.7985134
DNAJC5	50.61411763
NTF3	50.57885425
CCDC98	50.18526863
ARSG	50.07766386
SMARCD2	50.06196204
CLEC4G	50.0579348
ARSJ	49.99381538
NEU2	49.92147551
PTHLH	49.9192877
ARSD	49.91818734
ARSK	49.91161098
CLC4G	49.90174646
SUMF1	49.90174646
SLURP1	49.89845828

ARSB	49.85242382
ARSI	49.83105068
ARSH	49.82611842
PHKA2	49.76385938
ZMIZ1	48.96917239
SPDEF	48.737356
RNASEL	48.56965904
SMPD3	48.31341034
RAD54L2	48.29180819
NISCH	48.27865549
SLC4A8	48.18387889
ETV5	47.94326157
APOL2	47.67856342
CHMP2A	47.64560231
PRAM1	47.46695569
PPP2R5E	46.94324677
CRELD1	46.33620987
ITM2C	44.60351877
SGCB	44.36844453
TRPC1	44.17313798
NBEA	42.26199786
ORAI2	41.1787517
CHPF	40.84194463
NTS	40.70461411
DST	40.68898647
NRF1	40.68018364
PTPN3	40.66730523
TDGF1	40.63945592
C1orf174	40.22744939
GGA2	39.83522384
PRKAA1	39.81681911

PMF1	39.52063967
SGCD	39.16920222
EXPH5	37.94518708
C4orf17	37.90590947
CEP135	37.74750779
PDZK1IP1	37.31768652
PNPLA2	37.16839444
RNF19A	37.15186167
DLGAP4	37.06466103
MOS	36.77911026
EDG4	36.5681438
GMCL1	36.5126316
PPFIBP2	35.83506607
GH1	35.74966999
ATP1B1	35.53796107
LAMP2	35.23290391
RMND5A	34.86869696
MGAT1	34.78978074
SIL1	34.52179442
MAN2B1	34.51686216
EVC2	34.38040286
KIAA0226	34.36725016
AFF1	34.32285979
TRIP12	34.21434999
FLJ32214	34.18968867
OTUB1	34.03350033
C10orf2	33.99075404
CPXM2	33.97266908
METAP1	33.92992279
MAN1C1	33.91183782
MTMR11	33.86580336

ZNF484	33.80826029
KRT15	33.80504103
DIAPH3	33.75400539
UTX	33.7260559
TMEM57	33.71947954
CRYAB	33.60859772
UBA6	33.45642549
VPS8	33.35120386
CHRNB1	32.40844891
RABGGTA	32.40832423
ROGDI	32.31637999
UBE2L6	32.24643687
STK40	32.09442813
WIPI2	32.08785178
MAP7D1	31.81822137
CDH10	31.76396647
TNNT3	31.71800013
SLC4A7	31.58556939
CES1	31.19614229
BPY2C	30.54811434
BPY2B	30.45768951
BPY2	30.3376711
PHF17	30.2160086
DYNLT1	29.89707309
CPE	29.58758038
AMFR	29.41732324
FIS1	29.21014426
CD7	28.76906658
DENR	28.5541532
PRRX1	27.67241123
LOC137886	27.52520091

GCGR	24.61477487
COL18A1	23.86175804
LMX1A	23.31882628
ACTR8	23.06234858
CCDC89	22.98014418
TSR1	22.9505506
FOXL1	22.8420408
CES3	22.66283523
OC90	22.62008894
PMPCA	22.60035989
TWF1	22.58227492
PDZRN3	22.32086495
DTNA	22.28507468
CADPS2	18.65063896
GPSM1	17.01379484
RTN1	16.44910701
BRI3	16.29127458
SELPLG	16.08576359
GPLD1	14.78599486
GH2	13.8425705
GCG	13.8252584
SERPINB13	13.46659532
HOXD12	9.81557744
SLC9A6	9.506277573
CST7	7.480766347
CSTA	6.831351642
NUDT18	6.160403921
RENBP	4.590258271
RND2	1.926942018
SSPN	1.880907558
CD1C	1.178992497
