ABSTRACT

Yeast As a Model Organism To Assay the Functional Relevance of nsSNPS in Mitochondrial Disorders

Chidinma Okani

Director: Bessie Kebaara, Ph.D.

This thesis explores two yeast experiments. The first experiment, which was performed during the summers of 2011 and 2012, involved developing several yeast strains (including yeast strains with human wild-type orfeomes and yeast strains with human mutant orfeomes) and analyzing their growth rates with a flow cytometry growth competition assay. In this experiment, we hypothesized that 1) the yeast wild-type strains would grow better than the yeast deletion strains because the yeast deletion strain is lacking a growth gene and 2) the yeast strains with human wild-type orfeomes would grow better than the yeast deletion strains. The results of this experiment supported both hypotheses and validate our approach as a successful way to determine the relevance of certain nsSNPs in humans.

The second yeast experiment, performed by scientist Sze Chern Lim and colleagues, involved identifying a homozygous mutation in two cousins with OXPHOS deficiency by using various sequencing techniques. After the homozygous mutation was identified, a mutant *LYRM4* gene was able to partially complement for an ISD11 deletion in yeast. For this experiment, the scientists hypothesized that L-cysteine desulferase activity of NFS1 would be minimally present when co-expressed with a mutant ISD11 protein; the results of their *in vitro* experiments supported this hypothesis.

APPROVED BY DIRECTOR OF HONORS THESIS:

____

Dr. Bessie Kebaara, Department of Biology

APPROVED BY THE HONORS PROGRAM:

____

Dr. Andrew Wisely, Director

DATE: __________________________________

YEAST AS A MODEL ORGANISM TO ASSAY THE FUNCTIONAL RELEVANCE OF NSSNPS IN MITOCHONDRIAL DISORDERS

A Thesis Submitted to the Faculty of

Baylor University

In Partial Fulfillment of the Requirements for the

Honors Program

By

Chidinma Okani

Waco, Texas

December 2013

TABLE OF CONTENTS

List of Figures and Tables iii

Acknowledgments iv

Chapter One: Introduction 1

Chapter Two: Materials and Methods 24

Chapter Three: Results 43

Chapter Four: Discussion and Conclusion 49

Bibliography 53

LIST OF FIGURES AND TABLES

Table 1: The Polymerase Chain Reaction Reagents …………………………………....28

Table 2: Bacterial Transformation Reagents……………………………………………32

Table 3: Yeast Transformation Reagents………………………………………………..35

Table 4: Transformation Reagents Prepared For Six Reactions………………………...36

Figure 1: Growth Curves…………………………………………………………………...43

Figure 2: Deletion Strains Competing with FY5……………………………………...44

Table 5: Muscle of Patient 2……………………………………………………………..46

Table 6: Liver of Patient 1 and Patient 2………………………………………………...46

Table 7: L-cysteine Desulferase Activity………………………………………………..47

ACKNOWLEDGMENTS

This project would not have been possible without the support of many people. Many thanks to my thesis adviser, Dr. Bessie Kebaara, who offered invaluable feedback to me from the very beginning. Also thanks to my mentor and committee member Dr. Michael Springer who allowed me to work in his lab in summer 2011 and has offered invaluable guidance and support to me over the past two years. Thanks to the National Institute of General Medical Sciences and the FAS Center for Systems Biology for funding my summer research internship. And finally, thanks to my parents, brothers, and numerous friends who have offered encouragement to me throughout this process.

CHAPTER 1

INTRODUCTION

Human Disease

A disease is defined as "a definite pathological process having a characteristic set of signs and symptoms" (Dorland's Medical Dictionary). A disease may negatively affect either a body or its individual parts. *Genetic* diseases are illnesses caused by genomic abnormalities; these diseases manifest in several ways. They commonly appear as chromosomal abnormalities that either originate from the parent or occur with no family history. Genetic diseases may also appear as single gene disorders (commonly known as *Mendelian inheritance disorders*), an abnormality in a single gene; these disorders have a higher risk of being inherited. Furthermore, genetic disorders may arise through the joint effect of genes and the environment and are termed "multifactorial" disorders. A fourth type of genetic disorders, teratogenic disorders, occurs in fetuses exposed to teratogens (abnormality-causing substances such as alcohol or excessive radiation) during the first trimester of pregnancy (Types of Genetic Diseases).

Approaches to Determining the Molecular Origin of a Disease

There are two main approaches to determining the molecular mechanism of a disease: a genetic approach and a phenotypic approach. In a phenotypic approach, an investigator examines the phenotypic manifestation of a disease and work backward, step by step, in order to uncover the causative factors of the disease. The lack of knowledge concerning both the number of steps in the pathogenic process and the time span between the initial

pathogenic trigger and the disease manifestation pose challenges to the phenotypic approach. Furthermore, in the process of working backward, an investigator may uncover branching in the pathogenic pathway that may lead to confounding secondary changes. The phenotypic approach, which involves determining which biochemical steps influence the disease phenotype, has relied on model systems. For example, transgenic mice and gastrointestinal irritants have been manipulated in the study of the biochemical pathway of irritable bowel syndrome. The genetic approach differs from the phenotypic approach in its focus on the particular starting point of a pathogenic process rather than the end phenotype (an end point which can be reached in several ways) (Gusella and MacDonald 2002).

Three approaches are used to identify the genes associated with susceptibility to disease: pedigree, affected sib-pair linkage studies, and association studies of population samples. A pedigree is a chart that outlines the incidence and appearance of phenotypes of a particular gene in a family lineage; these diagrams are most commonly used to outline the family trees of humans, dogs, and horses. Affected sib-pair linkages studies examine phenotype and identical copies of the same ancestral allele in order to identify which alleles potentially cause genetic diseases. Association studies of population samples examine two groups of individuals and determine whether frequencies in both alleles at a single locus or multi-locus alleles differ between the groups; these studies are unique in that they rely on the direct comparison of genomes (Freimer and Sabatti 2004).

Genome-Wide Association Study: Its Promises and Shortcomings

Another popular genetic approach is genome-wide association study (GWAS), the examination of markers across DNA sets of many people in order to find genetic variants associated with a disease. With GWAS, researchers use genetic associations to create more efficient strategies to detect and prevent disease. Information from the Human Genome Project and the International HapMap Project (completed in 2003 and 2005 respectively) allows researchers to find the genetic components of common diseases. In this approach, researchers examine two groups of participants: people with the disease being examined and people without the disease. Researchers obtain DNA from each participant which is then purified from the blood and cells, placed on a tiny chip, and scanned in a laboratory for genetic variants called SNPs (single nucleotide polymorphisms) which are potentially associated with the disease at hand. Genetic variants found more frequently among the diseased participants are said to be "associated" with the disease (Genome-Wide Association Studies Fact Sheet).

GWAS has proven effective in discovering SNPs associated with the various genetic disorders; a notable example is the use of GWAS to identify SNPs associated with varying levels of severity in sickle cell anemia. Sebastian and his colleagues conducted GWAS to discover SNPs associated with either a "severe" or "mild" form of sickle cell in 1,265 patients. Using SNP analysis, the researchers determined that 40 SNPs were strongly associated with sickle cell severity; furthermore, 8 SNPs showed statistically insignificant effects while 19 SNPs showed no effects at all. Using a more advanced SNP Set Enrichment Analysis (SSEA), 27 genes were identified as containing "a strong

enrichment of significant SNPs." A notable gene found, TNKS, regulates telomere length (Sebastiani et al., 2010).

Despite its promises to discover disease-linked gene variants that would lead to individualized drug therapies, GWAS has many shortcomings. Namely, very few SNPs are associated with a disease. SNPS are numerous in an individual's DNA, occurring once in every 300 nucleotides on average (meaning that the human genome has roughly 10 million SNPs). SNPs that occur within a gene or in a region near a gene may play an important role in affecting a gene's function; however, most SNPs have no effect on a person's health. While GWAS has identified SNPs *associated* with many common diseases, researchers have been unable to discover the specific genetic changes that affect these common diseases. Most common genetic diseases originate from rare genetic variants that occur in remote regions of the genome that GWAS do not study. The low percent of genetic information that GWAS is able to glean about a disease is termed "missing heritability" (Kepler 2012).

Model Organisms

A model genetic organism is a non-human species that is used widely in genetic studies because it has characteristics, such as a short generation time and the availability of many genetic variants, that make it particularly useful for genetic study and analysis. In recent years, the genomes of many model genetic organisms have been completely sequenced, providing scientists with much useful information to facilitate genetic research (Pierce 2010).

Drosophilia melanogaster

Drosophilia Melanogaster, a fruit fly, is one of the most widely used and best known model organisms. it is the most popular member of the genus *Drosophilia*, a group of more than 1000 species of small flies that feed and harmlessly reproduce on fruit. *D. Melanogaster* began appearing in biological laboratories at the beginning of the twentieth century. Thomas Hunt Morgan, a renowned geneticists and evolutionary biologist, began using fruits flies in heredity studies in his laboratory at Columbia University. In his famous "Fly Room," he and his students used Drosophilia to explore many basic principles of heredity including sex-linked inheritance and epistasis.

The fruit fly has many characteristics that make it an ideal model organism for genetic study. It has a relatively short generation time; at room temperature, fruit flies complete an entire generation in ten days, allowing scientists to study several generations in the span of a few weeks. Additionally, fruit flies are easy to culture in a laboratory setting. They are usually raised in small glass vials and fed paste-like food with banana, cornmeal, and molasses components. Finally, *D. melanogaster* has a relatively small genome (175 million base pairs of DNA, only 5% of the size of human genome) which allows it to be easily analyzed.

D. melanogaster has played a significant role in the study of amyotrophic lateral sclerosis (ALS), a degenerative neurological disease that leads to the gradual weakening and atrophy of skeletal muscles. Though most cases of ALS appear in people with no family history of the disease, ten percent of ALS cases are inherited. In 2004, geneticists studied

a Brazilian family with multiple cases of ALS across several generations, due to a mutation in a gene called *VABP*, which encodes a membrane protein. Geneticists turned to *D. melanogaster*, which possesses a very similar gene to *VABP* and created transgenic fly models that aid in understanding the ways that *VABP* disruption leads to ALS (Pierce 2010).

Caenorhabditis elegans

Caenorhabditis elegans, another model organism, has become widely used in genetic study due to its simple body plan, production of numerous progeny, and ability to be cultivated in a laboratory setting. Most mature adults are hermaphroditic, with the ability to produce both eggs and sperm and undergo self-fertilization. A few are male and produce only sperm. Hermaphrodites have two sex chromosomes (XX) and the males possess one sex chromosome (XO). Consequently, the hermaphrodites that self-fertilize produce only hermaphrodites while the hermaphrodites that mate with males are capable of producing both male and hermaphroditic offspring.

The *C. elegans* genome is small compared with genomes of many multicellular organisms; it only has 103 million base pairs. Its complete genome sequence provides information about its gene structure and function. For example, scientists have used this genetic information to study apoptosis (programmed cell death) in *C. elegans* and have determined that the process is similar to that in humans. This discovery has increased our understanding of apoptosis in humans and the role it plays in the development and treatment of cancer (Pierce 2010).

Mus musculus

Mus musculus, the common mouse, possesses a close evolutionary relationship with humans and is commonly used in genetic study and analysis. As a mammal, the *Mus musculus* shares many genetic, behavioral, and physiological similarities to humans. Other advantages include a small size, rapid reproduction, adaptation to life in a laboratory, and toleration for inbreeding. The mouse genome, which contains 2.6 billion base pairs, is similar in size to the human genome; it is distributed across nineteen pairs of autosomes and one pair of sex chromosomes.

In the mouse, diploid germ cells undergo meiosis to produce sperm and oocytes. At puberty, male mice begin producing sperm while female mice go through an estrous cycle (physiological changes that females undergo during a mating season) every four days. If mating between males and females occurs during estrus, fertilization can occur. Gestation occurs over a three-week period and a generation of mice can be completed in eight weeks.

Many genetic techniques have been developed for use in mice, including the creation of transgenic mice (through the injection of DNA into a mouse embryo), knockout mice (through the disruption of specific genes), and knock-in mice (through the insertion of specific gene sequences into certain loci). These techniques utilize the manipulation of the reproductive cycle. The use of mice has led to many genetic discoveries including the genetic basis of various cancers and mutations leading to human birth defects (Pierce 2010).

The Discovery of Yeast

Saccharomyces cerevisiae (commonly referred to as "baker's yeast") has been widely utilized as a simple model organism in the field of genetics. In recent years, it has been utilized in the production of biofuels. In 1857, Louis Pasteur identified *S. cerevisiae* as the microorganism responsible for fermentation. Shortly after this discovery, *S. cerevisiae* became a subject of gene analysis. In addition to having properties that allow for the study of eukaryotic genetics, yeast cells are unicellular, allowing researchers to utilize techniques often reserved for bacteria (Pierce 2010).

Life Cycle of Yeast

Saccharomyces cerevisiae can exist as either haploid or diploid cells. Haploid cells are a product of the mitotic division of yeast which produces identical daughters through budding. Yeast also goes through sexual reproduction. In sexual reproduction, haploid cells of two different mating types (a and α) fuse and undergo nuclear fission to form a diploid cell. Once formed, this diploid cell is able to produce genetically identical haploid cells via mitotic budding. When the diploid cell is short of nutrients, the diploid cell undergoes meiosis, producing four different haploid nuclei; these nuclei differentiate into different cells, producing haploid spores. All the products of a single meiotic division can be isolating with *tetrad analysis,* a process by which the four meiotic daughter cells (tetrad) are analyzed in the *ascus* (the common structure they are enclosed in) (Pierce 2010).

Advantages of Yeast as a Model Organism

Although yeast is a eukaryote and has genetic and cellular systems similar to that of other eukaryotes, it is unicellular; like many bacterial systems, yeast requires little space and can be grown in a lab both inexpensively and in large numbers (Pierce 2010).

In haploid form, yeast cells contain a single allele at each locus, allowing each allele to be expressed in the phenotype. When diploid, no dominance is observed. Consequently, haploid cells easily display recessive alleles while diploid cells display interaction between alleles.

Despite possessing a unicellular structure, yeast cells possess many of the same genes found in humans and other complex eukaryotic organisms; many of these genes have similar or identical functions in yeast. As a result, the genetic study of yeast allows scientists to broaden their understanding of other more complex eukaryotic organisms, including humans (Pierce 2010).

The Yeast Genome

S. cerevisiae contains sixteen pairs of eukaryotic chromosomes which have a high rate of recombination; this gives yeast a much longer genetic map than most organisms. In addition to housing 12 million base pairs, *S. cerevisiae* contains between 2 million and 3 million bases pairs of rRNA genes. In 1996, yeast was the first eukaryotic organism to have its genome completely sequenced (Pierce 2010).

Genetic Techniques With Yeast

Researchers often utilize yeast plasmids to transfer target genes and DNA sequences into cells. Yeast cells possess a 2μ plasmid, a plasmid that has been utilized as a vector for transferring genes into yeast. This plasmid is transferred from parent cells to daughter cells in both mitosis and meiosis; furthermore, this plasmid replicates independently in the cell because its origin of replication is recognized by the yeast replication system.

In some instances, scientists have opted to use *bacterial* plasmids in yeast. Some of the bacterial plasmids undergo recombination with the yeast chromosome and in the process, transferring their sequences to the yeast chromosome. Shuttle vectors, which can be propagated in bacteria and yeast, make it possible to manipulate bacterial gene sequences and transfer the sequences into yeast cells where their function can be tested.

While bacterial plasmids are limited in the size of DNA they can carry, yeast artificial chromosomes (YACs) are more efficient and can hold DNA fragments as large as several hundred thousand base pairs (Pierce 2010).

The Physiology of Mitochondrial and Its Related Disorders

Mitochondria are specialized intracellular compartments that are responsible for creating more than 90% of the energy needed by the body for life and growth; they are found in every cell except red blood cells. The mitochondrial genome ensures that the "powerhouses" of human cells function correctly. They are more numerous than nuclear genomes and more prone to mutation (Tymoczko 2011).

Mitochondrial diseases result from failures of the mitochondria, causing less and less energy to be generated within the cell. These diseases result in cell injury and even cell death. If they occur in cells throughout the body, whole systems begin to fail, putting the life of the person in danger. Although these diseases mainly affect children, they are more common in adults than ever before. They cause the most damage to cells of the brain, heart, liver, skeletal muscles, kidney as well as the endocrine and respiratory systems. There are many symptoms of mitochondrial disease which include: loss of motor control, gastrointestinal disorders, poor growth, cardiac disease, liver disease, diabetes, respiratory complications, seizures, visual/hearing problems (Tymoczko 2011).

Mitochondrial disease can be inherited in several ways. With maternal inheritance, a mother with a mitochondrial DNA (mtDNA) mutation can pass this gene to all of her children; her children will be affected to varying degrees, though not necessarily in the same way the mother is affected. With autosomal recessive inheritance, a child can inherit mitochondrial disease if both parents are carriers the disease and pass on mutated mitochondrial nuclear DNA (nDNA); in this case, families have a one-in-four chance of having a child with the disease. With autosomal dominant inheritance, a parent with a dominant nuclear DNA mutation will pass on the gene to 50% of the children (How Is Mitochondrial Disease Inherited?).

Mitochondrial Oxidative Phosphorylation and Electron Transfer

Oxidative phosphorylation is a process by which mitochondria regenerate ATP through a series of intricate steps involving enzymes, energy release, and electron transfer. This process takes place through an *electron transport chain,* a transfer of electrons from a set of electron carriers (NADH and $FADH_2$) to a set of oxygen molecules (the electrons that convert these oxygen molecules to water molecules via reduction). The inner membrane of mitochondria contains five multiunit enzyme complexes (I-V) and two electron carriers (Coenzye Q and cytochrome c) that undergo oxidation-reduction reactions in the process of facilitating electron transfer. During this flow of electrons, protons are moved from the matrix to the intermembrane space of the mitochondria. This newly formed proton gradient is utilized to generate energy in the form of adenosine triphosphate (ATP) (Tymoczko 2011).

Mitochondrial Disorders

Mitochondrial respiratory chain diseases, the commonest of mtDNA disorders, are notable for their clinical heterogeneity (their very diverse clinical manifestations, spanning from a single organ to multiple organ systems). The commonest mtDNA-related clinical myopathy is chronic progressive external opthalmoplegia (CPEO), a disorder characterized by a progressive inability to move the eyes and eyebrows. MELAS (myopathy, encephalopathy, lactic acidosis, and stroke-like episodes), the commonest mitochondrial encephalomyopathy, often appears during childhood and causes a variety of phenotypic expressions including deafness, seizures, dementia, ataxia, and stroke.

Leber's hereditary optic neuropathy (LHON) is the commonest cause of blindness in males. Characterized by convoluted retinal arteries and optic atrophy, LHON manifests between late adolescence and early adulthood and takes complete effect in the fourth or fifth decades of life. A notable example of a multisystem respiratory chain disorder is Pearson's syndrome, which manifests in a reduction of healthy red and white blood cells and the failure of exocrine pancreatic functions (simultaneously affecting the circulatory and gastrointestinal systems) (Leonard and Schapira 2000).

MtDNA mutations are also associated with breast cancer, the most common cancer among women in the United States, and thyroid cancer, a cancer that more commonly affects people under 55 years of age (Cancer.org). Oncocytoma is a tumor made up of oncocytes, epithelial cells containing large amounts of mitochondria. Oncocytomas found in the thyroid and breast tissue are often malignant. In order to further understand whether specific mtDNA mutations were associated with the mitochondrial accumulation of oncocytes, Gasparre and colleagues sequenced the entire mtDNA of 45 thyroid tumors and 5 breast tumors as well as 52 control cases (21 non-oncocytic thyroid tumors, 15 breast tumors, and 16 gliomas). They found that 13 oncocytic tumors contained either nonsense or frameshift mitochondrial mutations while 2 samples in the non-oncocytic control group presented these mutations. Furthermore, a disruptive mitochondrial mutation was identified in one of the five breast tumors. Most notably, all of the disruptive mutations were found in complex I subunit genes and a significant association existed between these mutations and the oncocytic phenotype (Gasparre et al., 2007).

Although mitochondrial disorders may be inherited in several ways, many cases of the disease appear sporadically. For example, many patients with colon cancer caused by pathogenic mtDNA have no family history of colon cancer. These mtDNA mutations often occur in postmitotic tissue (neurons, osteocytes, and myoblasts) (Taylor et al, 2003). Furthermore, due to their phenotypic heterogeneity, mitochondrial respiratory chain disorders are often misdiagnosed. Because many physicians lack knowledge of the clinical manifestations of respiratory chain diseases, very few investigations are made. Furthermore, the lack of a sufficient gold standard for diagnosing these disorders renders a multidisciplinary effort necessary. Additionally, it is unclear whether these disorders should be considered distinct or if they should be grouped in categories Finsterer et al., 2001).

The complexity of mitochondrial disorders requires a biochemical, clinical, and histological approach to both classification and diagnostic testing. Confirmative molecular diagnostic techniques are challenging because of heteroplasmy (the large number and complexity of genomes and mtDNA). Usually, the first diagnostic step is the screening of point mutations and large deletions of the mtDNA. If causative nuclear genes can be identified, direct sequencing of white blood cell DNA is the next step. Furthermore, mtDNA depletion syndromes can be screened for by measuring mtDNA content of affected liver and muscle tissue (Wong et al. 2010).

Neurologists use two major algorithms to diagnose patients who may have a mitochondrial respiratory chain disorder. The Bernia criteria examines respiratory chain

enzymes of muscle and liver tissue in children and adults and classifies a number of features into "major" and "minor" categories. Neurologists then tally up these categories before sub-classifying patients as "definite," "possible," or "probable." The main problems with applying the Bernia criteria include the lack of child-specific criteria and the difficulty of dividing continuous data into discrete "major" and "minor" categories (Bernier et al, 2002). The Morava criteria rely on clinical and biochemical observation of children with a suspected mitochondrial respiratory chain disease. Muscle biopsies are collected from two groups of children: children with suspected mitochondrial disorders and healthy children. Based on a predetermined formula, patients are assigned a score on a scale of 1 to 12 with 1 being "unlikely," 2 to 4 being "possible," 5 to 7 being "probable," and 8 to 12 being "definite." Although the Morava criteria have a more child-specific classification system, it still involves invasive biopsy procedures which, when performed under anesthesia, pose a unique risk to patients with respiratory chain disorders and related abnormalities (Morava et al. 2006).

While several mitochondrial disease therapies have been proposed, none have been approved by the Food and Drug Administration (FDA). Consequently, all therapies currently used in the treatment of mitochondrial disease are unapproved treatments, off-label treatments, and minimally-effective dietary supplements—therapies that are limited to merely treating the symptoms of disease. The exceptions are treatments for diseases caused by dysfunctions in Coenzyme Q10 such as Parkinson disease, progressive supranuclear palsy, and statin-related myopathy (Hurko 2013).

Traditionally, scientists have utilized the Sanger method and restriction length fragment polymorphism (RLFP) analysis to sequence DNA—two methods that focus on individual genes and DNA fragments. However more advanced “next generation” sequencing (NGS) technologies now allow for the simultaneous sequencing of multiple DNA fragments. Calvo and colleagues performed NGS on 42 infants suspected of having mitochondrial respiratory chain disease. Specifically, they performed “MitoExome” sequencing, the sequencing of the mitochondrial DNA and exons of hundreds of nuclear genes that encode the mitochondrial proteome. This sequencing approach confirmed that 55% of patients had recessive genes or pathogenic mtDNA, 24% of patients had mutations in genes previously linked to disease, and 13% of patients had mutations in genes unrelated to disease (Calvo et al., 2012). In a similar case, Leiber and colleagues used NGS to target both mtDNA and the exons of over 1,600 nuclear genes of the mitochondria. In a patient with multiple disorders, this approach identified an unexpected missense mutation in gene *WFSI* (a mutation linked to Wolfram syndrome) and provided a possible link between *WFS1* and thiamine metabolism (Lieber et al., 2012).

A Comparison of Yeast and Mammalian Mitochondria

Yeast contains many mitochondrial proteins that are orthologs to human proteins. As many as 700 proteins in yeast are needed for the function of mitochondria. Out of 601 known yeast mitochondrial proteins, 222 of them are shared with humans. Many modern approaches used to study mammalian mitochondrial proteins involved the creation of yeast model systems. These approaches include the use of mass spectrometry, gene

expression analyses, fluorescent reporters for sub-cellular analysis, and deletion phenotyping (Tymoczko 2011).

Modeling Human Mitochondrial Disorders in Yeast

Knowledge of yeast mitochondrial synthesis has progressed in recent years. This knowledge has increased concerning the important mechanism that moderates preproteins synthesized in the cytoplasmic ribosomes through the mitochondrial inner and outer membranes. It also relates to the inner membrane insertion mechanism of polypeptides that are encoded in the mitochondria and proteins that participate in the assembly and control of ATP synthase and respiratory complexes. Recently, synthesis of iron-sulfur clusters and dynamics of the mitochondrion have emerged. Many yeast proteins that have been newly discovered have equivalent orthologs in human DNA. Consequently, *S. cerevisiae* has proven a fitting yet simple organism for approaching the molecular basis for an increasing number of human mitochondrial diseases that are caused by mutations in nuclear genes that have been identified by positional cloning.

Yeast Complementation Study 1:
Complex III Assembly and Mutations in Human Gene BCS1L

A complementation study is a study in which a gene product from one organism fully or partially replaces the function of a gene product in a separate organism. This approach allows scientists to both identify the human orthologs of yeast mitochondrial proteins and confirm the harmful effect of mutations affecting oxidative phosphorylation functions associated with human mitochondrial disorders.

Complex III, also called coenzyme Q: cytochrome *c* — oxidoreductase, is the third complex in the electron transport chain, a mitochondrial phenomenon that couples electron transfer between an electron donor and an electron acceptor in order to build a proton gradient that will be utilized to generate energy in the form of adenosine triphosphate (ATP). Complex III is formed by three catalytic subunits and seven non-catalytic subunits. The cytochromes b and c_1 and several non-catalytic subunits form a bc_1 precomplex needed for the assembly of the Rieske iron-sulfur protein (RIP), a component that plays a key role in electron transfer. Though several genes have been identified as coding for proteins that assist in the assembly of bc_1, only one of the genes has a function that is fully understood: the BCS1 gene. The BCS1 gene codes for a protein that is uniquely involved in the assembly of RIP. BCS1 acts as an ATP dependent chaperone that maintains the integrity of the precomplex prior to RIP assembly. Complex III deficiency is a rare cause of human mitochondrial disorder.

Dr. De Lonlay and his colleagues studied the genetic origin of Complex III deficiency in twelve patients from ten unrelated families. Six people from four Turkish families (three of whom shared a common ancestor) were found to carry BCSIL mutations that were not detected in the 150 control individuals. The phenotypic manifestation of Complex III deficiency in these families included metabolic acidosis at birth, neonatal tubulopathy, and hepatic involvement that was consistent with either liver failure or hepatic cytolysis.

A homozygous G to A transition was detected at nucleotide 830 in exon 5 of the gene, which resulted in a substitution of a conserved serine to asparagine residue in two

affected siblings and one fetus from the same family. Furthermore, two patients from unrelated consanguineous families carried the same homozygous C to T transition at nucleotide 296 in exon 1, which caused the substation of a leucine with a conserved proline residue. The sixth patient studied was a heterozygote containing a C to G transition at nucleotide 464 in exon 3, which changed a conserved valine to a methionine.

Since human BCS1 is able to partially restore the respiration in yeast *bcs1* mutants, four of the yeast mutations were introduced into the human BCS1L cDNA via site directed mutagenesis. Both the wild type and mutant genes were placed under the control of the yeast ADH1 promoter in the shuttle plasmid Yep351 and the prepared constructs were used to transform a respiratory defective haploid yeast strain with a null *bcs2* allele. Complementation was tested by examining respiration in isolated mitochondria and by measuring the growth of the mutant on the respiratory substrate glycerol. These experiments confirmed that the four mutations affected protein function. Furthermore, the complementation studies in yeast supported the conclusion that BCS1L mutations are responsible for the Complex III deficiency in patients with fatal tubulophathy, liver failure, and encephalopathy (Shoubridge 2001).

Yeast Complementation Study 2:
Complex IV Assembly and Mutations in Human Gene COX10

Complex IV, also called cytochrome c oxidase (COX), is a large transmembrane mitochondrial protein. The last enzyme of the electron transport chain, it catalyzes

electron transfer from cytochrome c to molecular oxygen. Complex IV is formed by thirteen unique subunits encoded by both nuclear and mitochondrial DNA.

Cytochrome oxidase deficiency is one of the most common causes of human mitochondrial disorders. This deficiency's phenotypic manifestations include hepatic failure, encephalomyopathy, and Leigh syndrome. Though mutations have not been detected in the COX subunits encoded by nuclear DNA, mutations have been detected in DNA products involved in COX maturation and assembly. One example is the gene products of the COX10 gene.

Dr. Valnot and his colleagues studied COX deficiency in members of a family with leuodystrophy and tubulopathy. Homozygosity mapping identified a region with two genes that appeared to be linked to COX deficiency: the SCO2 gene and the COX10 gene. The SCO2 gene encodes protein involved in the mitochondrial transport and insertion in Cytochrome oxidase. The COX10 gene encodes the heme A:farnesyl transferase that converts protoheme to heme O, a product that catalyzes the first step in the conversion of protoheme to heme A, the heme prosthetic group of COX1. Four affected individuals had COX10 sequences that contained homozygous C to A transversion at nucleotide 612 of the cDNA which corresponded to exon 4 of the gene; this changed an asparagine to a lysine residue in the protein. This substitution was heterozygous in both consanguineous parents and was absent in 200 chromosomes 17 from 100 unrelated controls of similar ethnic origin.

In order to confirm the harmful effect of the substitution in heme O-farnesyl transferase, both the human wild type and mutant genes were tests for their ability to complement the yeast cox10 null mutant. This yeast mutation was introduced into the wild type COX10 cDNA clone via PCR. Furthermore, the wild type and yeast mutant genes were cloned in three different types of plasmids. These plasmids were used to transform a selected yeast cox10 null strain. When present in low copy, growth of the mutant on rich glycerol medium was restored but the growth of the wild type was not. The inability of the mutant gene to complement the yeast cox10 null strain when expressed in low copy strongly supported the conclusion that the COX deficiency was caused by the substation (Shoubridge 2001).

Thesis Project: Objective and Hypotheses

This paper explores two yeast experiments. This first experiment, which was performed during the summers of 2011 and 2012, involved incorporating wild-type and mutant version of human genes (DLST, MPV17, ATP5A1, C6ORF57, and SDHAF2) into a yeast genome in order to develop human wild type yeast strains and human mutant yeast strains. Afterward, the growth rate of these two types of strains was analyzed in comparison to the growth rates of yeast wild type strains (with the deletion of a neutral locus) and yeast deletion strains. The comparison between different strains was accomplished by means of a growth competition assay in which two yeast culture are grown together in equal ratios and under identical environmental conditions and measured over time in order to quickly detect growth rate differences and establish a

fitness relationship. In this growth competition assay, this fitness relationship has implications for how mutations in the human genome may lead to unfavorable phenotypes.

For this experiment, we made two hypotheses. First, we hypothesized that the yeast wild-type strains would grow better than the yeast deletion strains because the yeast deletion strain is lacking a crucial growth gene. Secondly, we hypothesized that the human wild-type strains would grow better than the yeast deletion strain (because if the human wild-type could not at least partially restore the function of the deleted yeast gene, we would have no hope of determining if the human mutant gene was defective.

The second yeast experiment, performed by scientist Sze Chern Lim and colleagues, involves identifying a homozygous mutation in two cousins with OXPHOS deficiency. The mutation was in gene *LYRM4*, a gene which codes for the ISD11 protein that forms complex with and stabilizes the sulfur donor NFS1 (a cysteine desulferase needed for ISC assembly). In one patient with a deficiency of complexes I, II, and III (three complexes with ISCs), the homozygous mutation was identified via parallel sequencing of >1000 mitochondrial genes (MitoExmoe sequencing). In the patient's similarly affected cousin (who died as a neonate because of a more severe phenotype), Sanger sequencing was used to identify the same homozygous mutation (furthermore, complex IV was deficient in her skeletal muscle). After the homozygous mutation was identified, a mutant *LYRM4* gene was able to partially complement for an ISD11 deletion in yeast. For this experiment, the scientists hypothesized that L-cysteine desulferase activity of NFS1

would be minimally present (or not present at all) when co-expressed with a mutant ISD11 protein.

CHAPTER 2

MATERIALS AND METHODS

The Genes

FY5

FY5 is a prototrophic (grows with just ammonia as a nitrogen source version of BY4741 (S288C)

DLST

DLST (dihydrolipoamide S-succinyltransferase) encodes a mitochondrial protein that plays a role in catalyzing the conversion of 2-oxoglutarate to succinyl-CoA and CO_2 (NCBI)

MPV17

The MPV17 gene (also called "MpVI7 mitochondrial inner membrane protein") codes for a protein with a function that is currently unknown (may be involved in maintaining integrity of mtDNA). MPV17 proteins are found in the inner membranes of mitochondria. Though their function is unknown, these proteins are affiliated with various cellular activities including energy production and chemical signaling (NIH).

ATP5A1

ATP5A1 gene encodes a subunit of mitochondrial ATP synthase, an enzyme that catalyzes the synthesis of ATP in the mitochondria. ATP synthase consists of an F_1 core and an F_0 component reaching into the membrane (NCBI)

C6ORF57

This gene codes for the UPF0369 protein C6orf57 precursor, a mitochondrial protein (GeneCards).

SDHAF2

This gene is also called the succinate dehydrogenase complex assembly factor 2. It codes for a protein (SDHAFT protein) that associated with the succinate dehydrogenase (SDH) enzyme. This protein also helps the FAD molecule attach to the SDH enzyme (NIH).

Yeast Wild Type Strain

The BY4742 strains were the parental yeast strains from which each deletion strain was derived.

Yeast-ORFeome Strain: YBR 269-C huC6ORF57

This strain is a yeast wild type strain in which a crucial growth gene is deleted and replaced with a human wild type strain.

Yeast Deletion Strains

In the yeast deletion strains, a crucial growth gene has been removed. These strains are formed when a growth gene in a BY4742 strain is deleted and not replaced.

YBR269-C

This yeast strain contains a putative protein with an unknown function. High-throughput studies have detected this protein in mitochondria that has undergone purification (YeastGenome).

YOL071W

This yeast strain is a subunit of succinate dehydrogenase, an enzyme that couples succinate oxidation to ubiquinone reduction. This strain is required for FAD cofactor attachment to Sdh1p. The mutations found in the human ortholog PGL2 are linked to neuroendocrine tumors (paraganglioma) (YeastGenome).

YBL099W

This yeast strain is the alpha subunit of the F1 sector of mitochondrial F1F0 ATP synthase, an enzyme complex required for ATP synthesis (YeastGenome).

YDR148C

This yeast strain is Dihydrolipoyl transsuccinylase, an enzymatic component of the mitochondrial alpha-ketoglutarate dehydrogenase complex (this enzyme catalyzes the

oxidative decarboxylation of alpha-ketoglutarate to succinyl-CoA in the TCA cycle). This strain is also phosphorylated (YeastGenome)

YLR251W

This yeast stain is a protein required for ethanol metabolism. It is induced by heat shock and located in the inner mitochondrial membrane. It is homologous to MPV17, a mammalian peroxisomal membrane protein (YeastGenome).

Creating the Human ORFeomes with Polymerase Chain Reaction

An orfeome is the complete set of cloned open reading frames (ORFs) in a genome (an ORF is one of three possible ways to read a series of DNA nucleotides). ORFs correspond to the sequences on a gene that code for proteins that a cell needs. Consequently, an understanding of orfeomes is very important in the study of eukaryotic organisms. Creating a large number of copies of each of the five human orfeomes required the utilization of polymerase chain reaction (PCR), a scientific technique that amplifies a piece of DNA, creating thousands of copies of the desired DNA sequence. This technique relies on the use of a thermal cycler, an apparatus that undergoes cycles of repeated heating and cooling of the reaction for melting and enzymatic replication of DNA. This PCR method amplified DNA fragments of approximately 10 kilobases.

Table 1: The Polymerase Chain Reaction Reagents

Components	Description	µL (microliters)
Phusion	DNA polymerase enzyme that catalyzes the replication of the target human orfeome DNA sequence	1.5
HF buffer solution (5x)	an aqueous solution that provides a suitable environment for the enzyme to work most effectively	10
Deoxyribonucleotids dNTPs (10 mM)	DNA building blocks; Adenine, Thymine, Guanine, Cytosine	1
Forward primer	complementary to the 3'on one side of the DNA strand	1
Reverse Primer	complementary to the 3'on one side of the DNA strand	1
DNA template	one of the five human orfeomes	1
ddH_2O	Distilled water: allows the reaction mixture to reach the desired amount	34.5
Total		50

The first PCR setup consisted of eleven reactions (for five human orfeomes, their corresponding selection markers, and a working reaction that served as a positive control). The setup for the eleven reactions required the creation of a "mastermix," a solution consisting of Phusion DNA polymerase, deoxyribonucleotides (dNTPs), buffer solution, and distilled water at an optimal concentration to ensure efficient DNA amplification and minimize the variation between each reaction. The mastermix totaled 564 µl, an amount sufficient for twelve reactions; creating a mastermix for twelve reactions is a safety precaution to ensure that more than enough mastermix exists for each of the eleven reactions.

Each initial PCR was carried out in a reaction volume of 50 µl in reaction tubes that were 0.5 mL. After the eleven setups were complete, the reaction tubes were placed in a thermal cycler. The cycler had two thermal blocks, each consisting of holes where the reaction tubes containing the PCR mixtures could be placed. The cycling parameters consisted of a series of thirty-five cycles, each with the same series of five temperature alternations. In the *initialization step*, the reactions are heated to 95 degrees Celsius for five minutes; this step is often referred to as the "hot start." The purpose of the hot start is to reduce both nonspecific priming and the formation of primer dimers (primers that attached to each other due to complementary base pairing). The next step, the *denaturation step*, consists of heating the reaction to 98 degrees Celsius for thirty seconds. The heat disrupts the hydrogen bonds between complementary base pairs, causing the DNA melting and resulting in single stranded DNA. In the next step, the thirty-second *annealing step*, the reaction temperature was lowered to fifty-five degrees Celsius, approximately five Celsius degrees below the melting temperature of the forward and reverse primers. During this time, the primers anneal to the single-stranded DNA templates. Annealing occurs successfully only when the primer sequences and the template sequences are complementary to one another. Moreover, each primer must be 20-30 nucleotides long to ensure proper specificity. In this step, the polymerase binds to the primer-template junction and begins DNA synthesis. The following step is the *extension step*, in which the temperature is raised to 72 degrees Celsius, the optimal activity temperature of the DNA polymerase. In this step, the DNA polymerase synthesizes a new DNA strand by adding dNTPs that are complementary to the template.

The polymerase reads the DNA template from 3’ to 5’ and adds complementary dNTPs from 5’ to 3’side. This step is repeated thirty-five times. The final step, called the *final hold*, involves cooling the reactions to 4 degrees Celsius for an indefinite period of time. This step is often employed for short-term storage of the reaction.

The thirty thermocycles can be divided into three stages. In the first stage, *exponential amplification*, all the reagents are in abundance and the polymerase is fully functional. Consequently, DNA amplification occurs at 100% efficiency—the amount of replicated DNA doubles with each thermocycle. During the second stage, the *leveling-off stage*, the polymerase loses its activity during the later thermocycles and the concentrations of the reagents begin to decrease, hence limiting the rate and slowing down the reactions. In the final *plateau stage*, the polymerase loses its activity and the reagents are exhausted; no more amplification occurs.

Markers

The purpose of the marker gene is to determine if a DNA sequence has been successfully inserted into the DNA of an organism. Moreover, it protects the organism from a selective agent that would otherwise kill it or hinder its growth. The research project utilized selectable markers with kanamycin resistance; this marker was used for all the orfeomes. Kanamycin is a selective agent used in the isolation of plasmids (bacteria that have uptaken genes) coupled to a gene coding for kanamycin resistance.

pBluescript Vector

The pBluescript vector (a type of cloning vector) is a small piece of DNA that can maintain stability in an organism. It serves as a “backbone” for a foreign DNA fragment during a cloning procedure. It contains 21 unique restriction enzyme recognition sites.

PCR DNA Purification (Using a Microcentrifuge)

The purpose of DNA purification is to extract and purify the DNA of the appropriate human orfeomes and selectable markers and rid them of excess products (ex: primers) For the DNA purification protocol, the “QIAquick protocols” (protocols designed by the Qiagen assaying technology company) and reagents were used.

Bacterial Transformation

Bacterial transformation is the process by which bacteria uptakes DNA and expresses the characteristics of the DNA. For the bacterial transformation, the following reagents were needed: LB+Amp plates, pUC19, NEB competent cells, SOC media, glass beads, and PCR tubes. A series of preliminary steps preceded the bacterial transformation. First, the LB+Amp plates were warmed up to thirty-seven degrees. Two Hybex incubators were heated, one to thirty-seven degrees and the other to forty-two degrees. A tube of competent cells with enough cells for four reactions was thawed by resting on top of ice.

Table 2: Bacterial Transformation Reagents

Reagent	**Purpose**
LB+Amp plates	These plates contain Luria-Bertani, a bacterial nutrient medium that has been combined with agar. These plates have also been treated with the ampicillin antibiotic.
pUC19	A type of plasmid vector in which foreign DNA is introduced for the purpose of cloning
NEB (New England BioLabs) competent cells	E coli strains that are T1 phage resistant and are *endA* deficient, allowing for quality plasmid preparation
SOC media	Nutrient-rich media, usually used for E coli culturing
Glass beads	Added to the LB+Amp plates to thoroughly mix their contents
PCR tubes	These tubes house all the contents involved in bacterial transformation

In isothermal annealing, eight tubes were created—five tubes for each marker/gene/backbone, a tube for a pUC19 positive control, a tube for a Bluescript positive control, and a vector-alone control. 50 microliter aliquots were placed into each tube. After those tubes are incubated at fifty degrees Celsius for one hour, 4 microliters were transferred from each tube into seven individual tubes containing fifty microliters of *E.coli* competent cells. One microliter of pUC19 DNA was added to the positive control. The tubes were then tapped to mix their contents. Next, the tubes were briefly picofuged. Afterwards, the tubes were placed in ice for thirty minutes. Then they were heat shocked at forty-two degrees for thirty seconds before being placed on ice for two minutes. Afterwards, 200 microliters of SOC media was added to the tubes and the tubes were incubated at thirty-seven degrees in a shaker. After one hour in the shaker, the LB+Amp plates were removed from the incubator; approximately ten glass beads were added to

each plate. All 250 microliters of cells in SOC media were added to each LB+Amp plate. The plates containing the beads were shaken horizontally across the lab desk for eight seconds. The glass beads were then removed and the plates were incubated overnight at thirty-seven degrees. The next day, transformed colonies from each plate were picked and incubated in LB broth overnight in preparation for the miniprep the following day.

Minipreparation (Using a Microcentrifuge)

Minipreparation is the small-scale isolation of plasmid DNA from bacteria. The *QIAgen miniprep kit* was used to perform the minipreparation.

Isothermal Annealing

The purpose of isothermal annealing is to connect the human orfeome, marker, and backbone (of the vector/plasmid). After amplifying human orfeomes, markers, and Bluescript vectors in a PCR, a purification protocol was used to extract the DNA. From the purified DNAs, two microliters of human orfeome, two microliters of the corresponding marker, and one microliter of the corresponding Bluescript (for a total of five microliters) was combined with fifteen microliters of a premade substance (the substance was an isothermal assembly mix containing a phusion polymerase, taq DNA ligase, and a thermal label 5'exo nuclease). These twenty liters were then annealed in an incubator for one hours before transforming into *E. coli* competent cells.

Site-Directed Mutagenesis

Site-directed mutagenesis is the process of inducing mutations at specific DNA sequences and studying the effects of those mutations in an organism. In this project, site-directed mutagenesis involved inducing mutations in DNA with the use of primers and studying the effects of the mutated DNA when imbedded in the yeast genome. Primers differing in certain base pairs were used in PCR with our orfeomes in order to create multiple copies of mutated DNA. The primers used were purchased from Integrated DNA Technologies. Each primer was 41 base pairs long, the middle base pair was the desired mutation while the 20 base pairs on either side of the mutation complemented the DNA template to allow for efficient PCR. The DNA templates were provided by the Vamsi Mootha Lab, our collaborating lab. Isothermal annealing joined the mutated DNA to the corresponding Kanamcyin marker and Bluescript vector. Bacterial transformation and minipreparation allowed us to utilize E.Coli genetic machinery to create multiple copies of this newly formed construct. We were able to successfully create a mutated DLST gene and transform the gene into yeast (the strain was later phenotyped by summer intern in the summer of 2012).

Homologous Recombination (Yeast Transformation)

The purpose of homologous recombination is to replace a target allele with an engineered construct without affected other loci in the genome. This is the method by which DNA molecules integrate into yeast chromosomes during yeast transformation. After a specific gene is introduced into the yeast genome via a series of steps, the yeast will express the

characteristics of the added DNA (Yeast Experiments Site). As previously stated, we were able to successfully transform mutated gene DLST into yeast.

Reagents

A yeast transformation requires the following reagents: culture media, icebox, boiled salmon sperm DNA (ssDNA, 10mg/mL), Polyethylene glycol (PEG, 3500, 50% w/v), and Lithium acetate (LiAc, 1M and 0.1M).

Table 3: Yeast Transformation Reagents

Reagent	**Purpose**
YPD Culture Media	Increases yeast transformation efficiency, used for yeast strains exhibiting poor growth
Ice	Maintain integrity of ssDNA
Boiled Salmon Sperm DNA (ssDNA/10mg/mL)	A carrier that increases transformation efficiency
Polyethylene glycol (PEG, 3500, 50% w/v)	Likely dehydrates the membrane, enhance membrane permeability to calcium and other ions, and decrease membrane fluidity
Lithium acetate (LiAc, 1M and 0.1M)	Increases permeability of yeast cell wall

ssDNA

The ssDNA was preheated at a heat block to 100 degrees Celsius. One milliliter of the ssDNA was boiled for five minutes. Afterwards, it was stored at -20 degrees Celsius. At all times during the yeast transformation, the ssDNA was kept on ice.

Table 4: Transformation Reagents Prepared For Six Reactions

Reagent	**uL (microliters)**
PEG 3500 50% w/v; of; of (10 mg/ml); Plasmid DNA + dH_2O (~1microgram DNA) for a total of	1440
LiAc 1.0M	216
boiled ssDNA	60
Plasmid DNA + dH_2O (~1microgram DNA)	444
Total	2160 (or 2.160 mL)

Day 1

The yeast was inoculated into five milliliters of YPD and incubated on a rotary shaker at thirty degrees Celsius and 200rpm. A fifty milliliter flask of YPD was incubated (typically, one flask is incubated for every ten transformations performed.

Day 2

The overnight yeast culture was diluted 1:100 into the fresh incubated YPD media. These dilutions were incubated for five hours at thirty degree Celsius and 200 rpm. When the cells appeared cloudy (with an optical density between 0.3 and 0.5), the cells were poured into fifty milliliter falcon tubes and centrifuged at 3000g for five minutes. In order to wash the cells, twenty five milliliters of sterile water was added and the supernatant was discarded. The cells were washed again in twenty five milliliters of 100mM Lithium acetate. Afterwards, the cells were resuspended in one milliliter of 100mM Lithium acetate. The cell suspension was then transferred to a 1.5 milliliter microcentrifuge tube and centrifuged for thirty seconds at 13,000rpm; the supernatant was discarded. Next, 100mM of lithium acetate was added to complete a final volume of 500 microliters. The

cell suspension was vortexed and fifty microliters of each suspension was aliquoted into 1.5 milliliter microfuge tubes, one for each transformation. The tubes were centrifuged at 13000 rpm for thirty seconds and the supernatant was removed. Then, 360 microliters of the transformation mix was added to each tube and vortexed for thirty seconds to mix adequately. The transformation mix was incubated at thirty degrees Celsius for thirty minutes. Next, it was heat shocked at forty-two degrees Celsius for fifteen minutes. After centrifuging it again at 13000 rpm for thirty seconds and removing the transformation reagents with a pipet, 1.0 milliliters of YPD was pipette into each tube and the pellet was stirred with a vortex. The mixture was incubated at thirty degrees Celsius for five hours. Next, 200 microliter aliquots were plated onto selective media. The plates were incubated at thirty degrees Celsius for four days.

Growth Competition Assay

We developed a competition assay in which two yeast cultures are grown together in equal ratios and under identical environmental conditions and measured over time in order to quickly detect growth rate differences and establish a fitness relationship. Each cell of a 96-well plate was inoculated and the plate was incubated overnight at 30°C. after 24 hours, two different dilutions were prepared: one 100 fold dilution in a deep well plate (which was left to incubate for 24 hours to be read the next day) and a 10 fold dilution in a shallow well plate (to be read by a flow cytometer, a device that allows for simultaneous analysis of multiple properties of thousands of particles per second). In a flow cytometer, a laser light of a single wavelength is directed onto a focused stream of

fluid while several detectors aim that the point where the stream passes through the light. These detectors are able to pick up both scattered and fluorescent light and record information about the physical and chemical structure of each particle. Most importantly, these detectors pick up information about the relative growth rates of the competing yeast strains in each well. Fluorescent protein, targeted at a neutral locus in the yeast strain prior to the flow cytometry experiment, provide a convenient way to differentiate between the two competing yeast strains. In this experiment, we utilized yellow (YFP) and red (mCherry) fluorescent proteins).

ISD11

Clinical Background

Patient 1 weighed 3,170g at birth. He was born by normal vaginal delivery at 38 weeks of gestation. His parents were first cousins. Initial blood tests determined him to have metabolic acidosis with pH 7.14. He had maximum blood lactate of 14.4mmol/L, indicating lactic acidosis. Cerebrospinal fluid (CSF) lactate was 3.9mmol/L. his condition improved and he was transferred back to the referring hospital. However, he displayed poor weight gain despite adequate caloric intake. Over the next few weeks, he slowly improved in feeding and weighed a healthy 3,710g. Today, he is a healthy 20 year old (normal physical examination).

Patient 2 was the double first cousin. His labor was induced at 36 weeks gestation because of poor cardiotocograhy tracings. His birth weight was 2,685g. Within the first 24 hours, she developed respiratory distress and required supplemental oxygen for a total

of five weeks. she was later discharged at six weeks of age but was readmitted two weeks later after displaying a week cry, poor feeding, and weight loss. She had metabolic acidosis and her urine metabolic screen showed significant increase of lactate as well as severe ketosis. Her clinical condition progressively worsened and she died at 12 weeks of age.

MitoExome Sequencing and Variant Prioritization

The MitoExome genes sequenced consisted of the entire mtdNA and 1,381 nuclear-encoded mitochondrial-associated genes. This study was approved by the ethics committee at the Royal Children's Hospital in Melbourne. All samples were obtained with informed consent from families.

Sanger Sequencing

The following were performed: DNA isolation, RNA isolation, cDNA synthesis, inhibition of nonsense mediated decay (NMD) pathway, mRNA splicing, and sequencing of PCR.

Sequenom Genotyping Assay

86 Lebanese controls were genotyped for c.203G>T (the *LYRM4* mutation) as well as the c.93>A COX10 variant (a mutation in the *COX10* gene, a gene that encodes the heme A:farnesyl transferase that converts protoheme to heme O) using a multi-plexed MALDI-TOF mass spectrometry (Sequenom) assay.

Molecular Karyotyping

Molecular karyotyping of DNA samples was performed using the Illumia HuamnCytoSNP-12 array. Automated detection of Long Contiguous Stretch of Homozygosity (LCSH) was performed using the cnPartition v2.1.6 algorithm in the KaryoStudio software. SNP genotypes were created in GenomeStudio software (Illumina) using data from a set 102 intra-run samples. These DNA samples yielded a SNP call rate of 99.5%. Annotation used the NCB136/hg18 human genome assembly.

Construction of Yeast Strains

pISD11R71L (the yeast equivalent of the patient mutation) was constructed with isothermal annealing templated with pISD11 (the wild-type coding sequence). It was amplified with primers containing the p.R71L patient mutation. Template pISD11 was obtained from the Molecular Barcoded Yeast ORF (MoBY-ORF) collection. The pISD11R71L construct was confirmed by sequencing. Yeast strain MSE78Y was constructed using both PCR and homologous recombination into MSB89Y, a derivative of FY4 (a commonly used yeast strain). This deletion was confirmed by colony PCR. Construct pISD11R71L and template pISD11 were transformed into MSE78Y and sporulated in the presence of G418 antibiotic to maintain the plasmid. The Mating type was determined, and three independent yeast transformants were kept (as a way to confirm that the observed phenotype was due to the deletion). Furthermore, MSG30Y strain and the MSG31Y strain were constructed using similar methods.

Yeast Competitive Growth Assays

Strains MSG32-37 were each co-cultured independently with MSG30Y or MSG31Y. After saturation, co-cultures were diluted 1:100 into fresh media and allowed to reach saturation again in order to eliminate history dependent effects. For each growth assay, strains were seeded in quadruplicate in 96-well format. Every morning, each culture was diluted 1:100. Saturated cultures were diluted 1:10 into water and analyzed on a flow cytometer. Competitive growth was performed in four media: synthetic minimal 2% dextrose (SC), synthetic minimal 2% galactose (SGal), synthetic minimal 2% acetate (SAce), and YPD.

Cloning, Expression, Purification and Site-Directed Mutagenesis of Human NFS1 and LYRM4

For the construction of the ISD11-R68L variant (in order to examine the effect of the p.R68L mutation in ISD11), primers were designed by PCR mutagenesis. These primers were cloned into pACYCDuet-1. The purity of the NFS1/ISD11 complex and the presence of ISD11 were confirmed using SDS-PAGE and immunoblotting techniques.

L-cysteine Desulfurase Activity Assays

L-cysteine desulfurase activities of NFS1□1-55/ISD11, NFS1Δ1-55/ISD11-R68L or NFS1Δ1-55 were quantified by the methylene blue method. These reactions were initiated by the addition of L-cysteine (0.1-1.0 mM). The standard curve was recorded

with sodium sulfide. NFS1Δ1-55 concentrations were determined from the absorbance at 420 nm using the extinction coefficient of 10.9 mM^{-1} cm^{-1} for the native enzyme.

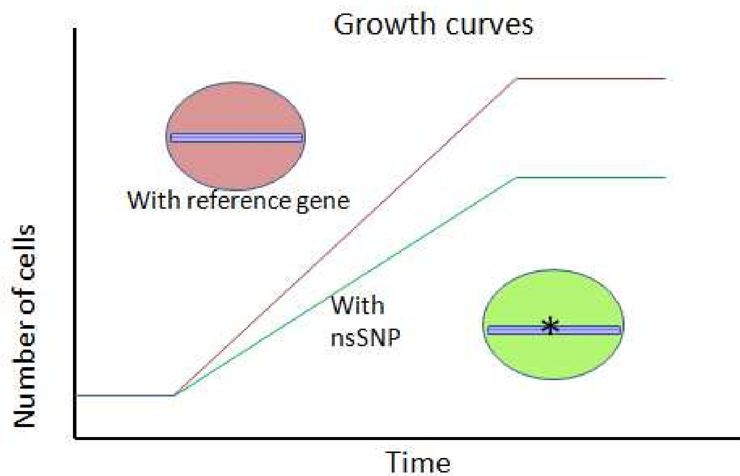
CHAPTER 3

RESULTS

Results: Summers 2011 and 2012

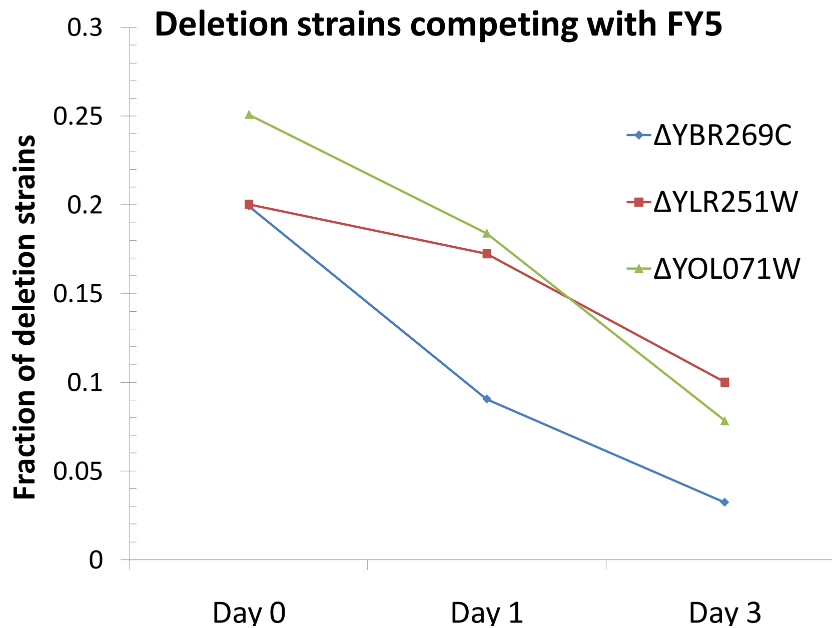
The graph below describes the growth rate of two yeast strains over time: the yeast strain containing a reference nsSNP and the yeast strain without the reference nsSNP. The graph indicates that over time, the yeast strain containing the reference gene had a higher growth rate than the yeast strain without the reference gene.

Figure 1: Growth Curves



The graph below describes three FACS growth competitions in which FY5 was competed against yeast strains with each gene of interest deleted (YBR269C, YLR251W, and YOL071W); the fraction of each plate well comprising deletion strains over a course of three days was plotted against time (days). Significantly, over the course of three days, all three deletion strains grew worse than the wild-type strains.

Figure 2: Deletion Strains Competing with FY5



Results: ISD11 Experiments

Homozygous LYRM4 Identified with MitoExome Sequencing

The MitoExome genes sequenced consisted of the entire mtdNA and 1,381 nuclear-encoded mitochondrial-associated genes. From these genes, 685 single nucleotide variants (SNVs) and short insertions or deletions (indels) were identified and prioritized based on a causative association with an inherited disease. Variants commonly found in public databases (based on an allele frequency cutoff of 0.005) were eliminated, leaving 18 variants. The remaining variants were further classified based on their likelihood to impact protein function (7 variants were further eliminated). When the remaining 11 variants were prioritized for recessive-type inheritance (containing either a homozygous variant or two heterozygous variant), a notable T mutation was found in *LYRM4*, a gene

that encodes the ISD11 protein required for the production of ISC. Since both P1 and P2 had deficiencies in OXPHOS complexes I, II, and III, the likelihood of a *LYRM4* gene mutation being a causative factor remained strong. Furthermore, the *LYRM4* mutation in P1 was further confirmed with Sanger sequencing and was found to be homozygous in cousin P2.

LYRM4 Was Within a Shared Region of Long Contiguous Stretch of Homozygosity

Illumina HumanCytoSNp-12 array (a high-throughput, whole genome scanning kit) identified two regions of Long Contiguous Stretch of Homozygosity (LCSH) in P1 and P2; LCSH's are regions in the genome possessing two identical copies of a gene, one from each parent. These two regions were chr6:4,762,144-11,270,637bp and chr21:28,127,764-29,481,474bp. Although three genes from the chromosome 6 region were identified as encoding mitochondrial proteins (*LYRM4, FARS2, TMEM14C*), *LYRM4* was the only candidate gene identified.

Levels of ISD11 and Proteins with ISCs Were Severely Decreased in Patients' Muscle and Liver

Using SDS-PAGE western blot, ISD11 was not detected in the liver of P1 and P2 or the skeletal muscle of P2. Furthermore, no skeletal muscle biopsy was available from P1 analysis. In P2 muscle, several subunits of complex I, II, and III were undetectable, while protein levels of several complex II, complex III and complex IV subunits decreased.

Table 5: Muscle of Patient 2

	Undetectable Subunits	Subunits with Decreased Protein levels	Percentage of Protein Level Decrease
Complex I	NDUFB8 and NDUFS3	N/A	N/A
Complex II	SDHB	SDHA	16%
Complex III	UQCRFS1	UQCRC2	35%
Complex IV	N/A	COX2 and COX1	25% and 20% respectively

Table 6: Liver of Patient 1 and Patient 2

	Reduced Subunits
Complex I	NDUFB8 and NDUFS3
Complex II	SDHB and SDHA
Complex III	UQCRFS1
Complex IV	COX1 AND COX2

The p.R68L Variant Failed to Fully Complement For an ISD11 in Yeast;
The p.R68L Variant Did Not Affect the Stability of the NFS1/ISd11 Complex

A yeast complementation assay was constructed in order to determine the effect of the p.R68L (missense mutation) in ISD11. Two *S.cerevisiae* strains were constructed with ISD11 promoter and terminator; one strain had the wild-type coding sequence (pISD11) and the other strain had the yeast equivalent of the patient mutation, p.R71L (pISD11R71L). Both pISD11 and pISD11R71L sequences were transformed into heterozygous ISD11 deletion strains. These strains were then allowed to sporulate, producing strains containing one copy of ISD11 on the plasmid. Competitive growth assays between the wild type and mutant strains showed a notable decrease in fitness in all four tested growth media. The impact of p.R68L variant on stability of the

NFS1/ISD11 complex was analyzed. The p.R68L variant in an ISD11 did not alter the stability of the NFS1/ISD11 complex.

Purified NFS1△1-55/ISD11-R68L Complex Had Minimal L-cysteine Desulferase Activity

L-cysteine desulferase activity of purified NFS1△1-55/ISD11-R68L experienced a significant decrease compared to NFS1△1-55/ISD11; it was comparable to the activity of purified NFS1△1-55 protein in the absence of ISD11. The L-cysteine Desulferase activity of the NFS1△1-55/ISD11 complex, NFS1△1-55/ISD11-R68L complex, and Purified NFS1△1-55 were analyzed in the presence and absence of FXN and ISCU, proteins that form a quaternary structure with NFS1/ISD11 complex and doubles L-cysteine desulferase activity.

Table 7: L-cysteine Desulferase Activity

	ISCU only	FXN only	ISCU and FXN
NFS1△1-55/ISD11 complex	Unaffected	Decreased by 1/3	Doubled
NFS1△1-55/ISD11-R68L complex	Slightly decreased activity	Slightly decreased activity	Substantial increase in activity
Purified NFS1△1-55	Unaffected (no activity)	Unaffected (no activity)	Detectable activity

Expression of NFS1△1-55/ISD11-R68L or NFS1△1-55/ISD11 Improved the Growth of iscS Mutant to 30% and 85% Respectively

NFS1 shares 60% of its amino acid sequences with E. coli *IscS*. *E.coli* cells were transformed with different plasmids and grew over a 10 hour period. The data recorded during this period indicates that the growth of the *iscS* mutant improved to 30%

compared to the *iscS* mutant transformed with a plasmid expression *iscS* (when expressing NFS1△1-55/ISD11-R68L). Contrastingly, cells expressing NFS1△1-55/ISD11 restored growth to 85%.

CHAPTER 4

DISCUSSION AND CONCLUSION

Summer 2011 and 2012 Experiments

The data indicates that the deletions significantly impair the fitness in yeast as determined by growth rates. This decline in fitness highlights the contribution these genes make to the normal growth of yeast cells. When the human gene was inserted into the cells, the difference in fitness (compared to the wild type yeast strain) was not significant, suggesting that the human gene at least partially complements the yeast gene. These results support our second hypothesis that stated that the yeast strains with human wild type orfeomes would grow better than the yeast deletion strains. Furthermore, when yeast wild type strains were competed against three different yeast deletion strains, the three deletion strains grew much worse than the wild type strains. These results support our first hypothesis which stated that the yeast wild-type strains would grow better than the yeast deletion strains. These results suggest a promising high-throughput method for these types of investigations.

ISD11 Experiments

Combined OXPHOS deficiency is one of the most common enzyme defects in patients with OXPHOS. However, combined deficiency of complexes I, II, and III is very rare. In this study, mutations in the *LYRM4* gene (the gene that encodes ISD11 protein) were identified as being a cause of deficiency in OXPHOS complexes I, II, III in two cousins from a Lebanese/Syrian consanguineous family. This conclusion is based on the

MitoExome sequencing of 1,000 genes encoding known mitochondrial proteins and corresponding Sanger sequencing. MitoExome sequencing prioritized a homozygous c.203G>T (p.R68L) *LYRM4* mutation as the likely deleterious recessive defect in Patient 1. Independent SNP analyses identified the *LYRM4* gene as being one of three genes encoding mitochondrial proteins in identical regions of the two cousins. The mutations were predicted to cause a missense change which affected a conserved amino acid residue in ISD11 and the ISD11 protein could not be detected in the skeletal muscle of Patient 2. Several subunits of OXPHOS complexes I, II, and III were decreased. The amounts of several iron-sulfur proteins from other biological pathways (mainly aconitase and ferrochelatase were also decreased). Functional studies of expressed wild-type and mutant proteins showed that in yeast isd11 deletions strains, mutant proteins are unable to rescue growth. Furthermore, purified NFS1□1-55/ISD11-R68L complex had minimal L-cysteine desulferase activity.

Though both patients had the same *LYRM4* mutation, they had significantly different clinical phenotypes. P2 notably had a complex IV deficiency in skeletal muscle that P1 did not have. Furthermore, a c.93C>A COX10 variant in a homozygous state was found in P1 but not P2, suggesting that these differences may be caused by two distinct genetic disorders in the family. Furthermore, the c.93C>A COX10 variant was homozygous in P1 and heterozygous in P2.

The difference in clinical outcome between the patients may be dependent of the availability of cysteine. There is a limited availability of sulfur donor cysteine during the neonatal period (due to the weakening of several crucial synthesis pathways).

In yeast, ISD11 plays an important role in ISC (iron-sulfur cluster) biosynthesis in mitochondrial, cystolic, and nuclear proteins. As previously stated, knocking down *LYRM4* in He-La cells affected both Fe-S cluster assembly in mitochondria and the cytosol and the overall iron homeostasis leading to abnormal iron accumulation in cells. Consequently, a defect in ISD11 would expectedly impact many pathways that require ISCs. This contrasts with the more specific effects that a mutation in NUBPL has on OXPHOS complex I (NUBPL has a role in the later stages of the assembly of ISC).

All patients with combined deficiency in complexes I, II, and III had mutations in one of the ISC biogenesis factors. However, this does not suggest that all patients with mutations in ISC biogenesis factors have combined deficiency in complexes I, II, and III (such as NUBPL patients with isolated complex I deficiency).

Conclusion

The results of this experiment suggest that the utilized alleles contain defects that are measurable in yeast and thereby suggest that a similar defect may be the causative defect in humans. Moreover, both experiments validate our approach as a successful way to determine the relevance of certain snSNPs in humans. Although further tests are needed,

the tests performed so far demonstrate a promising high-throughput method for our investigation.

BIBLIOGRAPHY

Adam, Alexander C et al. “The Nfs1 interacting protein Isd11 has an essential role in Fe/S cluster biogenesis in mitochondria.” *The EMBO journal* 25.1 (2006): 174–183.

"ATP5A1." *NCBI*. U.S. National Library of Medicine, n.d. Web. 9 Nov. 2013. <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&q=ATP5A1>.

Bernier, F. P. et al. “Diagnostic Criteria for Respiratory Chain Disorders in Adults and Children.” *Neurology* 59.9 (2002): 1406–1411. Web. 23 May 2013.

"C6orf57 Gene." *GeneCards*. N.p., n.d. Web. 9 Nov. 2013. <http://www.genecards.org/cgi-bin/carddisp.pl?gene=C6orf57>.

Calvo, Sarah E et al. “Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing.” *Science translational medicine* 4.118 (2012): 118ra10.

Dorland. *Dorland’s Illustrated Medical Dictionary*. Elsevier Health Sciences, 2011. Print.

Finsterer, J, C Jarius, and H Eichberger. “Phenotype variability in 130 adult patients with respiratory chain disorders.” *Journal of inherited metabolic disease* 24.5 (2001): 560–576. Print.

Freimer, Nelson, and Chiara Sabatti. “The Use of Pedigree, Sib-pair and Association Studies of Common Diseases for Genetic Mapping and Epidemiology.” *Nature Genetics* 36.10 (2004): 1045–1051. Web. 23 May 2013.

Gasparre, Giuseppe et al. “Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors.” *Proceedings of the National Academy of Sciences of the United States of America* 104.21 (2007): 9001–9006.

"Gene (DLST dihydrolipoamide S-succinyltransferase)." *NCBI*. U.S. National Library of Medicine, n.d. Web. 9 Nov. 2013. <http://www.ncbi.nlm.nih.gov/gene/1743>.

“Genome-Wide Association Studies Fact Sheet.” *National Human Genome Research Institute (NHGRI) – Homepage*. N.p., n.d. Web. 18 Apr. 2013. <http://www.genome.gov/20019523>.

Gusella, James, and Marcy MacDonald. “No Post-genetics Era in Human Disease Research.” *Nature Reviews Genetics* 3.1 (2002): 72–79. Web. 23 May 2013.

"High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method." *Nature*. N.p., n.d. Web. 9 Nov. 2013. <http://www.nature.com/nprot/journal/v2/n1/full/nprot.2007.13.html>.

"How Is Mitochondrial Disease Inherited?." *MitoCanada | Supporting families impacted by Mitochondrial Disease*. N.p., n.d. Web. 23 May 2013. <http://www.mitocanada.org/about-mitochondrial-disease/how-is-mitochondrial-disease-inherited-2/>.

Hurko, Orest. "Drug development for rare mitochondrial disorders." *Neurotherapeutics: the journal of the American Society for Experimental NeuroTherapeutics* 10.2 (2013): 286–306.

Kepler, Ann. "Shortcomings and Promises of GWAS." *Therheumatologist.com*. N.p., n.d. Web. 23 May 2013. <www.the-rheumatologist.org/details/article/1457969/Shortcomings_and_Promises_of_GWAS.html>.

"Journal of Basic MicrobiologyVolume 49, Issue 1, Article first published online: 16 SEP 2008." *Transcriptional and metabolic response in yeast Saccharomyces cerevisiae cells during polyethylene glycol-dependent transformation*. N.p., n.d. Web. 9 Nov. 2013. <http://onlinelibrary.wiley.com/doi/10.1002/jobm.200800123/pdf>.

Leonard, J V, and A H Schapira. "Mitochondrial respiratory chain disorders I: mitochondrial DNA defects." *Lancet* 355.9200 (2000): 299–304. Print.

Lieber, Daniel S. et al. "Atypical Case of Wolfram Syndrome Revealed Through Targeted Exome Sequencing in a Patient with Suspected Mitochondrial Disease." *BMC Medical Genetics* 13.1 (2012): 3. Web. 23 May 2013.

Morava, E et al. "Mitochondrial disease criteria: diagnostic applications in children." *Neurology* 67.10 (2006): 1823–1826.

"MPV17." *MpV17 mitochondrial inner membrane protein*. N.p., n.d. Web. 9 Nov. 2013. <http://ghr.nlm.nih.gov/gene/MPV17>.

"Part B : Yeast Experiments." *Department of Physics*. N.p., n.d. Web. 9 Nov. 2013. <http://www.phys.ksu.edu/gene/bb7.html>.

Pierce, Benjamin. *Genetics: A Conceptual Approach*. Macmillan, 2010. Print.

"SDHAF2." *Succinate dehydrogenase complex assembly factor 2*. N.p., n.d. Web. 9 Nov. 2013. <http://ghr.nlm.nih.gov/gene/SDHAF2>.

Sebastiani, Paola et al. “Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study.” *American journal of hematology* 85.1 (2010): 29–35.

Shi, Yanbo et al. “Human ISD11 is essential for both iron-sulfur cluster assembly and maintenance of normal cellular iron homeostasis.” *Human molecular genetics* 18.16 (2009): 3014–3025.

Shoubridge, Eric A. “Nuclear Genetic Defects of Oxidative Phosphorylation.” *Human Molecular Genetics* 10.20 (2001): 2277–2284. Web. 24 May 2013.

Taylor, Robert W. et al. “Mitochondrial DNA Mutations in Human Colonic Crypt Stem Cells.” *Journal of Clinical Investigation* 112.9 (2003): 1351–1360. Web. 23 May 2013.

Tymoczko, John L., Jeremy M. Berg, and Lubert Stryer. *Biochemistry: A Short Course*. Macmillan, 2011. Print.

“Types of Genetic Diseases.” *Packard Children’s Hospital at Stanford LPCH: Northern California Children’s Hospital* . N.p., n.d. Web. 18 Apr. 2013. <http://www.lpch.org/DiseaseHealthInfo/HealthLibrary/hrpregnant/typegene.html>.

Wong, Lee-Jun C et al. “Current molecular diagnostic algorithm for mitochondrial disorders.” *Molecular genetics and metabolism* 100.2 (2010): 111–117.

"YBL099W." *Yeast Genome*. N.p., n.d. Web. 9 Nov. 2013. <http://www.yeastgenome.org/cgi-bin/locus.fpl?locus=ATP1>.

"YBR269-C." *Yeast Genome*. N.p., n.d. Web. 9 Nov. 2013. <http://www.yeastgenome.org/cgi-bin/locus.fpl?dbid=S000000473>.

"YDR148C." *Yeast Genome*. N.p., n.d. Web. 9 Nov. 2013. <http://www.yeastgenome.org/cgi-bin/locus.fpl?locus=ydr148c>.

"YLR251W." *Yeast Genome*. N.p., n.d. Web. 9 Nov. 2013. <http://www.yeastgenome.org/cgi-bin/locus.fpl?dbid=S000004241>.

"YOL071W." *Yeast Genome*. N.p., n.d. Web. 9 Nov. 2013. <http://www.yeastgenome.org/cgi-bin/locus.fpl?dbid=S000005432>.

Zeharia, Avraham et al. “Acute infantile liver failure due to mutations in the TRMU gene.” *American journal of human genetics* 85.3 (2009): 401–407.