ESTABLISHING NOVEL RELATIONSHIPS BETWEEN GENETIC VARIANTS AND GENE EXPRESSION IN HUMANS

An Undergraduate Research Scholars Thesis

by

ISHITA MANDHAN

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor: Dr. James Cai

May 2013

Major: Computer Engineering (ECE Track)

TABLE OF CONTENTS

Page

ABSTRACT

Establishing Novel Relationships between Genetic Variants and Gene Expression in Humans.
(May 2013)

Ishita Mandhan
Department of Electrical Engineering
Texas A&M University

Research Advisor: Dr. James Cai
Department of Veterinary Integrative Biosciences

We observe phenotypic variations in the human population every day; however, the underlying genetic mechanism behind these variations is difficult to decipher. In this project I investigate the links between phenotypic and genetic variations in humans via three different studies. First, I focus on the impact of the loss-of-function (LoF) mutations on the gene expression. To study the impact of LoF mutations on gene expression, messenger RNA (mRNA) sequences were aligned generating a map of similarities and differences in the RNA seq around the LoF mutation being analyzed. The patterns of RNA seq alignment were studied using the Integrative Genomics Viewer and compared with the list of LoF mutations. The results from the study of LoF mutations confirmed the presence of mutations which impact the gene expression for a particular disease and also reinstated the RNA-seq data obtained. Therefore, the genetic variants present did not necessarily cause the disease in all cases and the gene was not expressed when LoF mutations were present. This RNA editing mechanism protects individuals from the expression of the LoF mutations. Second, I focus on the expression level of mitochondrial DNA (mtDNA)-encoded genes. FPKM (fragments per kilobase of exon per million fragments) was used in the study of mtDNA-encoded genes to measure the relative abundance of transcripts in the RNA seq

data. The study of mtDNA-encoded genes showed the presence of individual variations in the expression levels of the mtDNA and nuclear DNA encoded genes. I also found that the expression of nuclear genes is correlated to the expression of mtDNA-encoded genes.

ACKNOWLEDGEMENTS

I am very grateful to my advisor Dr. James Cai for his trust in me and his guidance throughout the duration of the project. This research would not have been possible without his constant encouragement and advice. I would also like to thank Gang Wang and Ence Yang from the Cai laboratory for their help in the different steps of this study. Last but definitely not the least; I would like to thank my parents for supporting me and my brother, Ishan Mandhan, for inspiring me to always work harder.

CHAPTER I

INTRODUCTION

The building blocks of life, DNA, are made up of sequences of nucleotides which have both coding and non-coding segments, called exons and introns, respectively. In the process of DNA transcription, the DNA is spliced at various sites such that only the exons form an mRNA sequence and the non-coding segments are removed. The mRNA then undergoes the process of translation resulting in amino acid sequences of the protein products. Each cell produces thousands of different mRNA molecules that are translated depending on the requirements of the cell. This thesis concerns two issues in regard to establishing discrete relationships between genetic variants and gene expression in humans.

First, I focus on the impact of the loss-of-function mutations on the gene expression. A genetic variant in the human genome refers to the difference in the genetic sequence of a human. These differences are known to occur numerous times in the human genome and are caused by spontaneous mutations. Variations on the molecular level refer to the differences in the nucleotide base pair sequences in the DNA or RNA that is sequenced and is under study. However, not every single variant explicitly contributes to a disease or abnormal trait being expressed. Sometimes combinations of variants are required to trigger changes in the body metabolisms which may be responsible for causing the disease or trait that is being analyzed. Next-generation sequencing (NGS) techniques have been developed to sequence the DNA at unprecedented speeds producing extremely large data sets of the human genomes for analysis. The technical development allows the large-scale sequencing of human genomes at the

population level, such as the 1000 Genomes Project (The_1000_Genomes_Project_Consortium 2010), became possible. In the meantime, this allows genome-wide association studies (GWAS) to compare thousands of people who have a particular disease to those who do not and try to find patterns in their genetic sequence which occur more frequently in the affected individuals.

Recently large-scale genomic data sets have been used to locate the loss-of-function mutations in multiple healthy individuals (MacArthur, Balasubramanian et al. 2012). Such variants have been regarded as rare and having a high probability of being deleterious, on the basis of their well-established causal roles in severe Mendelian diseases such as cystic fibrosis and Duchenne muscular dystrophy. The loss-of-function mutations, therefore, are of considerable scientific and clinical interest. In the systematic survey of the loss-of-function variants in the human genome by (MacArthur, Balasubramanian et al. 2012), an unexpectedly high number of genes, which have been rendered non-functional by loss-of-function mutations in human protein-coding genes, were found. The results show that human genomes on average contain about 100 genes which are non-functional and about 20 genes which carry inactivating mutations in both gene copies. This suggests that humans are able to tolerate such many loss-of-function mutations. However, the underlying mechanism of this tolerance is not completely known.

Second, I focus on the expression level of mitochondrial DNA (mtDNA)-encoded genes. Eukaryotic cells contain a small chromosome in the cell organelle called the mitochondrion, commonly called the cellular power plant. Each cell contains several mitochondria with its own mitochondrial DNA (mtDNA). Mutations in the mtDNA are common as the mtDNA is more susceptible to oxidative damage. Causing the most damage to cells in tissues with high energy

demands, mitochondrial defects are associated with a number of conditions and diseases, including ageing, cardiovascular disease, diabetes, and cancer. (Andreu, Arbos et al. 1998, Karamanlidis, Bautista-Hernandez et al. 2011, Sondheimer, Glatz et al. 2011)

Very little has been studied about the mtDNA and its transcription in comparison to nuclear DNA (nDNA). None of the large scale RNA-seq studied has focused on the transcription process in mtDNA. The lack of knowledge makes it harder to understand mitochondrial diseases and how they can be controlled and prevented. We study the mitochondrial gene expression in mRNA sequencing results which contain many mitochondrial transcripts in European populations. We hypothesize that the expression levels of mitochondrial DNA encoded genes are about as variable as those of nuclear DNA encoded genes and do not vary as much between individuals. The expression of mtDNA and nDNA encoded genes are compared and the patterns are then studied.

CHAPTER II

METHODS

Impact of loss-of-function mutations on gene expression

The dataset for this project was obtained from the 1000 genomes project (The_1000_Genomes_Project_Consortium 2010) and the sequences were extracted for all individuals having gene expression data. The genotype data of phase 1 of the 1000 genomes project was downloaded from the project official website at http://www.1000genomes.org. The nucleotide sequences were then converted to MATLAB matrices using integer equivalents to ATGC i.e., 1 corresponds to A, 2 corresponds to C, 3 corresponds to G and 4 corresponds to T.

A list of loss-of-function mutations was obtained from the supplementary data of the study of (MacArthur, Balasubramanian et al. 2012) (see Figure 1 for an example). This list contained the genomic locations, reference alleles, alternative alleles, and function effects of loss-of-function variants from existing literature. The indel mutations were filtered out and the main focus was on the single nucleotide variants (SNVs). Figure 1 shows a table of examples of loss-of-function SNVs extracted from the supplementary table of (MacArthur, Balasubramanian et al. 2012). These loss-of-function mutations were found in 1000 Genomes samples and are known to be associated with human complex diseases.

chr	pos	dbSNP	ref	alt	type	gene	phenotype	hets	homs	notes
2	162832842	rs35732034	C	T	splice	*IFIH1*	protection from type 1 diabetes	2 CEU	0	
2	162844751	rs35337543	C	G	splice	*IFIH1*	protection from type 1 diabetes	1 CEU	0	
3	38323747	rs753331	A	C	splice	*SLC22A14*	confirmed effects on mRNA splicing	22 CHB+JPT	6 CHB+JPT	
4	70933511	rs17147990	T	A	stop	*HTN3*	truncated histatin protein	19 YRI	2 YRI	
4	154844848	rs62323857	C	T	stop	*TLR2*	decrease in TLR2 protein function	1 JPT	0	
10	96530400	rs4986893	G	A	stop	*CYP2C19*	altered drug metabolism	5 JPT	1 JPT	
11	66084671	rs1815739	C	T	stop	*ACTN3*	altered muscle function	75 all	31 all	LoF in reference
17	36804625	.	C	T	splice	*KRT31*	truncated (functional) keratin	2 CEU	0	
19	53898835	rs1799761	AC	A	del	*FUT2*	non-secretion of ABO/Lewis antigens	2 YRI	0	
19	56226942	rs3745540	A	G	splice	*KLK12*	loss of protease activity	61 all	63 all	

[a] Approximate coordinates provided for large deletions

Figure 1. Examples of loss-of-function SNVs. Data source:(MacArthur, Balasubramanian et al. 2012).

Further analysis of the RNA-seq was carried forward using the various toolboxes developed by Dr. Cai (Cai, Smith et al. 2005, Cai, Smith et al. 2006, Cai 2008). The mRNA sequences were aligned for a particular range of base pairs in a chromosome obtained from the loss-of-function mutation data generating a map of similarities and differences in the RNA seq around the loss-of-function mutation being analyzed. Original RNA-seq reads produced by Montgomery, Sammeth et al. (2010) and were downloaded from NCBI Gene Expression Omnibus (GEO) under accession numbers GSE19480 and GSE25030 for CEU samples (Montgomery, Sammeth et al. 2010).

One explanation for the existence of loss-of-function mutations in healthy individuals might be RNA editing (Gott and Emeson 2000), a highly regulated molecular process in which the RNA sequence is chemically and enzymatically altered. One way to identify RNA editing events is through rigorous sequence data analyses to reliably identify RNA-DNA differences (RDDs)(Kleinman and Majewski 2012, Lin, Piskol et al. 2012, Pickrell, Gilad et al. 2012). We

systematically examined the RDDs at the site of known loss-of-function mutations in multiple 1000 genomes samples. We used a series of Matlab-based software toolboxes (Cai, Smith et al. 2005, Cai, Smith et al. 2006, Cai 2008) and other existing software tools, such as samtools (Li, Handsaker et al. 2009) and integrative genomics viewer (Robinson, Thorvaldsdottir et al. 2011), to conduct the analyses and produce visual results of the RNA-seq read alignments of different individuals at the same SNP in a particular chromosome.

Expression level of mtDNA encoded genes

RNA-seq data produced by Montgomery, Sammeth et al. (2010) were downloaded from NCBI Gene Expression Omnibus (GEO) under accession numbers GSE19480 and GSE25030 for CEU samples. Mitochondria and mitoplast RNAseq data produced by Mercer, et al. were downloaded from GEO under accession number GSE30772. For each data set, the Sequence Read Archives (SRA) files were downloaded and population specific reference genomes were obtained. The human gene annotation—Gencode v12 was downloaded from the website of the GENCODE project (http://www.gencodegenes.org/). Because large amount of RNA-seq reads were mapped onto mitochondrial genome, it was important to have accurate mitochondrial reference genomes.

FPKM (fragments per kilobase of exon per million fragments) was used in this study to measure the relative abundance of transcripts. To estimate FPKM, short reads were mapped to corresponding population-specific reference genome. The representative housekeeping genes, AATF, ADAR, BSG, E2F4, MGAT1, NDUFA7, PGD, RBM8A, SOD1, TACC1, TALDO1, UBE2D2, and YWHAB, were randomly selected from the list of housekeeping genes obtained from the study of Eisenberg and Levanon.

The weighted gene correlation network analysis (WGCNA) is a data reduction method that groups genes into modules which are clusters of densely interconnected genes in an unsupervised manner based on self-organizing properties of complex systems. R package WGCNA was used to create nuclear gene co-expression modules and estimate the correlation between modules and mitochondrial expression. Modules were identified among genes whose expression is highly correlated in CEU populations. We used the average expression level of mitochondrial genes in each sample as the trait. This allowed us to identify the modules highly correlated with mitochondrial gene expression. We then examined gene functions for the top three significant modules with genes positively correlated with mitochondrial genes and the top three significant modules with negatively correlated genes. For each module, we computed enrichment scores for the GO biological process and molecular function terms. The Gene Ontology (GO) analysis was conducted using DAVID.

CHAPTER III

RESULTS

The results intensified the idea of an RNA editing mechanism that functions at the transcriptional level to eliminate the deleterious effects caused by the loss-of-function mutations. Transcription is the process of formation of a RNA strand complementary to the DNA copy. The loss-of-function mutations that can be found at the DNA level have a parallel gene code of the mRNA which shows similar patterns. If the transcription of DNA sequence into mRNA occurs without any editing, we should be able to identify the same nucleotides of loss-of-function mutations in the mRNA transcripts. However this is not the case in the data that was analyzed in the study. The results of several different sites of mutation showed the possibility of transcription occurring with an RNA editing mechanism.

Figure 2 shows aligned mRNA reads around the site of the loss of function mutation for four individuals with ids ERR009096, ERR009097, ERR009099, and ERR009102 respectively. Thymine at the 31124849 base pair position on chromosome 6 of the CCHCR1 gene indicates the presence of mutation unless cytosine is present predominantly in turn covering up for the very little/no thymine. The CCHCR1 gene has been linked to the psoriasis phenotype in humans (Gandhi, Buttar et al. 2011). Due to the function of the compensatory mechanism for eliminating the Loss-of-function mutations, the mutational profile of mRNA transcripts at the site of the Loss-of-function mutation was different from that of the genomic DNA.

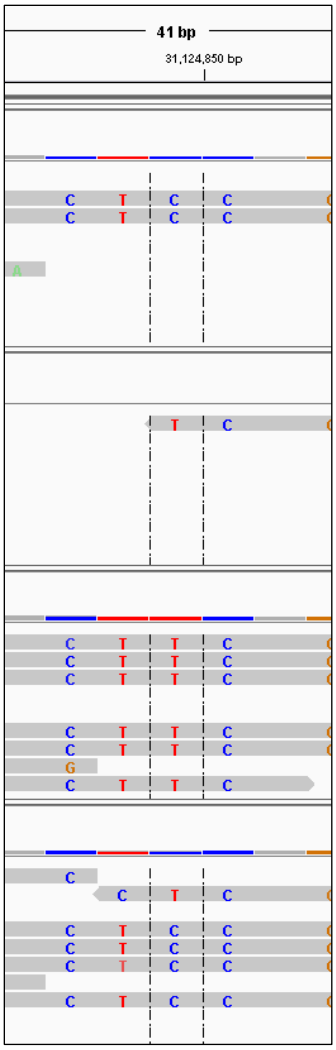

Figure 2. Nucleotide sequences of four individuls at the 31124849 base pair position on chromosome 6 of the CCHCR1 gene.

Aligned mRNA reads around the site of the loss of function mutation, i.e., at the 32411035 base pair position, for four individuals with ids ERR009096, ERR009097, ERR009099, and ERR009102 respectively can be seen in figure 3. This base pair position on chromosome 6 of the human HLA-DRA gene is essential for immune responsiveness to foreign antigens(Matsushima,

Itoh-Lindstrom et al. 1993). The absence of any nucleotide at the site of mutation denotes the presence of a nucleotide base as expected based off of the genomic DNA. The presence of cytosine in individuals 2 and 3 implies that these individuals show the mutation at the loss-of-function mutation sites but remain phenotypically unaffected.

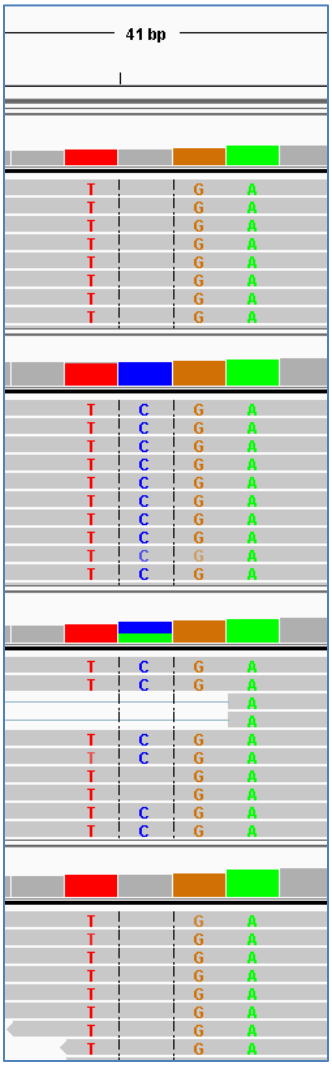

Figure 3. Nucleotide sequences of four individuls at the 32411035 base pair position on chromosome 6 of the HLA-DRA gene.

The gene expression level for 60 CEU individuals was compared for both mitochondrial and nuclear DNA encoded genes. The expression levels for mitochondrial DNA encoded genes showed greater variability than the nuclear DNA encoded genes. This marked difference can be seen by observing the closer quartile ranges for nuclear DNA encoded genes on graph b compared to the more distant quartile ranges for mitochondrial DNA encoded genes on graph a. Also, the between-individual variance in expression level in mt-DNA encoded genes is larger than nuclear DNA encoded genes. The expression of all 13 of the mt-DNA encoded housekeeping genes examined is positively correlated to each other. This could be attributed to the fact that most of them encode for parts of the electron transport chain and are therefore closely related.

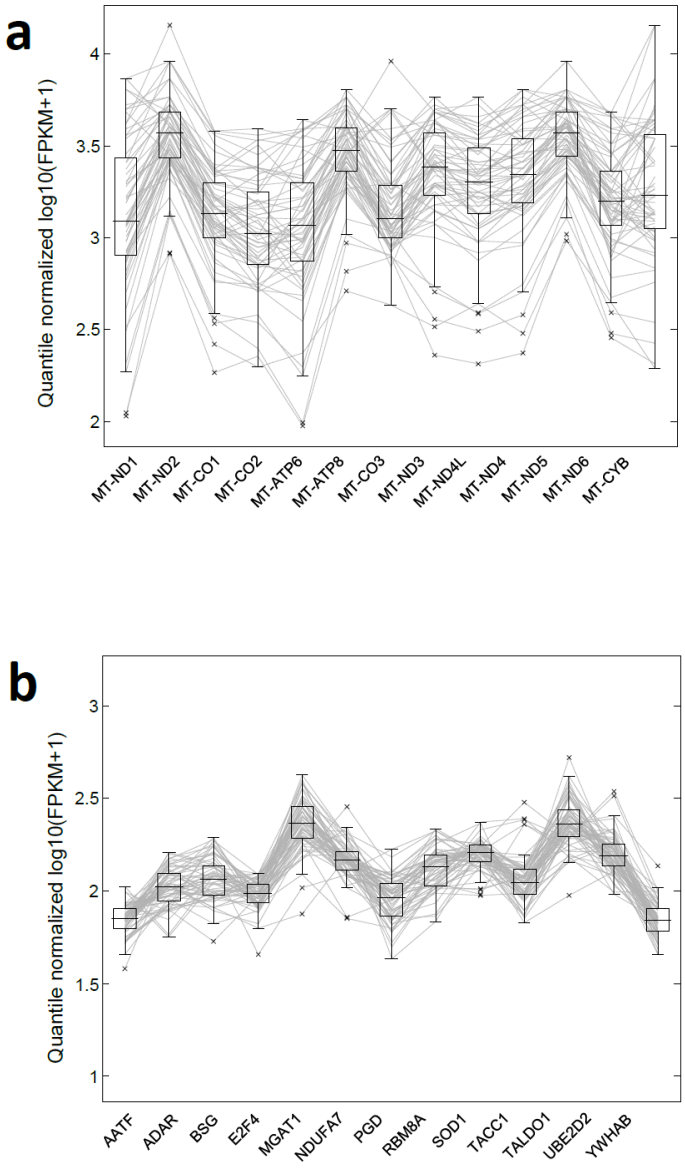

Figure 4. Between-individual variation of mitochondrial gene expression and normalized gene expression levels in 60 CEU individuals for mt-DNA encoded genes (a) and nuclear DNA encoded genes (b).

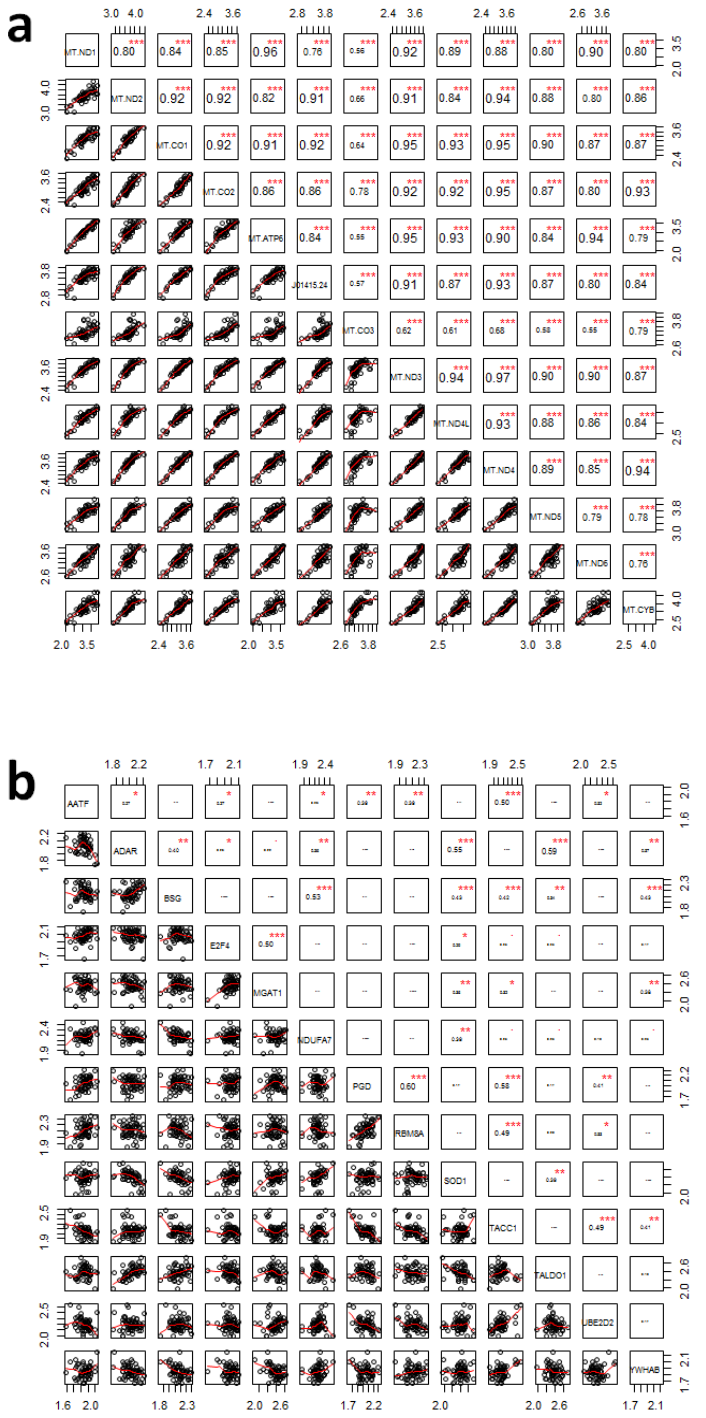

Figure 5. Pairwise correlations of expression level between mt-DNA encoded genes (a) and nuclear DNA encoded genes (b).

CHAPTER IV

DISCUSSION

Inactivation of a gene can have several different phenotypic effects, like inactivation of the gene without clinical impact(MacArthur, Balasubramanian et al. 2012). One possible reason behind such loss-of-function tolerant genes is the presence of an RNA editing mechanism that enables the site of mutation to be overlooked during transcirption inturn supressing the phenotypic expression of the mutation. Our results provide the direct evidence in humans that RNA-editing may play a role in attenuating the detrimental impact on gene transcription caused by the loss of function mutations. Nevertheless, there are several severe recessive genes which are caused by mutations. Further sequencing studies that analyze the impact of these mutations can be used to improvise on clinical treatments.

Despite tehcnical advancements like the spindle transfer technique(Tachibana, Amato et al. 2013), our understanding of the generic regulation of trasncription of mtDNA encoded genes is still fairly low. Some diseases that are caused by alterations in the mitochondrial DNA include Alzheimer's disease (Chandrasekaran, Giordano et al. 1994, Chandrasekaran, Hatanpaa et al. 1997) and nephrotic syndrome (Holthofer, Kretzler et al. 1999). It is essential to learn more about mitchondrial diseases to prevent and treat them. Our lack of knowledge of the trasncriptional process in the mitochondrial DNA should not be in the way of how these set of diseases are dealt with.

We studied the expression levels of 13 mitochondrial housekeeping genes and compared them with nuclear DNA encoded genes. We also compared the expression levels of mitochondrial genes between 60 individuals in the European population being analyzed, which has been neglected by a large majority of previous studies. Using existing data sets, we revealed the population-level expression variability of mtDNA-encoded genes in European populations. Next we tried to understand the correlation in expression levels between mitochondrial and nuclear genes. There is a clear need to integrate information from both mitochondrial and nuclear genetic systems by describing genetic and expression profiles of both genomes(Pesole, Allen et al. 2012). We found that the expressions of many nuclear genes are either positively or negatively associated with the expression of at least one of the mtDNA-encoded genes in CEU populations.

Mitochondria are not only important for the energetic status of the cell, but also the fatal organelles deciding about cellular life and death - apoptosis and mitochondria relationship is known in lower eukaryotic organisms like yeast(Eisenberg, Buttner et al. 2007). Previous comparative analysis across species also suggested that many pathways connected to the respiratory chain are still undiscovered(Vafai and Mootha 2012). Thus, the co-expression we detected may have helped identify these new pathways.

CHAPTER V

CONCLUSIONS

The analysis of loss-of-function variants in 43 healthy individuals of the CEU population showed that most individuals do contain several mutations in heterozygous state which do not necessarily impact us. These variants are likely to be in non-essential genes and so they do not affect the normal functioning and do not create any sort of phenotypic variations. In the case when homozygous state of a LoF mutation indeed happens, RNA-editing is found to play a role to attenuate the negative impact of the mutations to normal gene transcription.

By studying the expression levels of mitochondrial and DNA encoded genes in 60 CEU individuals we have realized the presence of a significant amount of variation in the expression levels across individuals in mtDNA encoded genes. We also learned that there was a similar trend of more varied expression levels for mitochondrial genes when compared to nuclear genes which showed lesser variability. One possible explanation for the variability is that mitochondrial DNA encoded genes are mainly responsible for housekeeping genes and so vary because of the different body metabolisms that need to be regulated at all steps. This could also explain the overall general trend of gene expression of the 60 individuals since we all follow basic housekeeping processes. Our analysis establishes the null expectation of mtDNA gene expression variability in human populations. Future studies of the genetic basis of gene expression of mitochondrial DNA encoded genes will be able to explain the mechanisms behind such variability.

REFERNCES

1. Andreu, A. L., M. A. Arbos, A. Perez-Martos, M. J. Lopez-Perez, J. Asin, N. Lopez, J. Montoya and S. Schwartz (1998). "Reduced mitochondrial DNA transcription in senescent rat heart." Biochem Biophys Res Commun **252**(3): 577-581.

2. Cai, J. J. (2008). "PGEToolbox: A Matlab toolbox for population genetics and evolution." J Hered **99**(4): 438-440.

3. Cai, J. J., D. K. Smith, X. Xia and K. Y. Yuen (2005). "MBEToolbox: a MATLAB toolbox for sequence data analysis in molecular biology and evolution." BMC Bioinformatics **6**: 64.

4. Cai, J. J., D. K. Smith, X. Xia and K. Y. Yuen (2006). "MBEToolbox 2.0: an enhanced version of a MATLAB toolbox for molecular biology and evolution." Evol Bioinform Online **2**: 179-182.

5. Chandrasekaran, K., T. Giordano, D. R. Brady, J. Stoll, L. J. Martin and S. I. Rapoport (1994). "Impairment in mitochondrial cytochrome oxidase gene expression in Alzheimer disease." Brain Res Mol Brain Res **24**(1-4): 336-340.

6. Chandrasekaran, K., K. Hatanpaa, S. I. Rapoport and D. R. Brady (1997). "Decreased expression of nuclear and mitochondrial DNA-encoded genes of oxidative phosphorylation in association neocortex in Alzheimer disease." Brain Res Mol Brain Res **44**(1): 99-104.

7. Eisenberg, T., S. Buttner, G. Kroemer and F. Madeo (2007). "The mitochondrial pathway in yeast apoptosis." Apoptosis **12**(5): 1011-1023.

8. Gandhi, G., B. S. Buttar, L. Albert, Q. Hasan and R. K. Aggarwal (2011). "Psoriasis-associated genetic polymorphism in North Indian population in the CCHCR1 gene and in a genomic segment flanking the HLA-C region." Dis Markers **31**(6): 361-370.

9. Gott, J. M. and R. B. Emeson (2000). "Functions and mechanisms of RNA editing." Annu Rev Genet **34**: 499-531.

10. Holthofer, H., M. Kretzler, A. Haltia, M. L. Solin, J. W. Taanman, H. Schagger, W. Kriz, D. Kerjaschki and D. Schlondorff (1999). "Altered gene expression and functions of mitochondria in human nephrotic syndrome." FASEB J **13**(3): 523-532.

11. Karamanlidis, G., V. Bautista-Hernandez, F. Fynn-Thompson, P. Del Nido and R. Tian (2011). "Impaired mitochondrial biogenesis precedes heart failure in right ventricular hypertrophy in congenital heart disease." Circ Heart Fail **4**(6): 707-713.

12. Kleinman, C. L. and J. Majewski (2012). "Comment on "Widespread RNA and DNA sequence differences in the human transcriptome"." Science **335**(6074): 1302; author reply 1302.

13. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

14. Lin, W., R. Piskol, M. H. Tan and J. B. Li (2012). "Comment on "Widespread RNA and DNA sequence differences in the human transcriptome"." Science **335**(6074): 1302; author reply 1302.

15. MacArthur, D. G., S. Balasubramanian, A. Frankish, N. Huang, J. Morris, K. Walter, L. Jostins, L. Habegger, J. K. Pickrell, S. B. Montgomery, C. A. Albers, Z. D. Zhang, D. F. Conrad, G. Lunter, H. Zheng, Q. Ayub, M. A. DePristo, E. Banks, M. Hu, R. E. Handsaker, J. A. Rosenfeld, M. Fromer, M. Jin, X. J. Mu, E. Khurana, K. Ye, M. Kay, G. I. Saunders, M. M. Suner, T. Hunt, I. H. Barnes, C. Amid, D. R. Carvalho-Silva, A. H. Bignell, C. Snow, B. Yngvadottir, S. Bumpstead, D. N. Cooper, Y. Xue, I. G. Romero, C. Genomes Project, J. Wang, Y. Li, R. A. Gibbs, S. A. McCarroll, E. T. Dermitzakis, J. K. Pritchard, J. C. Barrett, J. Harrow, M. E. Hurles, M. B. Gerstein and C. Tyler-Smith (2012). "A systematic survey of loss-of-function variants in human protein-coding genes." Science **335**(6070): 823-828.

16. Matsushima, G. K., Y. Itoh-Lindstrom and J. P. Ting (1993). "Activation of the HLA-DRA gene in primary human T lymphocytes: novel usage of TATA and the X and Y promoter elements." Mol Cell Biol **13**(8): 5132.

17. Montgomery, S. B., M. Sammeth, M. Gutierrez-Arcelus, R. P. Lach, C. Ingle, J. Nisbett, R. Guigo and E. T. Dermitzakis (2010). "Transcriptome genetics using second generation sequencing in a Caucasian population." Nature **464**(7289): 773-777.

18. Pesole, G., J. F. Allen, N. Lane, W. Martin, D. M. Rand, G. Schatz and C. Saccone (2012). "The neglected genome." EMBO Rep **13**(6): 473-474.

19. Pickrell, J. K., Y. Gilad and J. K. Pritchard (2012). "Comment on "Widespread RNA and DNA sequence differences in the human transcriptome"." Science **335**(6074): 1302; author reply 1302.

20. Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

21. Sondheimer, N., C. E. Glatz, J. E. Tirone, M. A. Deardorff, A. M. Krieger and H. Hakonarson (2011). "Neutral mitochondrial heteroplasmy and the influence of aging." Hum Mol Genet **20**(8): 1653-1659.

22. Tachibana, M., P. Amato, M. Sparman, J. Woodward, D. M. Sanchis, H. Ma, N. M. Gutierrez, R. Tippner-Hedges, E. Kang, H. S. Lee, C. Ramsey, K. Masterson, D. Battaglia, D. Lee, D. Wu, J. Jensen, P. Patton, S. Gokhale, R. Stouffer and S. Mitalipov (2013). "Towards germline gene therapy of inherited mitochondrial diseases." Nature **493**(7434): 627-631.

23. Vafai, S. B. and V. K. Mootha (2012). "Mitochondrial disorders as windows into an ancient organelle." Nature **491**(7424): 374-383.