A Systems Biology Approach to Model Skeletal Muscle Response To Aerobic Exercise Training

by Mamta R. Giri

B.E. in Bio-Medical Engineering, April 2010, University of Mumbai, India

A Thesis submitted to

The Faculty of
Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Master of Science

May 20, 2012

Thesis directed by

Eric P. Hoffman
Professor of Pediatrics

Acknowledgements

Fascination for expanding my knowledge in Genomics and Bioinformatics, I accepted the Masters Program in Biochemistry and Bioinformatics at The George Washington University in Fall, 2010. The last two years as a Graduate student at George Washington University gave me tremendous opportunity to grow and meet people that helped me work towards my goals. I have many people to thank for their mentorship and friendship during these two years.

First and foremost, I want to thank my family for their continuous love and support and giving me an opportunity to pursue my higher education here in the U.S.

While at George Washington University, I want to thank Dr. Jack Vanderhoek for his help and advice throughout, right from the very first day I came to GWU. I also want to thank Dr. Patricia Berg for giving me an opportunity to work in her laboratory and learn all the molecular techniques that would help me a long way. I would also like to thank Dr. Valerie Hu for being so kind and for agreeing to be my co-Advisor and reviewing my thesis.

I consider myself really fortunate that I got a chance to work with many great researchers at The Research Center for Genetic Medicine. I still remember the first day I came to CNMC when Dr. Eric Hoffman gave me a tour of the research space here. I was thrilled to see the Sequencer and wished I could master using this huge machine. During my time at CNMC,

I would like to thank Marie Pichaske, Dr. Joe Devaney and his team for their help and for making me so comfortable at Children's. I am also grateful to Dr. David Rowlands for his time, suggestions and help throughout my thesis work.

Finally, I want to thank my advisor Dr. Eric Hoffman for his amazing mentorship. I had known about Dr. Hoffman of how great of a scientist he is, but never thought I would be fortunate enough to work under his guidance. I truly appreciate the time he spent in mentoring me despite his really busy schedule. This thesis was only possible because of his support and encouragement.

Abstract of Thesis

A Systems Biology Approach to Model Skeletal Muscle Response to Aerobic Exercise Training

Type 2 diabetes is a condition with poorly regulated glucose metabolism. The incidence of Type 2 diabetes has increased tremendously in the past few decades in most countries and is highly prevalent in certain ethnic group particularly. The rise in incidence of Type 2 diabetes is connected to the 'Western life style', with calorie rich food, lack of exercise, and obesity. Type 2 diabetics show poor quality of life with higher risk of all the obesity-associated complications such as Stroke, Diabetes, Hypertension, Cardiovascular Disorders and even Cancer.

Efforts to reduce Type 2 diabetes, or prevent the onset of the disease, typically focus on diet, activity, and pharmacological interventions. Physical activity has been an effective approach in controlling and preventing or even reversing Type 2 diabetes. Physical activity leads to several beneficial structural and functional adaptations which are mediated through different processes and pathways in the skeletal muscle. In order to understand the association of physical inactivity with these complications, it is important to know the genes that are regulated by exercise which lead to improvements in these conditions. Moreover, these changes might be epigenetically regulated, which makes it crucial for therapeutic applications. Thus, it is important to integrate different levels of data, including both mRNA and DNA level regulation, to better understand these adaptive changes.

In this study, our goal was to define the adaptive changes in the skeletal muscle of patients with Type 2 diabetes after 16 weeks of aerobic exercise training. The patient group studied

was Polynesian; indigenous [Maori] and more recent residents of Pacific Island migrants to New Zealand that showed grade III obesity and Type 2 Diabetes Mellitus. We describe bioinformatics and data analysis by integrating mRNA, microRNA and epigenetic data from these subjects in muscle biopsies taken before and after the intervention.

The research was a pilot study to model the skeletal muscle response in T2DM subjects due to 16 weeks of aerobic exercise training. The approach used an integrated data to represent the molecular mechanism underlying this intervention and study the DNA (due to methylation) and RNA (due to miRNA) level of regulation. The networks generated showed an improvement in muscular development, endurance and carbohydrate metabolism following the intervention. We would further integrate the proteomics data into these networks using Bayesian approaches for a more probabilistic model of these adaptations followed by validation using methods like RT-PCR and/or functional assays. This would give an insight into the beneficial adaptive changes due to aerobic exercise training in the diabetic skeletal muscle.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Background:

Type 2 diabetes Mellitus (T2DM) is a metabolic disorder, associated with increased glucose level in the blood. The incidence of Type 2 diabetes has greatly increased in the Western countries due to caloric rich diet, lack of physical activity and obesity. According to the CDC [Centers for Disease Control and Prevention] report 2011, diabetes affects around 25.8 million people in the U.S. alone. The rate of diabetes is increasing tremendously causing health issues and increasing health care costs. Moreover, diabetes creates major long term vascular complications [Engelgau et al. 2004] and other microvascular complications which include diabetic retinopathy, peripheral neuropathy and diabetic nephropathy which ultimately may result in vision loss, amputations and end-stage renal disease. Exercise training, both resistance and aerobic are known to improve quality of life and has been considered as one of the best remedies in controlling and preventing or even reversing Type 2 diabetes and associated disorders. Skeletal muscle is an extremely flexible organ and adapts immediately to changes. With the advent of microarray technology, studies have been conducted to identify genes that are involved in the exercise induced adaptations in skeletal muscle [Tehran-Garcia et al. 2005]. However, the mechanisms that lead to these exercise induced effects are not very well understood. Physical activity plays an important role in the management of T2DM, improving glucose transport, insulin sensitivity and reducing the risk of cardiovascular disorders. Exercise is known to improve insulin sensitivity and cardiovascular fitness [Duscha et al. 2005, Church et al. 2007], reduces blood pressure [Halbert et al. 1997, Kelly et al. 2001], improves dyslipidemia [Durstine et al. 1994, Kraus et al. 2002], and metabolic syndrome [Johnson et al. 2007, Slentz et al. 2009]. In addition to functional improvements associated with motor recruitment and strength, training can increase skeletal-muscle glucose disposal [increased GLUT4 content], vascularization, mitochondrial volume,

fatty-acid turnover and oxidative capacity [Rowlands et al. 2009, Galgani et al. 2008]; these structural and metabolic adaptations increase the tissue metabolic capacity, which in time may contribute to euglycemia, and improved tissue insulin sensitivity. Exercise leads to these changes through the activation of MAPK signaling [Long et al. 2004, Zeirath et al. 2002] calcium activated signaling, regulation of GLUT4 expression [Lund et al. 22 1995], and regulation of PPARs [Russell et al. 2003]. It has also been shown that endurance exercise has greater potential of treating T2DM than resistance exercise [Cauza et al. 2005]. To better understand the effects of sedentary lifestyle and its association with these complications, it is important to define the molecular remodeling that occurs in Type 2 diabetic muscle following an exercise intervention. Thus, this study is conducted in an ethnic group with high risk for T2DM and grade III obesity to uncover the basis for molecular remodeling after aerobic exercise training.

Epigenomic Regulation:

There is increasing evidence that prolonged hyperglycemia of T2DM can induce epigenetic changes to the chromatin structure via regulation of various transcription factors and signaling pathways [Villeneuve et al. 2010]. In addition to hyperglycemia, T2DM is associated with a range of environmental factors [e.g. obesity, nutrition and lifestyle], where each factor could in itself induce epigenetic changes to the chromatin structure altering gene expression patterns. Clinical studies have shown diabetic complications even after normalization of blood glucose, indicating metabolic memory of prior glycaemic state [Villeneuve et al. 2010]. Furthermore, the T2DM diabetic pattern of histone acetylation and methylation appears to be reversible [Ling et al. 2009], which makes it important for therapeutic applications. There is also evidence that miRNA can have impact on skeletal muscle gene expression [Safdar et al. 2009, Neilsen et al. 2010]. Moreover, the fact that many miRNAs play an important role in the pathophysiology of T2DM make epigenetic approaches to increase our understanding of skeletal-muscle function and adaptation in T2DM useful [Fernandez-Valverde et al. 2011]. Therefore, our proposed study to

compare epigenetic mechanisms between the sedentary state and after 16 wks of AER training is likely to yield a better understanding of the epigenetic changes involved in the pathophysiology of T2DM.

Here, we will explore the epigenetic regulation of the transcriptomic response at the RNA level via microRNA [miRNA] expression analysis and at the DNA level via DNA methylation analysis. The overall aim of this study is to reveal new information on the systems biology of tissue response to AER training in the unstudied grade III obese population.

Data Integration and Systems Biology model:

The complex interplay of molecular pathways that are potentially involved in regulating muscle functionality, makes systems biology approaches a desirable option. Given the lack of knowledge about the exercise induced muscle remodeling, we, here use a systems biology model to gain an insight into the pathways/processes and mRNA and DNA level regulation leading to these beneficial adaptations in the skeletal muscle. Only one level of data is not sufficient to fully explain a particular mechanism in the complex biological system. To achieve this we use an integrated data analysis approach using miRNA, mRNA and epigenomic data to understand a broader picture of these exercise induced changes. In order to better understand the mechanism of cancer, an integrated approach incorporating miRNA and epigenetic regulation of genes that lead to disease has been used before [Zhu et al. 2011, Radpour et al. 2011]. There are studies that have tried to unveil the miRNA induced regulation of the transcriptome in human skeletal muscle in response to exercise training [Safdar et al. 2009, Keller et al. 2010], but this is the first report to integrate gene expression, methylation and miRNA expression into a model to understand the aerobic exercise induced adaptations in diabetic skeletal muscle.

Array Platforms and Bioinformatics Analysis:

In our analysis, we have used two widely used one color array platforms for microarray studies, Affymetrix and Illumina. We used Affymetrix GeneChip® [Affymetrix, SantaClara, CA] for microRNA profiling, Illumina® Gene Expression BeadChip Array technology [Illumina, Inc., San Diego, CA] for mRNA and Illumina Infinium HumanMethylation450 BeadChip for methylation profiling. The bioinformatics analysis approach used was specific to the platform. Considering the fact that there is no "best method" for bioinformatics analysis and that the type of analysis depends on the nature of data in hand, we tried to optimize every step of analysis to improve the signal/noise ratios and avoid false positives. In the case of the Affymetrix GeneChip, the array has 11 probe-sets per gene [perfect match/mismatch] and the expression value is the weighted average of 11 probes. Generating these values is a crucial step and we chose an algorithm that could provide higher reproducibility without loss of accuracy and a better differential sensitivity for low expressors. For Affymetrix, background correction is particularly important, which then depends on the sample to sample variability and can be set to MM (Mismatch probes) as background, global uniform background or GC content-based background. It is preferable to use PM (Perfect match probes) only for a homogenous sample group. In the Illumina BeadArray analysis, oligonucleotides are attached to microbeads which are then put onto microarrays using a random self-assembly mechanism and yields 30 copies of the same oligonucleotide which is used as internal technical replicate. The measurement of each probe type is the weighted average of the replicate measures. This higher sample call rates gives less likelihood for false positives and negatives. Illumina provides various internal controls and all the preliminary data Quality Control metrics are easily done in the GenomeStudio before any statistical analysis. This step eliminates any poor performing samples from being included in further analysis.

Overall, we selected the data pre-processing method and statistical model best suited to the particular platform used. Another purpose was to develop a method for data integration and build

networks that can define the integrated effect of the changes in these different levels of data after the training intervention.

Significance of the study:

This study is the first of its kind integrating different levels of data; mRNA, miRNA and DNA methylation in order to model the skeletal muscle response to aerobic exercise training in a cohort of South Pacific Islanders with type III obesity and T2DM.

Chapter 2 Methods

The study cohort:

The cohort of study in this work is a group of Polynesian New Zealanders with grade III obesity and T2DM. This population is a high risk group for diabetes. The group for the aerobic training is, n=9; mixed gender; average age 48 ± 6 y, with BMI 48 ±8 and weight 123 ±29 kilograms. These participants underwent 16 weeks of supervised aerobic exercise training [cycling and walking]. Blood and Muscle biopsies [Vastus Lateralis] were taken following an overnight fast before and after the exercise intervention, where the post training biopsies were taken after 48 hours of last exercise session.

Table 1: Subject and session information

ID	Group	Week 0	Week 16	Sessions Completed	% Completed	Age	Sex
15	Aerobic	Completed	Completed	40	83.33	52	Female
8	Aerobic	Completed	Completed	39	81.25	49	Female
14	Aerobic	Completed	Completed	44	91.67	59	Male
6	Aerobic	Completed	Completed	47	97.92	46	Female
19	Aerobic	Completed	Completed	22	45.83	47	Female
22	Aerobic	Completed	Completed	24	50.00	51	Female
26	Aerobic	Completed	Completed	25	52.08	47	Female
20	Aerobic	Completed	Completed	34	70.83	53	Female
7	Aerobic	Completed	Completed	39	81.25	53	Male

Identification of Outliers:

To indentify outliers in our dataset, different normalization techniques for each platform were performed on the dataset followed by hierarchical chip-based clustering [Euclidean Distance, Average Linkage UGMA], using The Hierarchical Clustering Explorer 3.0 to cluster samples based on distance. The distance which is driven by the stronger effect in the data was used to detect any poor performing sample or batch effects, if any, and these were noted. We considered both types of outliers, ones that showed major variation from the rest of the data [technical outlier] and the other that showed major "within sample" variations and not in accordance with our assumption, i.e. samples from same subject showing wide variation. Consistent outliers, ones that were noted despite the various normalization techniques used, were excluded from further analysis. Based on the clustering results, the best normalization technique based on signal/noise ratio was used for each of the three datasets (i.e., the gene expression, microRNA and methylation data).

Data Preprocessing and Statistical Analyses:

Gene Expression

The gene expression profiling performed on Illumina® Gene Expression BeadChip Array technology [Illumina, Inc., San Diego, CA] and scanned using HiScanSQ System to obtain decoded images. These images were then analyzed by GenomeStudio™ Gene Expression Module [Illumina, Inc., San Diego, CA]. Data were generated using the three methods in GenomeStudio, with internal normalization and no background correction, without any normalization but background corrected, and without any normalization and no background correction. The signal intensities and the control plots were used for the Quality Control metrics analysis in the GenomeStudio according to the guidelines provided by Illumina and outliers, if any, were noted. Following QC, data was imported into Partek® software, version 6.6 Beta

Copyright © 2012 Partek Inc., St. Louis, MO, USA. Different normalization methods were performed on the dataset to check for signal/noise ratios using unsupervised chip-based clustering in HCE 3.0 [The Hierarchical Clustering Explorer 3.0] to cluster samples based on distance [Euclidean distance, Average Linkage UGPMA]. The method that showed the best biological grouping of samples by cluster analyses was further used to find differentially expressed genes. The tested methods included one-way ANOVA, 2-way ANOVA, and multivariate 3-way mixed model ANOVA. In the mixed model analysis, chip and sample were initialized to have random effect on this dataset, to negate the effect of any chip bias and sample bias in the data. Since the number of replicates in our study was low, it gets difficult for the probes to pass multiple testing, so we used GEA ANOVA model that was performed using the approach as described in [Mansourian et al. 2004] in order to obtain more robust p values for a low sample size. Significant genes with a p value of $p<0.0005$ and fold change $>\pm 1.2$ were used considered significant.

MicroRNA Expression

The Affymetrix Cel files from Affymetrix GeneChip® [Affymetrix, Santa Clara, CA] platform were imported into Partek® software, version 6.6 Beta [Copyright © 2012 Partek Inc., St. Louis, MO, USA] using the RMA and GC-RMA algorithm separately. The data was also imported into ArrayStar [DNA Star, Madison, WI] and PLIER algorithm was performed on the data. Since the samples were all from the same tissue type, we used Perfect Match only while generating this value as recommended by Affymetrix for homogenous tissues samples. Further analyses were performed in Partek® software, version 6.6 Beta [Copyright © 2012 Partek Inc., St. Louis, MO, USA] on the dataset generated by the three different normalization methods. The data was then log transformed followed by quantile normalization. Unsupervised chip-based clustering was done using HCE 3.0 [The Hierarchical Clustering Explorer 3.0] to cluster samples based on distance [Euclidean distance, Average Linkage UGPMA]. The method that showed the best

biological grouping of samples by cluster analyses was further used to find differentially expressed genes. The tested methods included one-way ANOVA, 2-way ANOVA, and multivariate 3-way mixed model ANOVA. In the mixed model analysis, chip and sample were initialized to have random effect on this dataset, to negate the effect of any chip bias and sample bias in the data, to negate the effect of any chip bias and sample bias in the data. MicroRNAs [Human only] were considered significant with a p-value< 0.05 and a fold change of >±1.1. These selection criteria do not seem very stringent but multiple testing yielded very few significant probes owing to the low sample size. But then, it should be noted that it is a hypothetical model and would be further validated using RT-PCR and/or functional assays. These then were used to find the mRNA targets in the mRNA data by integrating in the IPA software [Ingenuity® Systems, www.ingenuity.com]. To find the mRNAs in our dataset most likely regulated by these miRNAs, we used the anti-correlated relationship filter and limited to only those targets that were either highly predicted [TargetScan] or experimentally observed [mirRecords].

Methylation data

The data from Infinium HD Human Methylation 450 BeadChip was normalized to controls and internal probes in GenomeStudio. The data, i.e. the Beta values, were imported into Partek® software, version 6.6 Beta [Copyright © 2012 Partek Inc., St. Louis, MO, USA]. The Beta value was logit transformed to give the M value, where the relation between Beta and M values is given in Equation 1. The rest of the analyses were performed using this value which is known to be more statistically valid for differential analysis of methylation [Laird et al. 2010, Du et al. 2010].

$$Beta_i = \left[\frac{2^{Mi}}{2^{Mi}+1}\right];\ M_i = log_2\left[\frac{Beta_i}{1-Beta_i}\right] \quad \text{--- Equation 1}$$

We performed normalization on the M value thus generated using Variance Stabilization and quantile normalization. Unsupervised chip-based clustering was done using HCE 3.0 [The

Hierarchical Clustering Explorer 3.0] to cluster samples based on distance [Euclidean distance, Average Linkage UGPMA]. The method that showed the best biological grouping of samples by cluster analyses was further used to find differentially expressed genes. The tested methods included one-way ANOVA, 2-way ANOVA, and multivariate 3-way mixed model ANOVA. In the mixed model analysis, chip and sample were initialized to have random effect on this dataset, to negate the effect of any chip bias and sample bias in the data, and probes were considered significant with a p value<0.005. These selection criteria do not seem very stringent but multiple testing yielded very few significant probes owing to the low sample size. But then, it should be noted that it is a hypothetical model and would be further validated using RT-PCR and/or functional assays. In order to find the genes that could be regulated by methylation in our dataset, we used two approaches. The two approaches used to find differentially methylated probes, one using ANOVA with the p<0.005 and the second using Tiling array analysis [Johnson et al. 2006] on the ANOVA results with p<0.01, average length 600 bp. First we tried to find probes that were either within the gene [exonic], in the CpG island, or in the enhancer region of the gene. From these, we selected the ones which were also in our data of differentially expressed mRNA. Again, we only considered the anti-correlated relationship between methylation and gene expression.

Pathway, Network and Function Analyses:

Significantly regulated mRNA data was used to interrogate the Ingenuity Pathway Analysis (IPA) Knowledgebase [Ingenuity® Systems, www.ingenuity.com] to find networks, functions and pathways enriched in the dataset. Considering the limitation of restricting the Ingenuity Knowledgebase filter to just skeletal muscle, we did not use this filter. In order to incorporate the

significant miRNAs and epigenomics data from our analysis into these networks, we added those into the IPA networks obtained by the mRNA network analysis.

Methylation driven networks:

In order to identify networks/functions that are regulated by the differentially methylated genes, we started with these differentially methylated genes and used the IPA "grow" function to get the IPA Knowledgebase-defined upstream and downstream targets of genes in our mRNA dataset. These molecule, were then used to interrogate IPA to find networks that are most likely driven by epigenetic regulation.

Chapter 3 Results

The figure below gives a summary of the integration workflow and the filters used.

Figure 1: Data Preprocessing and Statistical Analysis Workflow

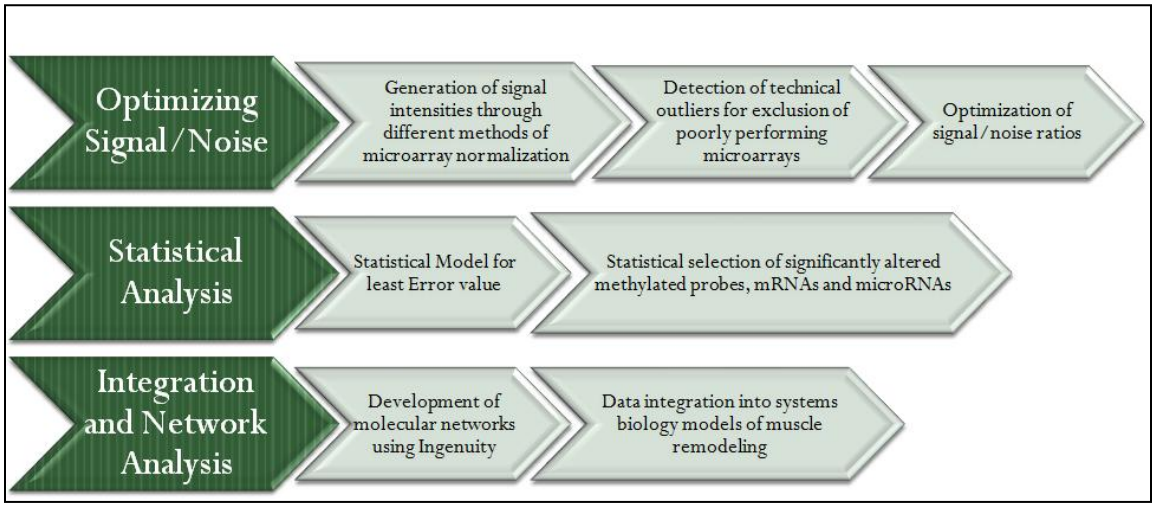


Figure2: Integration workflow

Figure 2a

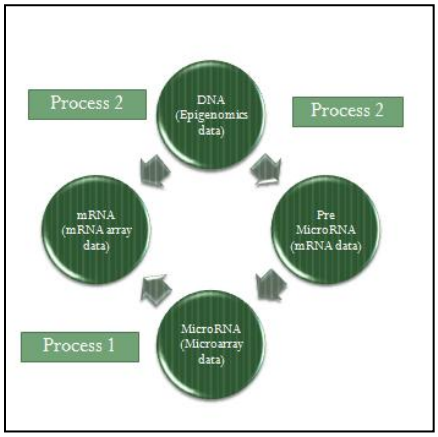
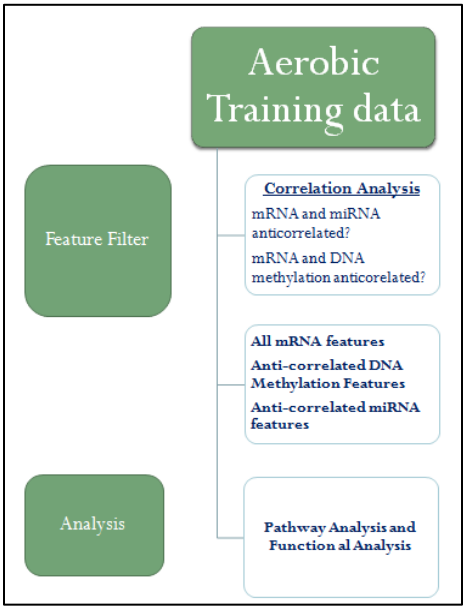


Figure 2b



In Figure2a the Process 1 is the integration of miRNA to mRNA, whereas Process 2 is the integration of Epigenomics data to the mRNA. In a third step involving both Process 1 and Process 2, we try to find the genes regulated by these two processes. Figure 2b shows the Filter selection for integration (Mankoo et al. 2011).

Data interpretation and integration:

The dataset utilized in this bioinformatics research was recently generated by Dr. David Rowlands of New Zealand, Dr. Eric Hoffman of Children's National Medical Center, and the research group in the Center for Genetic Medicine Research at Children's in Washington DC. My role in the research was to carry out the bioinformatics analyses, for which I was primarily responsible. The bioinformatics approach I pursued included:

Generation of signal intensities through different methods of microarray normalization

The generation of signal intensities was based on the specific array platform used and the provider recommendations were considered. For mRNA expression, the decoded images from HiScanSQ System were analyzed by GenomeStudio™ Gene Expression Module [Illumina, Inc., San Diego, CA] with internal normalization and no background correction, without any normalization but background corrected, and without any normalization and no background correction. The result showed that the internal normalization and background subtraction did not show much variation [data not shown]. While performing quality control for the data obtained from these methods, the quality of the data remained more or less the same, so we chose to use the data without any background correction and internal normalization for further analysis for signal generation in GenomeStudio. A similar method was applied to the methylation data when generating the Beta and signal intensities. However, the data with normalization to internal controls outperformed the others in the Quality control analysis. So we chose the Beta values which were generated after normalization to the internal controls in GenomeStudio for downstream analysis.

For the Affymetrix Cel files, generation of signal intensities is a crucial step. Affymetrix provides its own algorithms, such as MAS5, RMA, GC-RMA and PLIER, for the generation of signal intensities. We used all the four different approaches for the data generation for comparison.

Optimization of signal/noise ratios

Illumina Platform

From generating signal intensities to selecting an ANOVA model for statistical analysis, we tried to optimize the signal/noise ratios. For the signal intensities from the Illumina platform, initially we performed the quality control in GenomeStudio for the data generated using the three methods separately, with internal normalization and no background correction, without any normalization but background corrected, and without any normalization and no background correction. The same approach was used for both gene expression and methylation data. Analyzing the signal/noise ratios and other quality controls in GenomeStudio, on data from these methods, the quality of the data remained more or less the same for mRNA, so we chose to use the data without any background correction and internal normalization for further analysis. The mRNA outlier detected in GenomeStudio in the dataset was one of the technical replicates and was excluded from further analysis. Also, we analyzed the unsupervised chip-based clustering to detect if the samples grouped into the appropriate biological group [i.e., sample from the same subject] which would mean that the biological signal is more than the noise in the dataset. Based on the results of the unsupervised clustering results, we chose the normalization method of log transformation followed by quantile normalization for the mRNA dataset and Log-transformed values for methylation that showed better biological signals, representing the change in gene expression due to the exercise intervention.

Affymetrix Platform:

We analyzed all the methods of normalization and the biological signal from them by unsupervised chip-based clustering. The difference in the clustering results using all the three normalization methods did give very similar results. PLIER is the current Affymetrix recommended algorithm which outperforms RMA by introducing a higher reproducibility of

signal [lower coefficient of variation] without loss of accuracy [higher sensitivity to changes in abundance for targets near background and dynamic weighting of the most informative probes in a dataset to determine signal]. Another paper comparing the performance of the Probe set algorithm found PLIER to be more efficient in avoiding false positives [Seo et al. 2006]. So, we chose to select the dataset generated using the PLIER algorithm for our statistical analysis.

We also chose the mix model ANOVA where sample and chip were initialized to have random effect on the data. This model gave a better signal/noise ratio over the one-way and two-way ANOVA.

Detection of technical outliers for exclusion of poorly performing microarrays

For data from the Illumina platform, apart from detecting poorly performing microarray in GenomeStudio, we applied different normalization method to the signal intensities obtained from GenomeStudio such as quantile normalization, division by Q3, and variance stabilization, or no normalization at all. This was then followed by unsupervised clustering to detect outliers using HCE 3.0 [The Hierarchical Clustering Explorer 3.0] to cluster samples based on distance [Euclidean distance, Average Linkage UGPMA]. The assurance of Quality control is very important in microarray data analysis. Most approaches to normalizing expression levels assume that the overall distribution of RNA/methylation/ miRNA levels doesn't change much between samples or across the conditions. So we do not expect to see much “within sample” variation. This analysis helped us detect outliers in the data that did not cluster into the appropriate biological group [same sample clustering together in our case] consistently with the different normalizations, or if they were out-grouped, showed technical variation from the rest of the data. Since these outliers are driven by noise rather than by the biological signal, these samples were excluded from the downstream analyses. The mRNA, methylation and miRNA datasets showed one technical outlier each and these outliers were excluded from further analysis.

Statistical selection of significantly altered methylated genes, mRNAs and microRNAs [e.g. 'gene selection']

Finding differentially expressed mRNA:

mRNA: Since we had a small number of biological replicates, we chose to use the GEA ANOVA model which is known to outperform ANOVA in case of small sample size. Using the GEA ANOVA model on the dataset, we tried to minimize the chances of false positives by selecting a low p value. A standard cut-off for p value is $p < 0.05$ but as the size increases this cutoff is no more significant. In our case with a number of probes around 47000, where a p value of $p < 0.05$ could give large number of false positives, we choose a p value < 0.0005 and a fold change cut off of ± 1.2. This gave us 469 significant differentially expressed transcripts.

Finding Differentially Methylated probes:

The two approaches used to find differentially methylated probes that could have possible regulatory effect on the mRNA dataset, ANOVA with the $p<0.005$, giving differentially methylated probes and the second using tiling array analysis [Johnson et al.2006] on the ANOVA results with $p<0.01$, average length 600 bp, giving differentially methylated genomic regions. There was a clear overlap with the probes found in both methods but the ANOVA results and selection gave more significant probes than the tiling array and was used for our analysis. The list of significant differentially methylated genes is shown in Table 2.

Differentially expressed miRNA:

Using the RMA algorithm for normalization gave 120 significantly differentially expressed miRNA, whereas the method of normalization, PLIER followed by quantile normalization gave 20 significant differentially expressed human miRNAs with a p-value < 0.05 and fold-change $\pm$

1.1. We considered a p-value $p < 0.05$ since the size of the data was smaller (around 20000) when compared to mRNA and methylation. These selection criteria do not seem very stringent but multiple testing yielded very few significant probes owing to the low sample size. But then, it should be noted that it is a hypothetical model and would be further validated using RT-PCR and/or functional assays. All the miRNAs found significant by PLIER followed by quantile were consistent with that found in the list obtained with the RMA normalization adding confidence to the result. While RMA gave more miRNAs, they were not found to be significant with the PLIER normalization. Many of these miRNAs have been shown to be expressed in the skeletal muscle. Table 1 shows these miRNAs with their known role in the skeletal muscle.

Development of molecular networks using Ingenuity

mRNA expression Pathway Analysis:

Differentially expressed genes are shown in two tables, Table3, showing the top up- regulated genes and Table 4 showing the top down-regulated genes.

The significant differentially expressed genes were used for functional analysis using IPA software. The software detected 422 network eligible genes from the 469 used, based on the selected p-value and fold-change cutoffs. The top functions of these are shown in Table 5. These were found to be related to tissue development, organismal development, and skeletal and muscular development.

Methylation driven networks:

We started with the methylated genes to build a network that was most likely regulated by epigenetic changes, since the top mRNA networks are not necessarily regulated by epigenetic changes. This gives a better understanding of the functions associated with and regulated by methylation. These networks and the summary of the networks are shown figures 7 and 8.

Data integration into systems biology models of muscle remodeling.

To understand the epigenetic regulation of these networks, we incorporated the significant miRNAs and significant differentially methylated genes into these mRNA top networks using the workflow shown in figure 2. The network and their summary are shown in Table 5 and figures 3-7, where the network scores are based on the number of eligible molecules they contain. The higher the score, the lower is the probability of finding those molecules in the network by random chance. In constructing these networks, we considered only the anti-correlation relationship between mRNA expression and either miRNA expression or DNA methylation.

Table 2: Important MicroRNAs in our dataset and their known roles

Micro RNA	Importance	Reference
Mir-29b	Insulin resistance and play a crucial role in diabetes	He et al. 2007
Mir-181	aerobic exercise significantly up-regulates the micro RNA mir-181 which play an important role in muscle adaptation , miRNA is strongly up regulated during myoblast differentiation	Safdar et al. 2009
Mir-30c	down regulated by insulin and its down regulation, may be important physiological pathways in skeletal muscle	Eisenberg et al. 2008, Granjon et al. 2009
let7	regulate adipogenesis and thus play an important role in obesity and metabolic associated disorders	Sun et al. 2009
Mir-222/221	Modulate Differentiation and Maturation of Skeletal Muscle Cells	Baggish et al. 2011, Cardinelli et al. 2009

Table 3 Differentially Methylated Genes:

Table 3a: Hyper-methylation and mRNA down regulation:

Symbol	Fold change in mRNA	Function
OTUD1	-1.78871	
DUSP1	-1.571	Oxidative stress, cell cycle
FEZ2	-1.464	Axonal outgrowth
ZC3H3	-1.313	
ENPP7	-1.3	Negative regulation of cell proliferation, Metabolic activity

Table 3b: Hypo-methylated and mRNA up-regulated

Symbol	Fold Change in mRNA	Function
IRF1	1.299	Tumor suppressor
SMOC2	1.3	Promotes matrix assembly, cell proliferation and migration
MEOX2	1.3	myogenesis
CMIP	1.32	T cell signaling
IGF2	1.329	Growth promoter, amplifier of glucose mediated insulin secretion
HLA-DRA	1.377	Immune response
CXCL14	1.38	Immune, Inflammatory response
COL6A3	1.4	Tissue formation
LAMA5	1.47	Attachment, organization, migration of cells
VWF	1.48	homeostasis
EPAS1	1.51	Angiogenesis, hypoxia, heart rate
PLS3	1.52	Actin binding

Table 4 Top Up regulated mRNA

Molecules	Expression value	FUNCTION
COL4A1	2.343	Angiogenesis, differentiation, axonal guidance
BGN	2.108	Collagen assembly, extracellular matrix binding
ACTA2	1.982	Muscle filament assembly, apoptosis, muscle contraction
FABP4	1.847	Lipid metabolism, positive proliferation and inflammatory regulation
COL3A1	1.776	Axonal guidance, cell matrix adhesion, skeletal muscle development
MYH11	1.716	Axon guidance, muscle contraction, filament assembly
CD93	1.678	Cell adhesion, macrophage activation
COL4A2	1.624	Angiogenesis, extracellular matrix organization, axon guidance
FABP5	1.623	Glucose metabolism, lipid metabolism

Table 5 Top down regulated mRNA

Molecules	Expression values	Function
ACTC1	-2.192	Filament assembly, catabolic process, apoptotic process
FOS	-1.808	Regulation of proliferation, differentiation, apoptosis and cell death
MYH1	-1.761	Muscle contraction
RAN1	-1.70	Skeletal muscle fiber development, response to mechanical stimulus
KBTBD5	-1.669	
ANKRD2	-1.644	Muscle contraction
SMTNL1	-1.628	Negative regulation of vasodilatation, muscle organ morphogenesis
DUSP1	-1.571	Cell cycle, inactivation of MAPK activity, apoptosis
C21orf7	-1.566	
TPPP3	-1.565	Microtubule bundle formation

Table 6a Top mRNA Functions:

Molecular Function	# of Molecules
Cellular Movement	109
Cellular Growth and Proliferation	154
Cellular development	124
Cell Death	134
Cell Morphology	96

Table 6b Top Physiological Functions

Physiological Function	# of Molecules
Organismal Development	110
Cardiovascular system development and function	100
Tissue Development	157
Skeletal and Muscular system development and function	99
Embryonic development	95

Figure 3-9 Top mRNA Networks and Top Methylation Driven Networks: [Red shows up regulation of mRNA/miRNA Green down regulation of mRNA/miRNA, Yellow shows Hypo-methylation and up regulation and Orange shows Hyper-methylation and down regulation]

Figure 3 Network 1: Connective Tissue Disorder, Genetic disorder, Skeletal and Muscular Disorder [Score: 38]

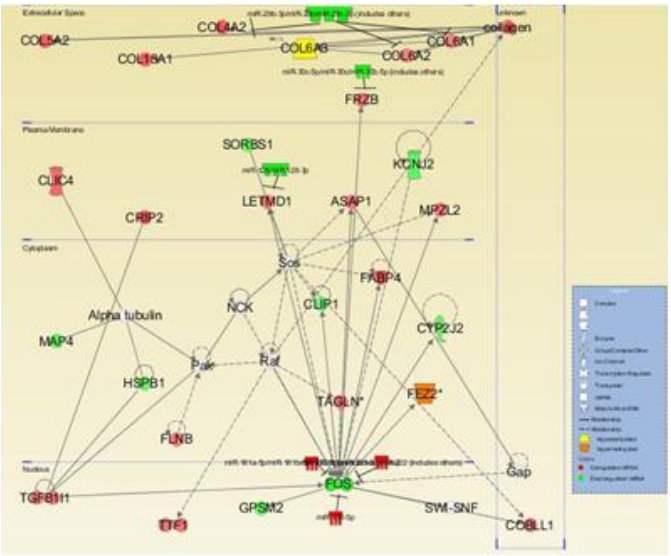


Figure 4 Network 2: Tissue development, Cell to cell signaling and Interaction, Cellular Assembly and Organization [Score: 36]

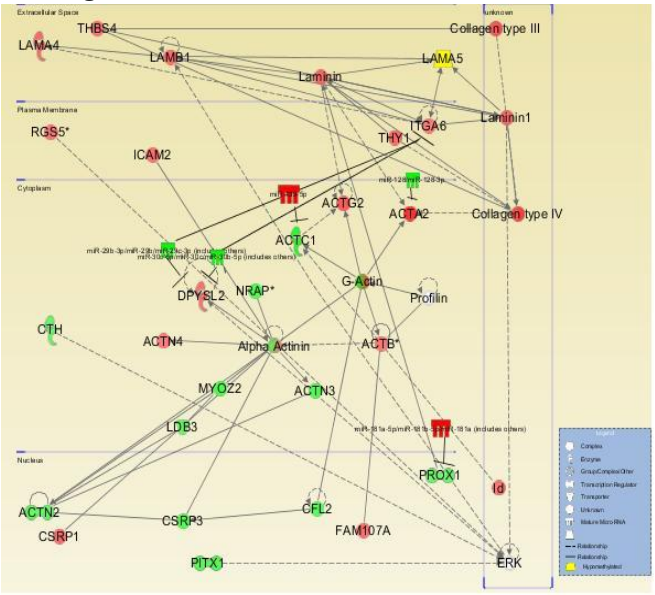


Figure 5: Network 3: Genetic Disorder, Hematological Disorder, Inflammatory Disease [Score: 32]

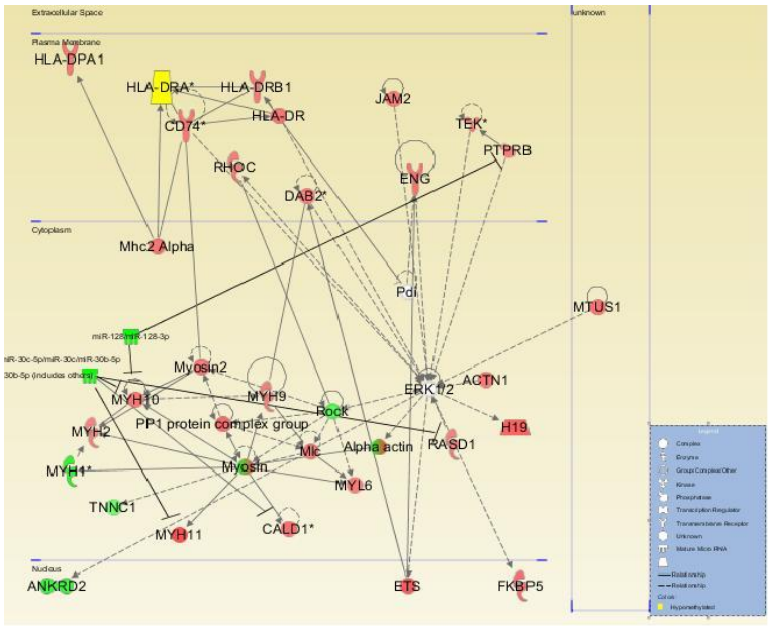


Figure 6: Network 4: Carbohydrate Metabolism, Molecular Transport, Inflammatory Disease [Score: 26]

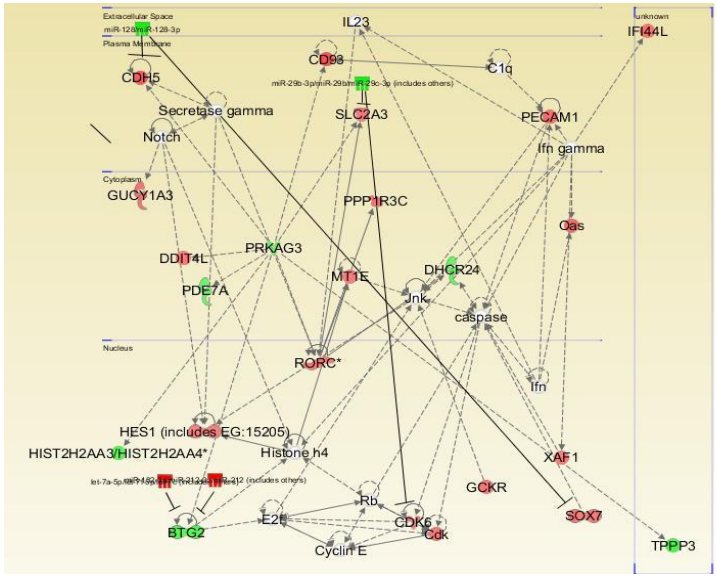
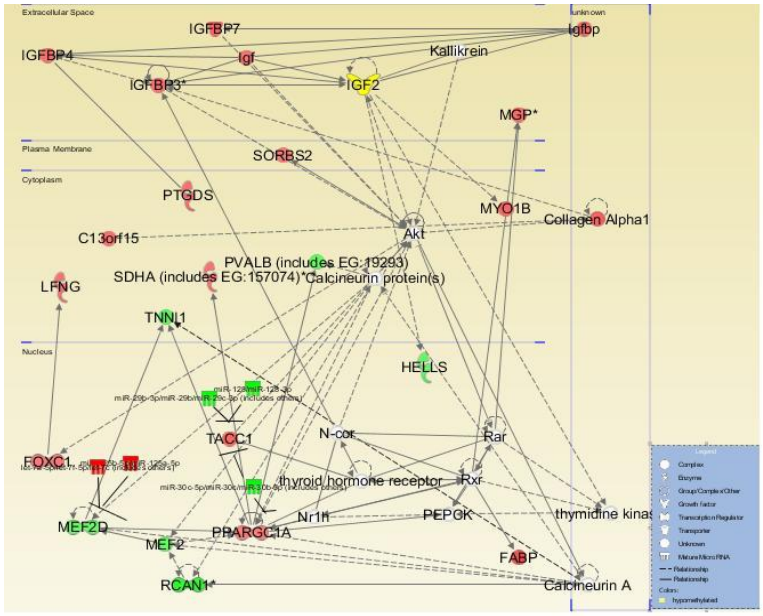


Figure: 7: Network 5: Skeletal and Muscular System development and Function, Cardiovascular system Development and Function, Cell Morphology [Score: 22]



Methylation driven Networks

Figure 8 Network 1: Connective Tissue Disorders, Genetic Disorder, Dermatological Disease and Condition [Score: 50]

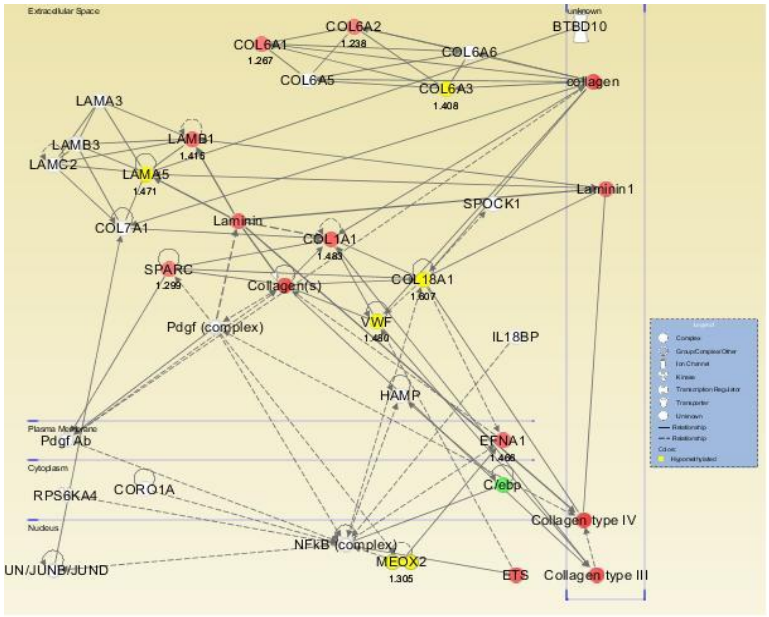
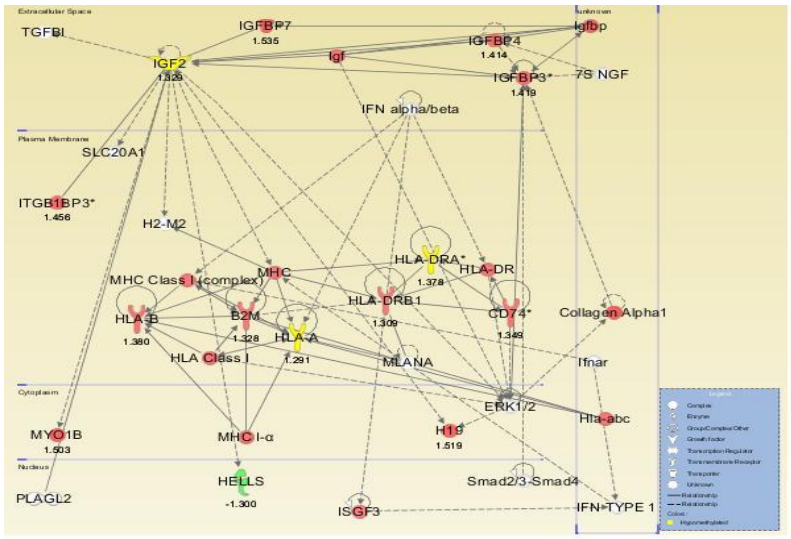


Figure 9 Network 2: Dermatological Disease and Condition, Immunological Disease, Inflammatory Disease [Score: 39]



Chapter 4 Discussion

There are various studies that try to determine the mechanism of the beneficial effect of exercise on muscle, especially in patients with type 2 diabetes mellitus. Our study was based on the effect of aerobic exercise on an ethnic group with T2DM using Integrated Systems Biology approaches, considering both miRNA and DNA level regulation of the networks. This study is the first of its kind trying to uncover the role of epigenomics in this exercise-induced muscle adaptation. However, it should be noted that there are certain limitations to in-silico analysis. Although we tried to eliminate false positives at every step of data normalization and statistical analyses, different experimental conditions might result in different interactions in the networks so formed. Finding targets that could be regulated by methylation was done using both the approaches using Tiling array analysis [Johnson et al. 2006] and by looking at probe level anti-correlation with the gene. Both the methods gave almost the same results adding confidence to our results. Although methylation is known to both activate and silence gene expression [Ndlovu et al 2011], we limited our study by only considering the anti-correlated relationship between gene expression and gene associated methylation. Incorporating epigenomics into the mRNA network to study its regulation could have been done in different ways. Since networks and functions regulated by methylation are not necessarily the top networks formed from significant mRNA data, we tried to build methylation-driven networks starting from the differentially methylated genes. Here, we used an approach starting with the epigenetically regulated genes and building a network from it, including its downstream and upstream targets in the mRNA dataset, to better study the regulation effect rather than just incorporating these differentially methylated genes in the network. To study the miRNA regulation, we limited our miRNA targets in the differentially expressed mRNA to only highly predicted and experimentally observed targets. This might miss some predicted targets but avoided false positives. Another limitation in integrating miRNA into the networks and in considering only the anti-correlation between miRNA and mRNA is that

miRNA does not necessarily act only on mRNA degradation but also can interfere with the translation process of mRNA [Filipowicz et al. 2008, Chekulaeva et al. 2009].

Analysis using the IPA software of the mRNA networks:

Using the selected p value and fold change cutoffs as filters, IPA software selected 422 molecules for the analysis out of 469 used to interrogate the software. There were many molecules common to multiple networks, and most of these were associated with tissue development, skeletal and muscular development, and organismal development.

Network 1: **Connective Tissue Disorder, Genetic disorder, Skeletal and Muscular Disorder [Score: 38]**

Network 1 shows the enrichment of genes involved in tissue development, dominated by the group of collagens which play an important role in formation of tissue from cells and hence play a role in muscular development. This result is consistent with previous studies showing exercise to play a positive regulatory effect on synthesis of collagens [Mackey at al. 2005]. There is hypo-methylation and up-regulation of COL6A3 which play an important role in organization of matrix components. FOS, which is predicted to be a target of three miRNAs, miR-181, miR-222 and miR-189, is down-regulated, and could have a possible role in the regulation of apoptosis. Although previous studies have shown FOS to be up-regulated following an exercise intervention, our results show its down-regulation, one possible reason for this could be regulation of apoptosis.

Network 2: Tissue development, Cell to cell signaling and Interaction, Cellular Assembly and Organization [Score 36]

Network 2 consists of mostly the genes involved in tissue development, cellular movement and these include the group of laminins and actinins. The network shows the up-regulation of the laminin gene family [Lama5, Lama4, LamaB1] which are known to mediate attachment, migration and organization of cells into tissue by interacting with other extracellular matrix component. This result is consistent with the previous studies showing up regulation of the laminin gene family after endurance exercise leading to muscle adaptation [Timmons et al. 2005]. TheLAMA5 gene, belonging to the alpha subfamily of laminin gene family and major component of basement membrane, was also found to be hypo-methylated and up regulated, thus promoting morphogenesis to tissue. Other members of the network [ACTB, ACTC1, CSRP3, DPYSL2, ITGA6, PITX1, PROX1, THBS4 AND THY1] also are involved in the development of tissue, with ACTN2, DPYSL2, ITGA6, LAMA5, THBS4, THY1 involved in cytoskeleton organization and actinins involved in muscle contraction and thus leading to muscular development. Micro-RNAs Mir-30c and Mir-29b predicted to target DPYSL2 and/or ITGA6 are down regulated. Since DPYSL2 (miRNA regulated) plays an important role in cytoskeletal remodeling (ref), this might play a regulatory role in cytoskeleton organization and/or cell adhesion.

Network 3: Genetic Disorder, Hematological Disorder, Inflammatory Disease [Score: 32]

Network 3 again shows the enrichment of tissue development function including cell proliferation mainly involving the interaction of myosins. Myosins are known to promote muscle contraction and thus muscular development [Uren et al. 2000], Human MYL/h6 is involved in development of skeletal muscle and angiogenesis whereas human TEK increases angiogenesis of blood vessels. Another member of the network, ENG, is associated with angiogenesis of blood vessels and in cytoskeleton organization [Sanz-Rodrigues et al. 2004]. Thus these genes seem to be functioning together towards tissue development after the exercise intervention. The network shows an up regulation of the genes regulated by ERK1/2. Although members of the ERK1/2 are not in the dataset, the genes centered around it [JAM2, PTPRB, ENG, RHOC, ACTN1, H19,

HLA-DRA, ETS, FKBP5, MTUS1, TEK, CD74] are up regulated and function in the differentiation and proliferation of cells. Hypo-methylation of HLA-DRA might be important from the diseased point of view, as it was shown in a mouse model to repress diabetes [Johnson et al. 2001]. MicroRNAs like miR-128 and miR-30c are down-regulated in the network, where miR-30 is previously known to play a role in physiological adaptation in skeletal muscle after exercise intervention [Eisenberg et al. 2008, Granjon et al. 2009]. This miRNA was predicted to target the myosins, showing regulation of muscular contraction and development which might be mediated by this microRNA.

Network 4: Carbohydrate Metabolism, Molecular Transport, Inflammatory Disease [Score: 26]

This network is enriched with molecules involved in carbohydrate metabolism. The network genes may improve the transport of D-glucose and synthesis of glycogen by up-regulation of SLC2A3 [GLUT3] and PPP1R3C. IPA functional analysis predicts reduction of glucose metabolism disorder and diabetes mellitus, with 7 molecules [BTG2, GCKR, HES1, MT1E, PDE7A, PPP1R3C AND PRKAG3] in the network associated with glucose metabolism disorder, and 5 of the 7 molecules, GCKR, HES1, MT1E, PDEFA, PPP1R3C, going in the direction of negatively affecting insulin resistance. This shows that the intervention helped in the improvement of glucose transport and breakdown in the diabetic muscle with impaired glucose metabolism. MicroRNA miR-29b, which is supposed to play a crucial role in diabetes [He at al. 2007] and predicted to target SLC2A3, is down-regulated. The transcription factor BTG2, which has an anti-proliferative property is down-regulated, which is a predicted target of two microRNAs in the network, miR-132 and let-7a, showing a possible positive regulation of proliferation.

Network 5: Skeletal and Muscular System development and Function, Cardiovascular System Development and Function, Cell Morphology [Score: 22]

This network is enriched with molecules associated with cell differentiation and tissue development, including the IGF family of molecules [IGFBP4, IGFBP3, IGFBP7, IGF2]. Although the dataset does not include AKT, there is an up regulation of molecules regulated by AKT like IGF2, IGFBP7, IGFBP4, IGFBP3, SORBS2, FOXC1, TACC1 and PPARGC1A. The up regulation of PPARGC1A is consistent with previous findings showing its up-regulation following exercise and is a key feature in mitochondrial biogenesis. Molecules associated with glucose metabolism disorder, IGFBP3, PPARGC1A and PTGDS, which are up-regulated in the network are known to decrease glucose metabolic disorder. Here the hypo-methylation of IGF2 might be really important. A previous study on the effect of aerobic exercise has shown the up-regulation of IGF2 and IFBP4, IGBP7 with positive physiological adaptation [Timmons et al. 2005]. Another study showed that the low level of IGF2 might be associated with weight gain and obesity [Sandhu et al. 2003]. Hypo-methylation of this imprinted gene in the TSS region, [which was highly methylated in all samples prior to exercise, considering the Beta values before exercise] shows that it might be an important regulatory gene. There seems to be a microRNA regulation of the TACC1 gene, which plays a role in differentiation, by miR-128, miR-30c and miR-29b.

<u>Methylation Driven Networks:</u>

Network 1: Connective Tissue Disorders, Genetic Disorder, Dermatological Disease and Condition [Score: 50]

This is also the top network which emerged from the mRNA network analysis. This network shows the enrichment of tissue development function, where the collagens and laminins group play a major role. Two of the members of the collagen family, COL6A3 and COL18A1, and

LAMA5 of the laminin family show hypo-methylation, giving insight into the epigenetic regulation muscle development. The homeobox gene MEOX2, which is important in skeletal muscle tissue development, also shows an up-regulation and hypo-methylation [Otto et al. 2010]. VWF, an endothelial marker, which is known to increase with exercise, is also hypo-methylated in this network and has a function in maintaining homeostasis.

Figure 7 Network 2: Dermatological Disease and Condition, Immunological Disease, Inflammatory Disease [Score: 39]

This network again shows genes involved in growth, proliferation, differentiation, immune response and glucose metabolism disorder. This network seems to be regulated by the hypo-methylation of IGF2, HLA-DRA and HLA-A, which might play a role in cell proliferation and/or glucose metabolism regulation.

Conclusion:

Aerobic exercise training plays an important role in the management of Type 2 diabetes undergoing several adaptive changes in the skeletal muscle. The research was a pilot study to model the skeletal muscle response in T2DM subjects due to 16 weeks of aerobic exercise training. The approach used an integrated data to represent the molecular mechanism underlying this intervention and study the DNA (due to methylation) and RNA (due to miRNA) level of regulation. The networks generated showed an improvement in muscular development, endurance and carbohydrate metabolism following the intervention. Although it should be noted that it is a hypothetical model and would need further validation using methods like RT-PCR and/or functional assays. There were several approaches for the data integration which we tried to compare and contrast before we followed one for the downstream analysis. The regulation via miRNA is expected provide a better understanding after the integration of proteomics data into these networks. And for integrating the proteomics data we would use Bayesian approaches for amore probabilistic model of these adaptations. These predictions will then be validated using methods like RT-PCR and/or functional assays and is expected to give insight into the beneficial adaptive changes due to aerobic exercise training in the diabetic muscle.

References:

1. **World Health Organization**. Obesity 2008. Available at: http://www.who.int/topics/obesity/en/

2. **Baggish, A. L**., Hale, A., Weiner, R. B., Lewis, G. D., Systrom, D., Wang, F., et al. [2011]. Dynamic regulation of circulating microRNA during acute exhaustive exercise and sustained aerobic exercise training. The Journal of Physiology, 589[Pt 16], 3983-3994. doi:10.1113/jphysiol.2011.213363

3. **Bornfeldt, K. E**., & Tabas, I. [2011]. Insulin resistance, hyperglycemia, and atherosclerosis. Cell Metabolism, 14[5], 575-585. doi:10.1016/j.cmet.2011.07.015

4. **Cardinali, B**., Castellani, L., Fasanaro, P., Basso, A., Alema, S., Martelli, F., et al. [2009]. Microrna-221 and microrna-222 modulate differentiation and maturation of skeletal muscle cells. PloS One, 4[10], e7607. doi:10.1371/journal.pone.0007607

5. **Cauza, E**., Hanusch-Enserer, U., Strasser, B., Kostner, K., Dunky, A., & Haber, P. [2005]. Strength and endurance training lead to different post exercise glucose profiles in diabetic participants using a continuous subcutaneous glucose monitoring system. European Journal of Clinical Investigation, 35[12], 745-751. doi:10.1111/j.1365-2362.2005.01573.x

6. **Chekulaeva, M**., & Filipowicz, W. [2009]. Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. Current Opinion in Cell Biology, 21[3], 452-460. doi:10.1016/j.ceb.2009.04.009

7. **Church, T.** S., Earnest, C. P., Skinner, J. S., & Blair, S. N. [2007]. Effects of different doses of physical activity on cardiorespiratory fitness " among sedentary, overweight or obese postmenopausal women with elevated blood" pressure: A randomized controlled trial. JAMA : The Journal of the American Medical Association, 297[19], 2081-2091. doi:10.1001/jama.297.19.2081

8. **de Groot, P**. C., Dekkers, O. M., Romijn, J. A., Dieben, S. W., & Helmerhorst, F. M. [2011]. Human Reproduction Update, 17[4], 495-500. doi:10.1093/humupd/dmr001

9. **Du, P**., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L., et al. [2010]. Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics, 11, 587. doi:10.1186/1471-2105-11-587

10. **Durstine, J. L**., & Haskell, W. L. [1994]. Effects of exercise training on plasma lipids and lipoproteins. Exercise and Sport Sciences Reviews, 22, 477-521.

11. **Duscha, B. D**., Slentz, C. A., Johnson, J. L., Houmard, J. A., Bensimhon, D. R., Knetzger, K. J., et al. [2005]. Effects of exercise training amount and intensity on peak oxygen consumption in middle-age men and women at risk for cardiovascular disease. Chest, 128[4], 2788-2793. doi:10.1378/chest.128.4.2788

12. **Engelgau, M. M**., Geiss, L. S., Saaddine, J. B., Boyle, J. P., Benjamin, S. M., Gregg, E. W., et al. [2004]. The evolving diabetes burden in the united states. Annals of Internal Medicine, 140[11], 945-950.

13. **Eisenberg, I.**, Eran, A., Nishino, I., Moggio, M., Lamperti, C., Amato, A. A., et al. [2007]. Distinctive patterns of microRNA expression in primary muscular disorders. Proceedings of the National Academy of Sciences of the United States of America, 104[43], 17016-17021. doi:10.1073/pnas.0708115104

14. **Fernandez-Valverde, S. L**., Taft, R. J., & Mattick, J. S. [2011]. Diabetes, 60[7], 1825-1831. doi:10.2337/db11-0171

15. **Filipowicz, W**., Bhattacharyya, S. N., & Sonenberg, N. [2008]. Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? Nature Reviews.Genetics, 9[2], 102-114. doi:10.1038/nrg2290

16. **Galgani, J. E**., Moro, C., & Ravussin, E. [2008]. Metabolic flexibility and insulin resistance. American Journal of Physiology.Endocrinology and Metabolism, 295[5], E1009-17. doi:10.1152/ajpendo.90558.2008

17. **Gallagher, E. J**., Leroith, D., & Karnieli, E. [2011]. The metabolic syndrome--from insulin resistance to obesity and diabetes. The Medical Clinics of North America, 95[5], 855-873. doi:10.1016/j.mcna.2011.06.001

18. **Granjon, A**., Gustin, M. P., Rieusset, J., Lefai, E., Meugnier, E., Guller, I., et al. [2009]. The microRNA signature in response to insulin reveals its implication in the transcriptional action of insulin in human skeletal muscle and the role of a sterol regulatory element-binding protein-1c/myocyte enhancer factor 2C pathway. Diabetes, 58[11], 2555-2564. doi:10.2337/db09-0165

19. **Gregg, E. W**., Cheng, Y. J., Narayan, K. M., Thompson, T. J., & Williamson, D. F. [2007]. The relative contributions of different levels of overweight and obesity to the increased prevalence of diabetes in the united states: 1976-2004. Preventive Medicine, 45[5], 348-352. doi:10.1016/j.ypmed.2007.07.020

20. **Halbert, J. A**., Silagy, C. A., Finucane, P., Withers, R. T., Hamdorf, P. A., & Andrews, G. R. [1997]. The effectiveness of exercise training in lowering blood pressure: A meta-analysis of randomised controlled trials of 4 weeks or longer. Journal of Human Hypertension, 11[10], 641-649.

21. **He, A**., Zhu, L., Gupta, N., Chang, Y., & Fang, F. [2007]. Overexpression of micro ribonucleic acid 29, highly up-regulated in diabetic rats, leads to insulin resistance in 3T3-L1 adipocytes. Molecular Endocrinology [Baltimore, Md.], 21[11], 2785-2794. doi:10.1210/me.2007-0167

22. **Johnson, E. A**., Silveira, P., Chapman, H. D., Leiter, E. H., & Serreze, D. V. [2001]. Inhibition of autoimmune diabetes in nonobese diabetic mice by transgenic restoration of H2-E MHC class II expression: Additive, but unequal, involvement of multiple APC subtypes. Journal of Immunology [Baltimore, Md.: 1950], 167[4], 2404-2410

23. **Johnson, W. E**., Li, W., Meyer, C. A., Gottardo, R., Carroll, J. S., Brown, M., et al. [2006]. Model-based analysis of tiling-arrays for ChIP-chip. Proceedings of the National Academy of Sciences of the United States of America, 103[33], 12457-12462. doi:10.1073/pnas.0601180103

24. **Johnson, J. L**., Slentz, C. A., Houmard, J. A., Samsa, G. P., Duscha, B. D., Aiken, L. B., et al. [2007]. Exercise training amount and intensity effects on metabolic syndrome [from

studies of a targeted risk reduction intervention through defined exercise]. The American Journal of Cardiology, 100[12], 1759-1766. doi:10.1016/j.amjcard.2007.07.027

25. **Katsiki, N**., Ntaios, G., & Vemmos, K. [2011]. Maturitas, 69[3], 239-243. doi:10.1016/j.maturitas.2011.04.010

26. **Keller, P**., Vollaard, N., Babraj, J., Ball, D., Sewell, D. A., & Timmons, J. A. [2007]. Using systems biology to define the essential biological networks responsible for adaptation to endurance exercise training. Biochemical Society Transactions, 35[Pt 5], 1306-1309. doi:10.1042/BST0351306

27. **Keller, P**., Vollaard, N. B., Gustafsson, T., Gallagher, I. J., Sundberg, C. J., Rankinen, T., et al. [2011]. A transcriptional map of the impact of endurance exercise training on skeletal muscle phenotype. Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology, 110[1], 46-59. doi:10.1152/japplphysiol.00634.2010

28. **Kelley, G. A**., & Sharpe Kelley, K. [2001]. Aerobic exercise and resting blood pressure in older adults: A meta-analytic review of randomized controlled trials. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 56[5], M298-303.

29. **Kraus, W. E**., Houmard, J. A., Duscha, B. D., Knetzger, K. J., Wharton, M. B., McCartney, J. S., et al. [2002]. Effects of the amount and intensity of exercise on plasma lipoproteins. The New England Journal of Medicine, 347[19], 1483-1492. doi:10.1056/NEJMoa020194

30. **Kurukulasuriya, L. R**., Stas, S., Lastra, G., Manrique, C., & Sowers, J. R. [2011]. Hypertension in obesity. The Medical Clinics of North America, 95[5], 903-917. doi:10.1016/j.mcna.2011.06.004

31. **Laird, P. W**. [2010]. Principles and challenges of genomewide DNA methylation analysis. Nature Reviews.Genetics, 11[3], 191-203. doi:10.1038/nrg2732

32. **Laye, M. J**. [2011]. Erratum to JPHYSIOL/2010/189860 muscle specific MicroRNAs are regulated by " endurance exercise in human skeletal muscle soren nielsen, camilla scheele," " christina yfanti, thorbjorn CA akerstrom, anders rinnov nielsen, bente klarlund" " pedersen, and matthew james laye.". The Journal of Physiology, doi:10.1113/jphysiol.2011.205450

33. **Ling C**, Groop L. [2009]. Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes 58[12]:2718-25

34. **Long, Y. C**., Widegren, U., & Zierath, J. R. [2004]. Exercise-induced mitogen-activated protein kinase signalling in skeletal muscle. The Proceedings of the Nutrition Society, 63[2], 227-232. doi:10.1079/PNS2004346

35. **Lund, S**., Holman, G. D., Schmitz, O., & Pedersen, O. [1995]. Contraction stimulates translocation of glucose transporter GLUT4 in skeletal muscle through a mechanism distinct from that of insulin. Proceedings of the National Academy of Sciences of the United States of America, 92[13], 5817-5821.

36. **Mansourian, R**., Mutch, D. M., Antille, N., Aubert, J., Fogel, P., Le Goff, J. M., et al. [2004]. The global error assessment [GEA] model for the selection of differentially expressed genes in microarray data. Bioinformatics [Oxford, England], 20[16], 2726-2737. doi:10.1093/bioinformatics/bth319

37. **Ndlovu, M. N**., Denis, H., & Fuks, F. [2011]. Exposing the DNA methylome iceberg. Trends in Biochemical Sciences, 36[7], 381-387. doi:10.1016/j.tibs.2011.03.002

38. **Nguyen, D. M.,** & El-Serag, H. B. [2010]. The epidemiology of obesity. Gastroenterology Clinics of North America, 39[1], 1-7. doi:10.1016/j.gtc.2009.12.014

39. **Otto, A**., Macharia, R., Matsakas, A., Valasek, P., Mankoo, B. S., & Patel, K. [2010]. A hypoplastic model of skeletal muscle development displaying reduced foetal myoblast cell numbers, increased oxidative myofibres and improved specific tension capacity. Developmental Biology, 343[1-2], 51-62. doi:10.1016/j.ydbio.2010.04.014

40. **Pansuria, M**., Xi, H., Li, L., Yang, X. F., & Wang, H. [2012]. Frontiers in Bioscience [Scholar Edition], 4, 916-931.

41. **Radpour, R**., Barekati, Z., Kohler, C., Schumacher, M. M., Grussenmeyer, T., Jenoe, P., et al. [2011]. Integrated epigenetics of human breast cancer: Synoptic investigation of targeted " genes, microRNAs and proteins upon demethylation treatment.". PloS One, 6[11], e27355. doi:10.1371/journal.pone.0027355

42. **Rowlands, D. S**., Johnson, N. A., Thomson, J. A., Chapman, P., & Stannard, S. R. [2009]. Exogenous glucose oxidation is reduced with carbohydrate feeding during exercise after starvation. Metabolism: Clinical and Experimental, 58[8], 1161-1169. doi:10.1016/j.metabol.2009.03.016

43. **Russell, A. P**., Feilchenfeldt, J., Schreiber, S., Praz, M., Crettenand, A., Gobelet, C., et al. [2003]. Endurance training in humans leads to fiber type-specific increases in levels of peroxisome proliferator-activated receptor-gamma coactivator-1 and peroxisome proliferator-activated receptor-alpha in skeletal muscle. Diabetes, 52[12], 2874-2881.

44. **Safdar, A**., Abadi, A., Akhtar, M., Hettinga, B. P., & Tarnopolsky, M. A. [2009]. miRNA in the regulation of skeletal muscle adaptation to acute endurance exercise in C57Bl/6J male mice. PloS One, 4[5], e5610. doi:10.1371/journal.pone.0005610

45. **Sandhu, M. S**., Gibson, J. M., Heald, A. H., Dunger, D. B., & Wareham, N. J. [2003]. Low circulating IGF-II concentrations predict weight gain and obesity in humans. Diabetes, 52[6], 1403-1408.

46. **Seo, J.**, & Hoffman, E. P. [2006]. Probe set algorithms: Is there a rational best bet? BMC Bioinformatics, 7, 395. doi:10.1186/1471-2105-7-395

47. **Slentz, C**. A., Houmard, J. A., & Kraus, W. E. [2009]. Obesity [Silver Spring], 17 Suppl 3, S27-33. doi:10.1038/oby.2009.385

48. **Teran-Garcia, M**., Rankinen, T., Koza, R. A., Rao, D. C., & Bouchard, C. [2005]. Endurance training-induced changes in insulin sensitivity and gene expression. American Journal of Physiology.Endocrinology and Metabolism, 288[6], E1168-78. doi:10.1152/ajpendo.00467.2004

49. **Timmons, J. A**., Larsson, O., Jansson, E., Fischer, H., Gustafsson, T., Greenhaff, P. L., et al. [2005]. Human muscle gene expression responses to endurance training provide a novel perspective on duchenne muscular dystrophy. FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology, 19[7], 750-760. doi:10.1096/fj.04-1980com

50. **Uren, D**., Hwang, H. K., Hara, Y., Takeda, K., Kawamoto, S., Tullio, A. N., et al. [2000]. Gene dosage affects the cardiac and brain phenotype in nonmuscle myosin II-B-depleted mice. The Journal of Clinical Investigation, 105[5], 663-671. doi:10.1172/JCI8199

51. **Villeneuve, L**. M., & Natarajan, R. [2010]. The role of epigenetics in the pathology of diabetic complications. American Journal of Physiology.Renal Physiology, 299[1], F14-25. doi:10.1152/ajprenal.00200.2010

52. **Zalesin, K. C**., Franklin, B. A., Miller, W. M., Peterson, E. D., & McCullough, P. A. [2011]. Impact of obesity on cardiovascular disease. The Medical Clinics of North America, 95[5], 919-937. doi:10.1016/j.mcna.2011.06.005

53. **Zhu, J**., Jiang, Z., Gao, F., Hu, X., Zhou, L., Chen, J., et al. [2011]. A systematic analysis on DNA methylation and the expression of both mRNA and microRNA in bladder cancer. PloS One, 6[11], e28223. doi:10.1371/journal.pone.0028223

54. **Zierath, J. R**. [2002]. Invited review: Exercise training-induced changes in insulin signaling in skeletal muscle. Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology, 93[2], 773-781. doi:10.1152/japplphysiol.00126.2002

55**.** **Mankoo, P. K**., Shen, R., Schultz, N., Levine, D. A., & Sander, C. (2011). Time to recurrence and survival in serous ovarian tumors predicted from integrated genomic profiles. PloS One, 6(11), e24709. doi:10.1371/journal.pone.0024709