

Microarray Analysis of Differentially Expressed Genes Between Diabetes vs Healthy

Allam Appa Rao, Shyambabu M, Srinubabu Gedela

Received April 20, 2008; Accepted May 15, 2008; Published May 25, 2008

Citation: Allam AR, Shyambabu M, Srinubabu G (2008) Microarray Analysis of Differentially Expressed Genes Between Diabetes vs Healthy. *J Proteomics Bioinform S1: S055-S065*. doi:10.4172/jpb.s1000010

Copyright: © 2008 Allam AR, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The expression profiling of diabetes vs. healthy is a method of identifying genes potentially involved in the pathogenic process. Microarray analysis enable one to determine the relative level of expression of practically all genes in a genome, allowing the prediction of cellular plans for protein synthesis to be established. We therefore took the approach of using microarray analysis to provide a list of genes that are differentially expressed between diabetes vs. healthy. Statistical methodologies are employed for interpretation of microarray results. The present paper discusses the introduction to microarray analysis and statistical methods along with the application of our present study on differentially expressed genes of diabetes vs healthy.

Introduction

Completion of the human genome project, and the availability of complete genomes for model organisms, provided unprecedented prospects to the scientific community to carry out investigations regarding the greater mysteries of life at the molecular level, i.e. “from the bottom”. The availability of several genomic blueprints has allowed new approaches that are based on comprehensive molecular analyses (and which enhance the understanding of biological systems) to be devised especially for biomedical applications. These new approaches offer the potential to describe specific types of genetic changes as well as patterns of altered gene expression and functions that define, for instance, actual medical problems in the context of, but not entirely based on, symptoms. It is anticipated that these new methods will lead to identification of previously unknown features of individual disease characteristics and profile progression and response to treatment on the molecular basis. One of the most powerful tools that has been developed as a vehicle for carrying out such comprehensive analyses is the DNA microarray, or the “Gene chip”, which consists of a flat solid support with multiple probes that can be used to yield analytical signals (Suzuki et al, 2007).

Since its inception, DNA microarray technology has gained widespread popularity for several reasons, including

the fact that it allows a global snapshot of an organism’s gene expression at a given point time to be obtained. This is important because it is widely believed that thousands of genes and their products in a given living organism function in concert in a complicated and well-coordinated way to support its activities. Thus, a technology that allows such a global picture to be obtained enhances the understanding of the molecular-level biology of an organism and is highly desirable from that perspective. Traditional molecular biology methods of research have generally worked on a single experiment basis, determining the functions of a specific gene in given physiological, chemical and/or biochemical conditions, which means that the throughput is very limited and a comprehensive picture is hard to obtain.

Biological Background

Perhaps one of the most fundamental biological precepts is the crucial role played by proteins as functional molecules of living cells. They are known to be responsible for energy production, biosynthesis of macromolecular components, maintenance of the structural architecture of the cell and response to external stimuli. Specialization of cellular functions occurs when certain specific proteins are produced to direct the essential activities of a given cell type. These proteins may be synthesized when the need arises for non-routine functions such as response to environmental results.

On the other hand, ‘housekeeping proteins’ are required for basic processes such as replication, transcription, translation, protein folding and primary metabolism. Recently Xi et al (2007) used a high throughput method called DNase- chip to identify 3,904 DNaseI HS sites from six cell types across 1% of the human genome. A significant number (22%) of DNaseI HS sites from each cell type are ubiquitously present among all cell types studied

Although it is not clear at what levels housekeeping proteins are produced, there is a general agreement that specialized proteins are produced in fluctuating concentrations, and are important for influencing many of the unique cellular dynamics. Thus, understanding the well-orchestrated molecular networks that control the synthesis, stability and degradation of these proteins is important in appreciating most vital biological functions of cells (Sato and Brivinalov, 2006). Understanding these regulatory networks provides insight into possible molecular interventions in cases of cellular malfunction. This is the driving force behind the advent of studies culminating in the recent high throughput technologies in general molecular biology.

Consequently, two options are available for investigating molecular dynamics of the cell: (i) analyzing the complete set of proteins in the cell (proteomics) or (ii) studying the variation in transcription of genetic information that leads to the production of these proteins (transcriptomics). While proteomics provides a snapshot of the status of the current molecular machinery of a cell, transcriptomics allows one to identify the cell’s strategy for protein synthesis in the conditions under which it is being investigated. The goals of both transcriptomics and proteomics are most often met using high-throughput technologies such as DNA microarrays and mass spectrometry. Although protein array technology has also been developed for proteomics, its use is currently not as widespread as its DNA counterpart. DNA microarrays enable one to determine the relative level of expression of practically all genes in a genome, allowing the prediction of cellular plans for protein synthesis to be established. The greater goal of genomics is to determine the functional pathways influenced by the interactions of all the expressed genes in a genome under a specific set of conditions. Unfortunately, this goal has not been met in the past, perhaps due to the lack of technological ability to survey a large number of gene transcripts or proteins simultaneously, and the scarcity of genes whose DNA sequences had been determined (Romero et al, 2006).

The genome is a blueprint for the biology of cells and its transcription is a regulatory step leading to cellular functional diversity. A genome is defined as the entire repertoire of genes in an organism’s chromosomes, while genes are described as sequences of DNA nucleotides capable of encoding biological information. Some genes encode proteins, others functional RNAs such as ribosomal RNAs and transfer RNAs, required for the translation process itself. The fluctuations in the amount of expressed genetic information lead to a cascade of events influencing the cell’s function. If such functions are routine, then it would be expected that the amount of genetic information expressed would stay relatively stable. In principle, for any given expressed gene in a cell, it is possible that a protein, whose function is required by the cell, will be synthesized. In practice however, the quantitative correlation between gene expression and protein synthesis is quite poor due to differences in mRNA stabilities and translational efficiencies.

However, a comprehensive evaluation of whole – genome expression is expected to be very informative with respect to cell dynamics. Consequently, by evaluating fluctuations in the levels of thousands of expressed genes, greater confidence can be placed on inferences concerning the functional needs of the cell.

The molecular transmission of information in eukaryotes follows a pathway between DNA, RNA and proteins. The biological information provided in the DNA nucleotide sequence of a gene is transcribed into mRNA, which is ultimately translated into protein. The mRNA primary transcript is complementary of the DNA sequence and must be correctly spliced to remove non-coding intronic sequences in order to yield the mature mRNA, which consists of information – coding (exonic) segments of a gene. In addition to the coding region, the mature transcript contains a 5’ untranslated region. (UTR), a 1’UTR and a polyadenylation signal which specifies the addition of a polyadenosine tail to the 3’ end. Translation into proteins is performed on ribosomes and starts at an initiator methionine codon (ATG). An initiator transfer RNA (tRNA) forms a complex that results in the beginning of the nascent peptide. Sequentially, complexes are formed between codons and the appropriately charged tRNA and amino acids are added (with the ribosome moving from codon to codon along the mRNA) until a stop codon is encountered. The order of amino acids added during translation is determined by the order of codons on the mRNA between a start codon and a stop codon, known

as an open reading frame (ORF). A caricature of this process is shown in Figure 1.1.

With respect to cellular activity and function therefore, transcription is one of the most fundamental biomolecular processes. This process is controlled of their own genes and other genes, and thereby stabilizes cellular activity. Thus, if transcription is inhibited, entire molecular pathways and cellular functions could be disrupted. This underscores the need to evaluate this process as a whole, in order to integrate inherent cellular dynamics and complexities. In principle, it is possible to monitor transcription by characterizing fluctuations in the relative concentration of mRNA. The link between variations in mRNA abundance and cellular activity is a matter of biological fact, notwithstanding the poor quantitative correlation. When a cell is faced with atypical conditions such as starvation, stress, or infection by disease, it responds by activating a transcriptional program that ensures maintenance of cellular homeostasis. Thus, if a snapshot of the status of transcription in a cell is taken subsequent to a perturbation, it is possible to determine the candidate proteins required to counter the imbalance by analyzing the abundance of mRNAs. Investigations were largely based on single gene studies, and relied heavily on the prior

identification of the genes that played known roles in specific cellular functions. This, unfortunately, underestimated the complexity of the transcriptional programs involved. Undoubtedly, viewing transcription as a complex biological motor that drives the most fundamental cellular processes like growth, development, response to abuse and even death, became a necessity in the understanding of these processes.

Single – Channel Microarrays

Single – channel microarrays represent perhaps some of the best known commercial platforms for DNA microarray technology, epitomized by the Affymetrix Gene Chips (Downey et al, 2006). These are made by synthesizing, in situ, thousands of short nucleotide sequences based on ESTs, cDNAs or genomic DNA on silicon wagers. For purposes of expression monitoring, fluorescent labeled cDNA are hybridized to the array to allow probe-target interactions through base-pairing.

Although these arrays have a number of positive features, there are also several drawbacks. Perhaps chief among these is cost, since the technology is currently proprietary and therefore not subject to market influences. Another important limitation is that the availability of the arrays is restricted to a small number of specific organisms that have been extensively sequenced and that are of general interest. Layout designs are standardized, although custom arrays can be produced at a cost. The requirement of knowledge of exact DNA sequences for the probes has also put these arrays at a relative disadvantage in terms of the discovery of novel genes. In addition, due to the short lengths of the probes, it is anticipated that, when attached to a surface, the bases nearest to the surface will be strictly inaccessible due to duplex formation with complementary molecules in mixture.

Although single–channel arrays are widely used; the focus of our present study will be on two-channel arrays, which appear to have established themselves to greater extent in research/academic laboratories because of their lower cost and greater flexibility.

Two – Channel Microarrays

The basis of two – channel microarray platforms is the comparison of mRNA abundance in similar cell samples fewer than two distinct physiological conditions on a single chip (Sjogren et al, 2007). The approach for accomplishing this can be described in four individual steps. First, mRNA

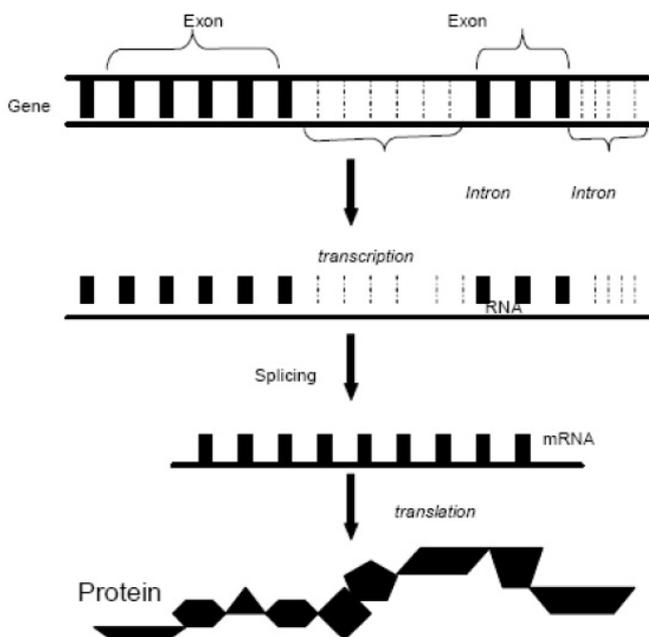


Figure 1.1: Molecular transmission of information in eukaryotes starts from the transcription of a gene to RNA followed by splicing, to eliminate the non – information coding introns, and ultimately to the translation of mRNA to a functional protein.

from samples under two conditions, where one condition is taken to be the reference (I e.g. I normal physiological state), is independently extracted. The amount of mRNA in the two samples is usually normalized through absorbance measurements of the total RNA. In the second step, mRNA from the two extracts is separately copied into cDNA *in vitro*, using an enzymatic reaction known as reverse transcription. During this synthesis, a deoxynucleotide triphosphate labeled with one of two color fluorophores (red or green) or an aminoallyl 1 deoxynucleotide triphosphate, which is subsequently chemically coupled to a fluorophore, is added into the respective reaction mixtures and is incorporated into the synthesized cDNA. Third, equal aliquots of the two-labeled cDNAs are co-hybridized onto a single array containing single – stranded DNA probes. Finally, fluorescence signals emitted by the targets are collected when the array is scanned with lasers set at wavelengths corresponding to the excitation frequencies of the two fluorophores. For every hybridization experiment, the emitted fluorescence is captured and stored as a 16-bit tagged image file format (tiff). The relative abundance of mRNAs in the two samples is calculated as the ratio of the fluorescence intensities of the two dye-labeled cDNAs that hybridized with each probe. The general experimental setup is represented in Figure 1.2. The theory is that each probe will recognize and bind all of its complementary partners in the sample through base pairing since the probes are in relative excess. The non – hybridized transcripts are subsequently washed off so that the emitted fluorescence is exclusively due to hybridized targets. The principle of co-hybridization of transcripts and determination of relative rather than absolute amounts of transcripts is a consequence of the practical aspects of the experimental setup for the spotted microarray platform. Relating the measured fluorescence intensity of hybridized transcripts to absolute gene expression levels is impractical because; (a) the concentration and length of probes among spots on a slide is variable, (b) probe attachment is susceptible to aberrations that lead to non-uniform spot morphologies, and (c) reference standards containing known amounts of transcription products are not generally available. Regarding (a), variation in the amount of probe can occur when the probes are obtained from a library of expressed genes that vary in length. While (b) is not a concern with *in situ* synthesized microarrays, it is a fundamental problem in spotted microarrays. Spotting of probes is performed robotically using pins (print heads) that pick up DNA from 96 – or 384 – well microtitre plates

by capillary action. These deposit probe aliquots sequentially onto many glass microarray slides. Due either to non – uniform surface properties of the glass slides, or temporal wear of the print heads, the shapes of the spots may vary across a slide and among slides. Thus, when the fluorescence intensity is evaluated for each spot, it is common for such morphological anomalies to result in high signal variability. Finally, in view of (c), the lack of reference standards leads to the situation where one of the physiological conditions from which the two cell samples are derived must be considered as a reference state or considered.

This allows transcriptional readjustments in the cells under perturbed chemical or physical environments to be evaluated based on this reference. Thus, analysis of two – channel microarrays involves computing the relative fluorescence intensities of the two dyes for each probe, where the reference sample acts as an internal standard. Ratios are believed to alleviate potential experimental variability resulting from unequal concentrations of probe, cross- hybridization and micro – spotting anomalies. Although this may mitigate some of the variability, other sources of these errors is important in appreciating the context in which two – color microarrays are measured and analyzed.

One of the most widely used methods for ratio calculation is the ratio of medians. This is a method where differential expression is measured as a ratio of the median of pixel intensities within a spot mask for both dyes. The median is intended to represent the center for the distribution of pixel intensities comprised in the spot mask. Perhaps one of the major advantages of this approach is that the measured ratios are robust to influence from a few pixels with extreme values at either end of the distribution. Unfortunately, when spots are characterized by substantial regions (>50%) of low intensity pixels, as in the case of “donuts”, it is anticipated that the low intensity pixels will dominate the spot mask and result in ratios with a high uncertainty.

Another common measure of differential expression involves evaluating the ratio of the mean of pixel intensities within the spot mask. Calculation of mean values is straight forward and less affected by extended regions of low intensity fluorescence, but they are more susceptible to the influence of extreme values at either end of a population, i.e., outliers in pixel population. For this reason, the ratio of means is generally less robust. A less frequently used approach to measuring the relative fluorescence is to calculate pixel –

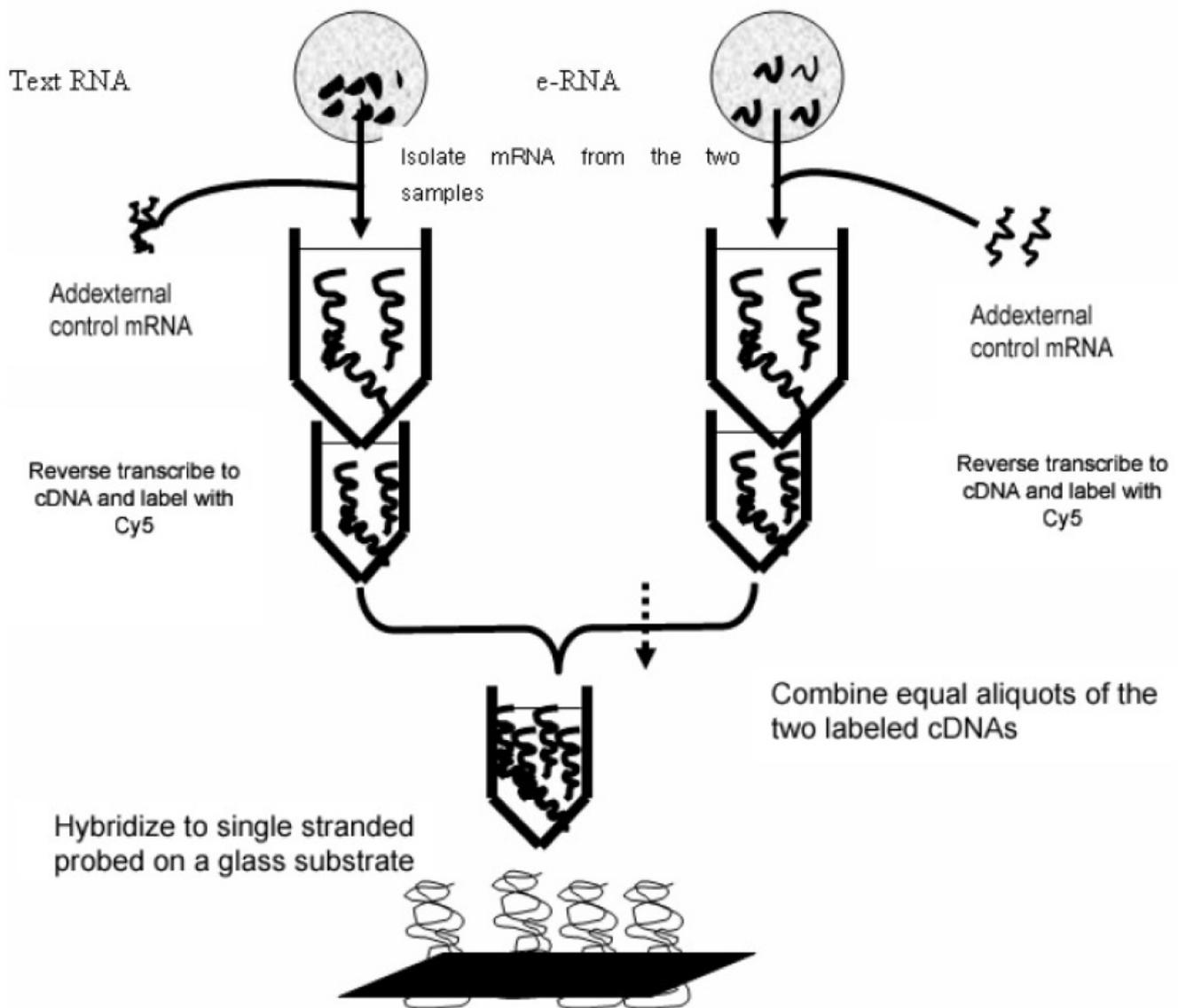


Figure 1.2 : Spotted microarray experimental set –up. mRNA extracts (targets) from cells under two distinct physiological conditions are obtained, reverse transcribed to cDNA and then labeled with different (Cy3 and Cy5) dyes. Equal aliquots of the dye – labeled nucleic acid extracts are combined and applied to a glass substrate onto which single stranded cDNA or oligonucleotide complements (probes) of the mRNA (dye- labeled cDNA) are immobilized.

by-pixel ratios of intensities across the spot and then report the differential expression as the arithmetic mean or median of the ratios. This is referred to as the “mean of ratios” or “median of ratios”, respectively (Bakewell DJ, and Wit E, 2005). A major drawback of this approach, especially when using means, is the high sensi- tivity of the summary statistic to pixels.

Experimental Design Issues

One of the unfortunate consequences of the technical and conceptual simplicity of microarray technology is its

capacity to yield data sets that are biased by inadequate design considerations. In the absence of well – established experimental designs for microarrays, poorly designed experiments continue to yield multiply confounded data with which one is unable to answer the question for which the experiment was conducted. The general objective of designing an experiment is to curtail effects of confounding factors by generating data that span rich and diverse sample spaces, have minimum effects of unwanted variation and provide the potential for maximum efficiency for probing the hypotheses under investigation. Yet, in microarrays,

there is often the false hope that due to the volume of data generated per experiment, confounding factors and unwanted variation will be somewhat mitigated. The focus of interest in microarray studies is typically genes that are differentially expressed in different subjects, different tissues, cells exposed to varying physical/biochemical conditions, or those undergoing growth, development, and degeneration. Some of the common reasons for evaluating these variables are to discover the roles of genes in an organism, to group genes according to common functions, to understand the relationships among genes in a biological system (systems biology), to classify biological specimens (e.g tumor cells) on the basis of gene expression, and to identify important biomarkers in disease progression.

Thus, analysis of these experiments involves identification of genes that display uncharacteristic tendencies of increased or decreased expression and achieving this goal must involve careful experimental design to avoid spurious observations confounded by unrelated experimental variables at multiple levels. Microarray experiments can be regarded multilayered in the sense that they involve several nested levels at which variability may be introduced. In general microarray experiments must be designed into three layers: (1) the selection of experimental units, (2) the design of mRNA extraction, labeling and hybridization, and (3) the arrangement of probes on the glass slides. Whereas the first layer controls the span of the biological design space, the second and third layers account for the analytical (technical) variability at the lower levels of the experimental process and will be the focus of this section.

Higher Level Data Analysis

At the primary level of data analysis, which might be considered as data preprocessing from a chemometrics perspective, the steps are largely the same from one application to another: gridding and segmentation, gagging, image processing, background subtraction, ratio calculation and normalization. Although the details of these steps may differ, in the end the usual result is a vector of ratios and their associated gene identifiers for a series of samples, forming a two – way data matrix for further analysis. At this stage, a variety of methods can be used to coax the desired information from the data, depending on the nature of the experiment. Typical goals include: (1) the identification of genes exhibiting deferential expression (up – or down – regulation) relative to some reference state, (2) the clustering

or classification of genes based on their expression across multiple samples, (4) the identification of genes that may be used as biological markers (e.g for a mutation, a disease, or resistance to some medication), and (5) elucidation of gene function and mechanisms of interaction, i.e. gene networks. In these studies, the term “expression profile” is generally used to describe the normalized ratio (test/ reference) or log-ratio of signals across all genes for a sample represented on a particular microarray. From a chemometrics point of view, it could be considered a kind of “genetic spectrum” except that there is no naturally contiguous ordering of channels.

Changes, not absolute ratios, are important in time course experiments. In other words, a change of 0.5 to 1 is equivalent to a change of 1 to 2 None the less, a consistent point of reference should be chosen. It is also important to note that, due to the proportional error structure, it becomes more useful to determine the normalization factor, a , (Following quation) through a regression of the ratios on the log scale using the model:

$$\log_2 y_i = \log_2 a + \log_2 x_i + I$$

Functional Classification of Differentially Expressed Genes

To determine biological significance of differentially expressed genes, functional classification was performed using Gene Ontology. Gene Ontology reports along with z-score are provided in supplementary material for your reference. Numbers in parentheses indicate number of up-regulated/down-regulated genes and total number of genes (in uploaded data), present in that particular ontology respectively. Z-scores give statistical significance, indicating relative representation up- regulated/down-regulated genes in each function. To determine pathways associated with differentially expressed genes, pathway analysis was performed. Pathway reports are provided in supplementary material. Numbers in parentheses indicate number of up-regulated / down- regulated genes and total number of genes (in uploaded data), present in that particular pathway respectively.

1. Obese Vs Tendency towards Obesity (O Vs HO)

1.1 Molecular function: Genes involved in transforming growth factor beta binding, Sodium : amino acid symporter activity, adenosylhomocysteinase activity, transferase activity, transferring acyl groups, caspase activator activity, NAD(P)H oxidase activity, steroid 21- monooxygenase activity, malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+)

activity, glutamate de- carboxylase activity upregulated in O Vs HO. Genes involved in creatine:sodium symporter activity, glycolipid transporter activity, glycolipid binding, 3-hydroxyisobutyrate dehydrogenase activity, leukemia inhibitory factor receptor activity, superoxide-generating NADPH oxidase activity, chemokine receptor activity, interleukin-22 receptor activity are downregulated in O Vs HO.

Biological process: Genes involved in establishment of cellular localization, cuticle biosynthetic process, hydrogen peroxide, biosynthetic process, vesicle docking are upregulated in O Vs HO. Genes involved in synaptic vesicle membrane organization and biogenesis, response to stimulus, anatomical structure development are down regulated in O Vs HO.

Cellular component: Genes localized in CAAX – protein geranylgeranyltransferase complex are upregulated

in O Vs HO. Genes localized in Golgi transport complex, vesicle, oncostatin-M receptor complex, perikaryon are downregulated in O Vs HO.

Diabetes with History Vs Diabetes without History

D&PH Vs D&NPH1

Molecular function: Genes involved in MHC class II receptor activity, gamma-aminobutyric acid:hydrogen symporter activity, chemokine receptor activity, interleukin-4 receptor activity, interleukin-7 receptor activity, arachidonate 5-lipoxygenase activity, complement receptor activity are upregulated in D&PH Vs D&NPH1. Genes involved in ammonia ligase activity, transaldolase activity, 4-alpha-glucanotransferase activity, choline:sodium symporter activity, interleukin-8 receptor activity are downregulated in D&PH Vs D&NPH1.

Genes Involved in Inflammatory Response

| Condition | Inflammatory Genes (Differentially Expressed) |
|--|--|
| Diabetes with family history vs healthy individual (D&PH vs H) | <i>ALK, GCH1, IFIH1, IFIT1, IL11RA, ITGB2, MAP3K4, MMP19, MMP3, RPS27A, SLK, TNFRSF12A, UBC</i> |
| Diabetes without family history vs healthy individual (D&NPH1 vs H) | <i>CCL3, CDKN1A, CXCL12, HLA-A, IL11RA, KRT8, LTB, MAP3K4, MMP10, MMP19, MMP20, MMP3, RPS27A, TNFSF10, UBC</i> |
| Diabetes without family history vs healthy individual (D&NPH2 vs H) | <i>CCL16, CCR8, CXCL11, CXCL12, FNI, GCH1, HLA-A, IL11RA, LTB, MMP19, MMP3, RHOA, S100A12, SLK, SYK, UBC</i> |
| Obese vs Healthy (O vs H) | <i>CCL13, CXCL12, HLA-A, IL6, KRT8, MMP19, MMP27, MMP3, RPS27A, UBC</i> |

| | |
|--|--|
| Diabetes with family history vs obesity (D&PH vs O) | <i>ALK, CCL13, CCR4, CCR8, HGF, HLA-A, IFIT1, IL20, IL6, IL8RA, ITGB2, KRT8, MC1R, MMP20, MMP27, SLK, TNFRSF12A, UBC</i> |
| Obese individual vs individual with tendency for obesity (O vs HO) | <i>CCL13, CCL16, CCR8, HLA-A, IFIT1, IL6, KRT8, MC1R, MMP16, MMP27, TNFRSF12A, UBC</i> |
| Diabetes with family history vs Diabetes without family history 1 (D&PH VS D&NPH1) | <i>ALK, CCL13, CCR8, CDKN1A, EDN1, FGF1, IFIT1, IL12RB1, IL20, IL22, IL2RG, IL8RA, ITGB2, MMP20, SLK, TNFRSF12A, UBC, XCR1</i> |
| Diabetes with family history vs Diabetes without family history 2 (D&PH VS D&NPH2) | <i>ALK, BLR1, C5, CCL15, CCL16, CCR7, CCR8, CXCL11, CXCL12, FNI, FTH1, GBP1, HLA-A, IFIT1, IL12A, ITGB2, KIT, LTB, MMP20, PPARD, RHOA, RPS27A, TAC1, TLR4, TNFAIP6, TNFRSF11A, TNFRSF12A</i> |

Biological process: Genes involved in cell activation, macromolecule biosynthetic process, hydrogen peroxide biosynthetic process, immune response, regulation of glycolysis are upregulated in D&PH Vs D&NPH1. Genes involved in blastocyst growth, aromatic compound biosynthetic process, nitric oxide biosynthetic process, regulation of glycolysis are downregulated in D&PH Vs D&NPH1.

Cellular component: Genes localized in ribonucleoside-diphosphate reductase complex, interleukin-18 receptor complex, interleukin-1 receptor complex, mitochondrion interleukin-5 receptor complex are upregulated in

D&PH Vs D&NPH1. Genes localized in proteasome activator complex, isoamylase complex, CAAX-protein geranylgeranyltransferase complex, protein kinase CK2 complex, oxoglutarate dehydrogenase complex, MHC class I peptide loading complex are downregulated in D&PH Vs D&NPH1.

D&PH Vs D&NPH2

Molecular function: Genes involved in structural constituent of ribosome, MHC class II receptor activity, ferroxidase activity, NAD(P)H oxidase activity are upregulated in D&PH Vs D&NPH2. Genes involved in 4-alpha-glucanotransferase activity, phosphomannomutase

SUMMARY

| S.N o. | Reference sample* | Test sample | % of DE genes for 2-fold change | Modified fold change (MFC) | % of DE genes for MFC | Up- regulated genes | Down- regulated genes |
|-----------|----------------------|----------------|--|-------------------------------------|-----------------------------|---------------------------|-----------------------------|
| 1 | H | H&O | 6.74 | 2.36 | 4.00 | 1247 | 331 |
| 2 | H | O | 11.0 | 2.94 | 4.40 | 962 | 989 |
| 3 | H | D&NPH1 | 7.69 | 2.37 | 4.38 | 1249 | 477 |
| 4 | H | D&NPH2 | 10.6 | 2.96 | 3.00 | 861 | 356 |
| 5 | H | D&PH | 6.73 | 2.36 | 3.33 | 1211 | 368 |
| 6 | H&O | O | 8.38 | 2.38 | 5.79 | 814 | 1469 |
| 7 | H&O | D&NPH1 | 5.26 | 2.14 | 4.05 | 539 | 1058 |
| 8 | H&O | D&NPH2 | 5.75 | 2.43 | 3.07 | 541 | 672 |
| 9 | H&O | D&PH | 3.85 | 2.07 | 3.36 | 502 | 822 |
| 10 | O | D&NPH1 | 7.75 | 2.07 | 7.13 | 1479 | 1333 |
| 11 | O | D&NPH2 | 8.62 | 2.39 | 5.95 | 1534 | 813 |
| 12 | O | D&PH | 6.93 | 2.17 | 5.93 | 1338 | 1002 |
| 13 | D&NPH1 | D&NPH2 | 5.58 | 2.18 | 4.08 | 948 | 662 |
| 14 | D&NPH1 | D&PH | 4.15 | 2.0 | 4.16 | 686 | 682 |
| 15 | D&NPH2 | D&PH | 6.10 | 2.0 | 6.10 | 676 | 979 |

Table 3.1: Number of up regulated and down regulated genes in each treatment category.

activity, receptor signaling protein tyrosine kinase activity are downregulated in D&PH Vs D&NPH2.

Biological process: Genes involved in intracellular sequestering of iron ion, ribosome biogenesis and assembly, hydrogen peroxide biosynthetic process are upregulated in D&PH Vs D&NPH2. Genes involved in hemostasis, developmental growth, lipid glycosylation, regulation of glycolysis are downregulated in D&PH Vs D&NPH2.

Cellular component: Genes localized in ribosome, ferritin complex are upregulated in D&PH Vs D&NPH2. Genes localized in CAAX-protein geranylgeranyltransferase complex, isoamylase complex, apolipoprotein B mRNA editing enzyme complex, lipopolysaccharide receptor complex, proteasome activator complex are downregulated in D&PH Vs D&NPH2.

Pathway Analysis

Diabetes Vs Normal (D&PH Vs H)

Genes involved in Inositol phosphate metabolism, Starch and sucrose metabolism, Nitrogen metabolism, Oxidative phosphorylation, Androgen and estrogen metabolism, Glycan biosynthesis and metabolism pathways, Metabolism of cofactors and vitamins pathways, MAPK signaling pathway, ECM-receptor interaction, Neuroactive ligand-receptor interaction, Regulation of actin cytoskeleton, Cell communication pathways, Nervous system pathways, Neurodegenerative disorders pathways are upregulated in D&PH Vs H.

Genes involved in Glycolysis / Gluconeogenesis, Propanoate metabolism, Carbon fixation, Biosynthesis of steroids, Fatty acid metabolism, Histidine metabolism, Phenylalanine metabolism, Tyrosine metabolism, Urea cycle and metabolism of amino groups, Cell cycle, Insulin signaling pathway, PPAR signaling pathway, Anti- gen processing and presentation are downregulated in D&PH Vs H.

Genes involved in Cell adhesion molecules (CAMs), Cytokine-cytokine receptor interaction, Insulin signaling pathway, Immune system pathways are downregulated in O Vs H.

Diabetes Vs Obese (D&PH Vs O)

Genes involved in Inositol phosphate metabolism, Oxidative phosphorylation, Amino acid metabolism pathways, Ubiquinone biosynthesis, Signal transduction pathways, Signaling molecules and interaction pathways, Nervous system pathways are upregulated in D&PH Vs O.

Diabetes with History Vs Diabetes without History

D&PH Vs D&NPH1

Genes involved in signal transduction, Regulation of actin cytoskeleton, Antigen processing and presentation, Complement and coagulation cascades, Axon guidance, Neurodegenerative disorders pathways are upregulated in D&PH Vs D&NPH1. Genes involved in carbohydrate pathways are downregulated in D&PH Vs D&NPH1.

D&PH Vs D&NPH2

Genes involved in Oxidative phosphorylation, Metabolism of cofactors and vitamins pathways, Immune system pathways, Nervous system pathways, Metabolic disorders pathways are upregulated in D&PH Vs D&NPH2.

Genes involved in Lipid metabolism pathways, Amino acid metabolism pathways, Glycan biosynthesis and metabolism pathways, Ubiquitin mediated proteolysis, Signal transduction pathways, Signaling molecules and interaction pathways, Insulin signaling pathway, PPAR signaling pathway are downregulated in D&PH Vs D&NPH2.

References

1. Bakewell DJ, Wit E (2005) Weighted analysis of microarray gene expression using maximum-likelihood. *Bioinformatics* 21: 723-729. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Downey T (2006) Analysis of a multifactor microarray study using Partek genomics solution. *Methods Enzymol* 411: 256-70. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
3. Filzmoser P, Garrett RG, Reimann C (2005) Multivariate outlier detection in exploration geochemistry. *Computers and Geosciences* 31: 579-587. » [CrossRef](#) » [Google Scholar](#)
4. Sridhar GR (2007) Psychiatric co-morbidity & diabetes. *BMC Bioinformatics* 8: 387-394. *Indian J Med Res* 125: pp311-320.
5. Garrett RG (1989) The chi-square plot: A tool for multivariate outlier recognition. *Journal of Geochemical Exploration* 32: 319-341.
6. Gervini D (2003) A robust and efficient adaptive reweighted estimator of multivariate location and scatter. *Journal of Multivariate Analysis* 84: 116-144. » [CrossRef](#) » [Google Scholar](#)

7. Kathryn EW, Gökhan SH (2005) Inflammation, stress, and diabetes. *J Clin Invest* 115: 1111-1119. » [Google Scholar](#) » [CrossRef](#)
8. Loguinov AV, Saira Mian I, Vulpe CD (2004) Exploratory differential gene expression analysis in microarray experiments with no or limited replication. *Genome Biology* 5: R18. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
9. Paturi VR, Xinfang L, Patrick P, Mark T, Nandgaonkar S (2004) Gene expression profiles of peripheral blood cells type 2 diabetes and nephropathy in Asian Indians. *Genome Biology* 5: P9. » [CrossRef](#) » [Google Scholar](#)
10. Romero R, Espinoza J, Gotsch F, Kusanovic JP, Friel LA (2006) The use of high-dimensional biology (genomics, transcriptomics, proteomics, and metabolomics) to understand the preterm parturition syndrome. *BJOG* 113 Suppl 3: 118-35. » [Google Scholar](#) » [CrossRef](#) » [Pubmed](#)
11. Rousseeuw PJ, Van ZBC (1990) Unmasking multivariate outliers and leverage points. *Journal of American Statistical Association* 85: 633-651. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
12. Sato N, Brivanlou AH (2006) Microarray approach to identify the signaling network responsible for self-renewal of human embryonic stem cells. *Methods Mol Biol* 331: 267-83. » [Pubmed](#) » [Google Scholar](#)
13. Sjogren A, Kristiansson E, Rudemo M, Nerman O (2007) Weighted analysis of general microarray experiments.» [Google Scholar](#) » [CrossRef](#) » [Pubmed](#)
14. Suzuki T, Tian QB, Kuromitsu J, Kawai T, Endo S (2007) Characterization of mRNA species that are associated with postsynaptic density fraction by gene chip microarray analysis. *Neurosci Res* 57: 61-85. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
15. Tény GE, Christophe H, Dong Y, Olivier BB, Barend M (2006) NADPH Oxidase- Derived Overproduction of Reactive Oxygen Species Impairs Postischemic Neovascularization in Mice with Type 1 Diabetes. *Am J Pathol* 169: 719-728. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
16. Xi H, Shulha HP, Lin JM, Vales TR, Fu Y (2007) Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* 17: 3-8. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
17. Zao HY, Yue PY, Fang KT (2004) Identification of differentially expressed genes with multivariate outlier analysis. *J Biopharm Stat* 14: 629-646.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)