Exploring Differential Nature of Human and Chimpanzee Organs through Linear Correlative and MI Measures

Aurpan Majumder
Dept. of ECE, NIT Durgapur, Durgapur, India
Email: aurpan.nitd@gmail.com

Mrityunjay Sarkar
Dept. of ECE, D.I.A.T.M, Durgapur, India
Email: mrityu1488@gmail.com

***Abstract*—Change in gene expression values (due to mutation) is responsible for the evolution of complex organs in different creatures. Here we are carrying out our analysis in between two very closely related primates Human and Chimpanzee. Both Chimpanzee and Human are extant taxa evolution from a common ancestor due to the change in expression levels of genes in different organs. Change in expression levels of genes across various organs happen to be the key driving factor towards evolution of closely related primates. In this analysis we are comparing the extra features/functionalities in Human with respect to Chimpanzee via correlation based linear dependency as well as mutual-information based non-linear dependency between the relevant differential genes. In general it is clear from the analysis that maximum number of genes with changed expression level occurs in the TESTIS than in the BRAIN. This result is very surprising because it is our common assumption that most of the changes should occur in brain due to enhanced cognitive abilities and some recent literatures which focus on Transcription Factor (TF) genes of brain with differential expression between Human and Chimpanzee as a key factor of evolution. So, this turns out to be an important finding.**

***Index Terms*—correlation, mutual information, fuzzy c means clustering, kegg pathways**

I. Introduction

Humans and Chimpanzees have originated from a common ancestor. It is known that the evolution of these closely related primates from a common ancestor had occurred 5-7 million years ago [1], [2]. Humans differ from Chimpanzees in a number of important anatomical and physiological respects. So, the most important question is that how Humans are different from their closely related primate Chimpanzee and why? Till now many theories have been proposed [3], [4] where most of them have focused on sequence and structure of genomes specific to some species and depending upon this many conserved functionalities have been explored.

Manuscript received November 25, 2013; revised January 27, 2014.

However, Verki *et al*. [5] has proposed the concept of not only candidate gene analysis but also genome wise analysis to find out evolutionarily, physiologically, and bio medically important differences between two primates

The differences between Human and Chimpanzee results in enhanced cognitive abilities and accordingly there is substantial increase in the size of Human brain. Khaitovich *et al.* [6] had first proposed comparisons for not only in *brain*, but also in other organs namely *heart*, *kidney*, *liver*, and *testis*. Function of Transcription Factors (T. F) in this aspect has been discussed by Nowick *et al.* [7]. Pathways are the key factors to explore different functionalities between the corresponding organs of Human and Chimp because pathways exhibit a form of grouping of different genes meant to explore certain biological functionality. Thus, across closely related primates if the pathways are different for a particular organ the functionalities will also be different. Hence, in search of different pathways in our work we do proceed via linear and non-linear dependencies across differential genes and have shown that for some organs linear analysis shows significant pathways while for others non-linear half becomes far more significant. Below, a brief description of the two important terminologies in this aspect (namely correlation and mutual information) is being given.

A. Correlation

The most familiar linear measure of dependence between two quantities is the Pearson product-moment correlation coefficient, or "Pearson's correlation". The population correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is defined as:

$$\rho_{XY} = corr(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$

where E is the expected value operator, *cov* means covariance, and, *corr* a widely used alternative notation for Pearson's correlation.

doi: 10.12720/jomb.4.1.12-18

If we have a series of n measurements of X and Y written as x_i and y_i where $i = 1, 2, ..., n$, then the *sample correlation coefficient* can be used to estimate the population Pearson correlation r between X and Y. The sample correlation coefficient is written as

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2]}} \quad (2)$$

B. Mutual Information

Formally, the mutual information of two discrete random variables X and Y is:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) \quad (3)$$

where $p(x, y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively.

In the case of continuous random variables, the summation is replaced by a definite double integral:

$$I(X;Y) = \int_{Y} \int_{X} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) dx dy \quad (4)$$

where $p(x, y)$ is now the joint probability *density* function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability density functions of X and Y respectively. Many estimators are available for the calculation of Mutual Information [8]. A detailed description and a comparative study of their performances are given in [9], [10]. Here we have applied *Parametric Gaussian Estimator* [11], [12] where the Mutual Information between two variables X and Y is calculated by

$$I(X, Y) = \left(1/2\right) \log\left(\frac{\sigma_X^2 \sigma_Y^2}{|C|}\right) \quad (5)$$

Here σ_X^2 and σ_Y^2 are the variance of X and Y respectively and $|C|$ is the determinant of covariance matrix. Other methods described in different literatures [9], [10] (like Miller Madow, Shrinkage, Schurmann-Grassberger) does not give any noticeable difference in the performance [13] and hence we selected parametric gaussian estimator, the fastest one for our simulation.

The rest of the paper is as follows. In next section we have discussed about the *Methodology*. Then in *Results and Comparison* section pathways from two different methods are compared. In this section we have also given functionalities (how they work differently between Human and Chimps) of some significant pathways in the corresponding organs. The paper concludes with *Conclusion and Future work*.

II. Methodology

Our dataset contains expression profiles of 6 Humans and 5 Chimps. For this analysis tissues are collected from individuals who had suffered sudden death. This is a whole genome base data where in total 21,000 transcripts were measured by 51,460 probe sets. Details about the experimental settings are available in [14]. All primary expression data are publically available at (http://www.ebi.ac.uk/arrayexpress/experiments) with accession number E-AFMX-11.

We have the sample of 5 tissues (brain, heart, kidney, liver, testis) from 6 Humans and 5 Chimps [14]. The calculation of gene wise k_s, k_i, k_a is performed from the expression data obtained from the Affymetrix arrays using RefSeq-identifiers [15] given in Affymetrix annotation tables [16]. There were in total 11,781 genes, k_i is an estimate of the substitution rate of a genomic region that avoids the large sampling errors, k_s is the substitution rate per synonymous site when closely related primates were compared [17], k_a is the promoter divergence [18]. Measurement of expression divergence for each tissue and each gene were associated with corresponding k_a/k_i ratio and the impact of the ratio on expression divergence were calculated by R package ANOVA [19]. This ratio is a significant predictor of expression divergence having RSS (Residual Sum of Square) = 0.0023 and $p<10^{-06}$.

In any organ between two species, if a gene shows expression divergence then against the gene name for that particular organ a 1 is assigned otherwise a 0 is assigned.

In this dataset, first we pick up only those genes which do not give any expression change in any tissue of Chimp with respect to Human (corresponding entries are 0). Next, we pick up those genes which give an expression change only in brain (not in remaining 4 tissues) of Humans with respect to Chimps (corresponding entry is 1). We repeat the same for the remaining 4 tissues. After making such an ensemble of genes we first do linear clustering. While simulating we have implemented through a form of partition clustering (Fuzzy C Means (FCM) [20]) instead of hierarchical clustering.

Due to the following reasons we have chosen FCM over other methods:

1) When this clustering operation is performed then across iterations the association between two genes can change (taking other genes as a background). Hierarchical clustering is unable to detect this unique property where if two genes get associated in any iteration then for the remaining iterations they will be associated forever, but partition clustering checks the necessity of grouping for each and every iteration and if in some stage it finds that any gene is not suitable with respect to other genes present in a cluster then it will omit that one. On the other hand, if it finds that a gene has close association with other genes present in any cluster then the relevant gene is added as a member of that cluster.

2) Now in partition clustering, we are using FCM instead of simple K-means because the association of each gene over all the clusters can be judged only through FCM, whereas the K-means counterpart can detect the

association of each gene for an individual cluster and assign the same to that particular cluster only.

Let a gene X show same level of (strong) association over two clusters A and B having different functionalities. So FCM algorithm will assign X in both A and B, suggesting that X has possibly got similar impact over the functionalities described by A and B, but simple K-means will assign X in either A or B, neglecting the effect of X on the functionalities described by the other. Here, at first clustering is governed by Pearson correlation based gene to gene dependency.

The optimum number of clusters is found utilizing the concept of Xie-Beni [20] index which is meant for judging the quality of fuzzy clustering. It may be expressed as:

$$XB=\frac{\sum_{i=1}^{k}\sum_{j=1}^{n}u_{ij}^{2}\left\|\vec{Z}_j-\vec{m}_i\right\|^2}{n\times\min_{i\neq j}\left\|\vec{m}_i-\vec{m}_j\right\|^2} \quad (6)$$

The numerator of the Xie-Beni index is basically the objective function minimized by the FCM algorithm. The denominator of *XB* is basically the separation between the *i*-th and *j*-th clusters. The more separated the clusters are, the larger is the denominator and hence the smaller is the value of *XB*. Thus, the optimal number of clusters can be obtained by minimizing the *XB* value.

For genes which do not undergo any change in any tissue we are getting 25 numbers of clusters. Number of optimum clusters for genes with differential expression (only) in brain, heart, kidney, liver and testis are 4, 8, 8, 6 and 20 respectively.

Next, we go for information theoretic Fuzzy C Means clustering through mutual information (henceforth to be designated as MI) based dependency between every pair of genes. For genes which do not undergo any change in any tissue we are getting 28 numbers of Xie-Beni optimised clusters. Whereas, the optimal number of clusters for differential genes (only) in brain, heart, kidney, liver and testis are 4, 10, 10, 4 and 25 respectively.

A. Pre-processing

Before going for the above mentioned analysis we are replacing 0 by a very small value of (10^-30) and missing values are imputed by *KNNimpute* [21] method.

The KNN-based method selects genes with expression profiles similar to the gene of interest to impute missing values. Let us consider gene X that has one missing value in experiment 1, this method would find K other genes, which have a value present in experiment 1, with expression most similar to X in the remaining experiments. A weighted average of values in experiment 1 from the K closest genes is then used as an estimate for the missing value in gene X.

III. Results and Comparison

Cluster significance is found by the number of significant KEGG pathways [22]. If a pathway has 'p-value' ≤ 0.05 and at least 3 genes of any organism are associated with that pathway then the pathway is considered to be significant. Here, at first we are searching for the number of significant pathways with respect to differentially expressed genes, i.e. genes which are mostly responsible for the difference between Humans and Chimps in the various organs.

In Pearson correlation based clustering number of significant pathways for genes changed in brain only equals 8 & for MI based clustering it equals to 9. Similarly, for heart this becomes equal to 17 and 6, for kidney 10 and 5, for liver 6 and 7and finally for testis 47 and 38 respectively.

After getting the pathways, we observed that not only the number of significant pathways in brain and heart are higher in linear correlative than the MI based approach, but also the p-value significance of the pathways are better in the linear half. However, it is just the reverse for liver and testis where p-value significance as well as total number of participating genes in significant pathways are better through MI based analysis (Table I and II). The result is somehow moderate in kidney.

TABLE I. Significant KEGG Pathways of Different Organs by Linear Correlative Method

Region	Pathways	'p-value'	Genes
Brain	Propanoate metabolism	0.000144	4
	Pyruvate metabolism	0.000331	3
	Valine, leucine and isoleucine degradation	0.000394	4
	Metabolic pathways	0.002333	13
Heart	Propanoate metabolism	2.79E-07	5
	Valine, leucine and isolucine degradation	1.59E-06	5
	Adherens junction	0.000164	4
	beta-Alanine metabolism	0.000166	3
	Fatty acid metabolism	0.000873	3
	Metabolic pathways	0.001734	8
	Glycolysis Gluconeogene-sis	0.002067	3
	Oocyte meiosis	0.003091	4
	Pyruvate metabolism	0.004176	4
Kidney	Metabolic pathways	0.002311	10
	Alzheimer's disease	0.003808	5
Liver	Metabolic pathways	0.001132	9
Testis	Metabolic pathways	0.000249	17
	Huntington's disease	0.000405	5
	Propanoate metabolism	0.000562	4
	Oxidative phosphorylation	0.001362	5
	Parkinson's disease	0.001362	5
	Focal adhesion	0.001518	6
	Endocytosis	0.001581	3
	O-Glycan biosynthesis	0.002119	3
	Peroxisome	0.002472	3
	Citrate cycle (TCA cycle)	0.003541	3
	Alzheimer's disease	0.004148	5
	Glutathione metabolism	0.004148	3
	Valine, leucine and isoleucine degradation	0.004703	3
	Regulation of actin cytoskeleton	0.005807	4

TABLE II. SIGNIFICANT KEGG PATHWAYS OF DIFFERENT ORGANS BY MI METHOD

Region	Pathways	'p-value'	Genes
Brain	Pyruvate metabolism	0.0001666	3
Heart	Propanoate metabolism	0.000238	3
	Valine, leucine and isoleucine degradation	0.00048	3
	Adherens junction	0.001186	3
	Metabolic pathways	0.003312	8
	Regulation of actin cytoskeleton	0.008891	3
Kidney	TGF-beta signaling pathway	0.001471	5
	Oxidative phosphorylation	0.003848	6
	Metabolic pathways	0.005802	8
Liver	Ubiquitin mediated proteolysis	0.001284	4
	Tight junction	0.004688	5
	Ribosome	0.005688	4
	Metabolic pathways	0.00896	13
Testis	O-Glycan biosynthesis	2.56E-05	3
	Notch signalling pathway	0.002167	3
	Ubiquitin mediated proteolysis	0.002348	5
	Glutathione metabolism	0.00288	3
	Metabolic pathways	0.003646	16
	Insulin signalling pathways	0.004276	5
	Oocyte meiosis	0.005206	4
	Tight junction	0.006313	4
	Pyruvate metabolism	0.006996	3
	Wnt signalling pathway	0.007161	4
	Long-term depression	0.0087	3
	Alzheimer's disease	0.0093	4
	Huntington's disease	0.0093	4
	Long-term potentiation	0.0093	3
	Neurotrophin signalling pathway	0.0093	4
	Renal cell carcinoma	0.0093	3
	Ribosome	0.0098	3
	Adherens junction	0.0098	3
	Melanogenesis	0.01	3
	Parkinson's disease	0.01	3

B. Biological Importance of some Significant Pathways

In this section, we will discuss about some significant pathways which were detected by linear correlative as well as mutual information based methodologies. These pathways are assumed to be those which are primarily responsible for the change between Humans and Chimpanzees in the five tissues. Thus, here we will try to focus on the different functionalities of these pathways in Humans and Chimps. Pathways are discussed according to those tissues where they are significant. If any pathway is repeated in any organ (according to the outcome of clustering) by any method then we take the most significant one. Unfortunately, for some pathways we did not find any significant difference published till date.

1) Propanate metabolism: Found significant in:

- Brain (By linear correlative method with 'p-value' of 1.44E-04 having 4 genes)
- Heart (By linear correlative and MI method with 'p-value' of 2.79E-07 and 2.38E-04 having 5 and 3 genes)
- Testis (By linear correlative method with 'p-value' of 5.62E-04 having 3 genes)

Function: The metabolism of propionic acid (propanoate) begins with its conversion to propionyl coenzyme A (propionyl-CoA), which is the usual first step in the metabolism of carboxylic acids.

2) Pyruvate metabolism: Found significant in:

- Brain (By linear correlative method and MI method with 'p-value' of 3.3E-04 and 1.67E-04 both having 3 genes)
- Heart (By linear correlative method with 'p-value' of 4.17E-03 having 4 genes)
- Testis (By MI method with 'p-value' of 6.99E-03 having 3 genes)

Function: Pyruvate is a constituent of all the media used for Human in vitro fertilization (IVF), and it promotes the development of fertilized Human oocytes (zygotes) to blastocysts in culture.

Whereas in Chimps it plays a central role in balancing the energy needs of various tissues in the body under conditions in which oxygen supply is limited.

3) Metabolic pathways: Found significant in:

- Brain (By linear correlative method with 'p-value' of 2.33E-04 having 13 genes)
- Heart (By linear correlative method and MI method with 'p-value' of 1.73E-03 and 3.31E-03 respectively both having 8 genes)
- Kidney (By linear correlative method and MI method with 'p-value' of 2.31E-03 and 5.8E-03 respectively having 10 and 8 genes)
- Liver (By linear correlative method and MI method with 'p-value' of 1.1E-03 and 8.9E-03 respectively having 9 and 13 genes)
- Testis (By linear correlative method and MI method with 'p-value' of 2.49E-04 and 3.64E-03 respectively having 17 and 16 genes)

Function: In Humans mostly we see protein metabolism which determines and transmits amino acids, followed by conversion of the non-nitrogenous part of those molecules to glucose or lipids and helps in the removal of ammonia from the body by the synthesis of urea.

In Chimps we mainly observe carbohydrate metabolism which maintains concentration of glucose in blood within a narrow, normal range by glycogenolysis and gluconeogenesis.

4) Oxydative phosphorylation: Found significant in

- Kidney (By MI method with 'p-value' of 3.84E-03 having 6 genes)
- Testis (By linear correlative method with 'p-value' of 1.36E-03 having 5 genes)

Function: Oxidative phosphorylation (or OXPHOS in short) is the metabolic pathway in which the mitochondria in cells use their structure, enzymes, and energy released by the oxidation of nutrients to reform ATP. It is functional in MITOCONDRIAL DNA (mtDNA).

From [23] we can say that "*Oxygen* consumption, a sensitive index of respiratory function, showed that mtDNA from Chimpanzee and Pigmy Chimpanzee

replaced the Human mtDNA and restored respiration to essentially normal levels."

5) *TGF-beta signalling pathway:* Found significant in
- Kidney (By MI method with 'p-value' of 1.47E-03 having 5 genes)

Function: The transforming growth factor-ss (TGF-beta) signalling pathway plays a pivotal role in diverse cellular processes. TGF-beta switches its role from a tumour suppressor in normal or dysplastic cells to a tumour promoter in advanced cancers. TGF-beta signalling has been considered a useful therapeutic target.

6) *Tight junction(s):* Found significant in
- Liver (By MI method with 'p-value' of 4.68E-03 having 5 genes)
- Testis (By MI method with 'p-value' of 6.31E-03 having 4 genes)

Function: In Humans tight junctions are intercellular junctions adjacent to the apical end of the lateral membrane surface. They have two functions, the barrier (or gate) function and the fence function. The barrier function of tight junctions regulates the passage of ions, water, and various macromolecules, even of cancer cells, through paracellular spaces. The barrier function is thus relevant to edema, jaundice, diarrhea, and blood-borne metastasis. The fence function maintains cell polarity.

From [24] we can say that Zonula Occludens protein (ZO-1) which is a Tight Junction Protein is associated associated with different cellular functions with experimentally validated splicing level differences between Humans and Chimpanzees.

7) *Ubiquitin mediated proteolysis:* Found significant in:
- Liver (By MI method with 'p-value' of 1.28E-03 having 4 genes)
- Testis (By MI method with 'p-value' of 2.34E-03 having 5 genes)

Function: In Humans it performs mainly transcription regulation, DNA repair, receptor modulation, immune response, signal transduction and quality control. Whereas in Chimps protein ubiquitination mainly functions as a signal for 26S proteasome dependent protein degradation. The addition of ubiquitin to proteins being degraded is performed by a reaction cascade consisting of three enzymes, named E1 (ubiquitin activating enzyme), E2 (ubiquitin conjugating enzyme), and E3 (ubiquitin ligase). Each E3 has got specificity to its substrate, or proteins to be targeted by ubiquitination.

8) *Insuline signalling pathway:* Found significant in:
- Testis (By MI method with 'p-value' of 4.27E-03 having 5 genes)

Function: In Humans the insulin transduction pathway is an important biochemical pathway beginning at the cellular level affecting homeostasis. Signalling through the insulin/IGF (Insulin growth factor) 1 - like receptor pathway is a significant contributor to the biological aging process.

The main difference (between Human and Chimp related to this pathway) is due to FOXO1 transcription factors. It is the key target of the insulin/IGF signaling pathway. In [25], [26] it is stated that FOXO1 expression is prominent in Human and Chimpanzee which is evolved to serve divergent functions in male and female gonad: such as germline in testis.

9) *O-Glycan biosynthesis:* Found significant in:
- Testis (By linear correlative method and MI method with 'p-value' of 2.12E-03 and 2.56E-05 both having 3 genes)
- Testis (By linear correlative method and MI method with 'p-value' of 2.12E-03 and 2.56E-05 both having 3 genes)

Function: The most common O-linked glycans are the mucin-type glycans. The mucin-type ***O***-glycans have several cancer-associated structures, including the T and Tn antigens, and certain Lewis antigens.

10) *Oocyte meosis:* Found significant in:
- Heart (By linear correlative method with 'p-value' of 3.09E-03 having 4 genes)
- Testis (By MI method with 'p-value' of 5.26E-03 having 4 genes)

Function: In Humans the quality of oocytes plays a key role in a proper embryo development. In Humans, oocytes of poor quality may be the cause of women infertility and an important obstacle in successful i*n vitro* fertilization (IVF).

In Chimps women with patterns of high oocyte loss experience earlier menopause. Chimpanzees in captivity live longer, and thus, similar to Humans, they may experience follicular depletion that precedes death by many years

Apart from the above mentioned pathways some other pathways are also found to be significant. These are:

11) *Valine leucine and isoleucine degradation:* Found significant in:
- Brain (By liner correlative method with 'p-value' of 3.94E-04 having 4 genes)
- Heart (By Linear correlative as well as by MI method with 'p-value' of 1.59E-06 and 4.8E-04 having 5 and 3 genes)
- Testis (By Liner correlative method with 'p-value' of 4.73E-03 having 3 genes)

12) *Adherens junction:* Found significant in
- Heart (By linear correlative method and MI method with 'p-value' of 1.64E-04 and 1.18E-02 having 4 and 3 genes)
- Testis (By MI method with 'p-value' of 9.8E-02 with 3 genes)

13) *Parkinson's disease:* Found significant in
- Testis (By linear correlative method and MI method with 'p-value' of 1.36E-03 and 1E-02 having 5 and 3 genes)

In our work we have found that: *cox6c, cox7b, ndufa2, ndufc2, ndufs3, ube2g2, ndufs4* genes are responsible for Parkinson's disease.

Their functionalities and impact on Parkinson's disease have been given in [25].

14) *Alzheimer's disease:* Found significant in
- Kidney (By linear correlative method with 'p-value' of 3.81E-03 having 5 genes)

- Testis (By linear correlative method and MI method with 'p-value' of 4.15E-3 and 9.3E-03 having 5 and 4 genes)

We have found that *atp5g3, calm1, ndufa5, ndufb2, ppp3r1, cox6c, cox7b, apoe, ndufb7, ndufc1, ndufs4, psen2* genes are responsible for Alzheimer's disease (A.D).

Roll of the above mentioned genes in A.D have been discussed in [27], [28]

15) Huntington's disease Found significant in

- Testis (By linear correlative method and MI method with 'p-value' of 4.05E-04 and 9.3E-03 having 5 and 4 genes)

In our work we have found that: *ndufb1, ndufb4, sod2, polr2k, ap2a2, polr2l, ndufs4* genes are responsible for Huntington's disease.

This result has been validated by [29], [30] which give the functionalities of some above mentioned genes in Huntington's disease.

Overall Functionalities of all the above mentioned pathways and genes associated with them can be found in [31].

IV. Conclusion and Future Work

In this work, we focused on significant pathways for differentially expressed genes in the five tissues and the novelty lies through non-linear MI based and linear correlative analysis being conducted via fuzzy clustering. By comparing them we found some new pathways in the different tissues via MI based non linear analysis.

Our work has shown that genes associated in some disease related pathways may be associated in some physiological functions under normal conditions too. The probable answer of this dual nature can be found in differential coexpression analysis [32] which may be caused from different biological conditions indicating rewiring of transcriptional networks in response to disease or adaptation to different environments. For example, a group of genes may be under the control of a common regulator (a transcription factor or a epigenetic modification) that is active in one condition but absent in the other.

As given in [33] differential network analysis reveals differences in connectivity and module structure between two networks. Functional annotation of these genes may lead to different biological pathways, having some of them involved in pathogenesis.

As a future work we can check and compare the coexpression effect of differentially expressed genes under different conditions. Let there be two conditions. Then for first condition we can check the coexpression between differentially expressed genes. Next, we need to go for biological validation of the result. The same procedure can be repeated for the second condition. Finally, comparing the results (between two conditions) this can give us an idea about the effects of conditional dependency on differential expression. Here, we can also go for some sort of an untargeted approach, where not only the within module, but module to module differential coexpression will also be checked to get more precise condition specific correlated differential genes/modules.

We can extend the analysis to searching for genes showing topological differences between two conditions which can further enlighten the key differences between two species.

References

[1] M. Brunet, *et al.*, "A new hominid from the Upper Miocene of Chad, Central Africa," *Nature* 418, pp. 145-151, July 2002.

[2] F. C. Chen and W. H. Li, "Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees," *American Journal of Human Genetics*, vol. 68, no. 2, pp. 444-456, January 2001.

[3] M. Goodman, "The genomic record of humankind's evolutionary roots," *American Journal of Human Genetics*, vol. 64, no. 1, pp. 31-39, January 1999.

[4] A. Varki and D. L. Nelson, "Genomic comparisons of human and chimpanzees," *Advanced Research Journal*, vol. 36, pp. 191-209, September 2007.

[5] A. Varki and T. K. Altheide, "Comparing the human and chimpanzee genomes: Searching for needles in a haystack," *Genome Research*, pp. 1746-1758, December 2005.

[6] P. Khaitovich, *et al.*, "Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees," *Science*, vol. 309, pp. 1850-1854, September 2005.

[7] K. Nowick, T. Gernat, E. Almaas, and L. Stubbs, "Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain," in *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 52, October 2009, pp. 22358-22363.

[8] A. Kraskov, H. Stogbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, June 2004.

[9] S. Khan, S. Bandyopadhyay, A. Ganguly, and S. Saigal, "Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data," *Physical Review E*, August 2007.

[10] C. J. Cellucci, A. M. Albano, and P. E. Rapp, "Statistical validation of Mutual Information calculations: Comparison of alternative numerical algorithms," *Physical Review E. Statistical Nonlinear and Soft matter Physics*, June 2005.

[11] A. A Margolin, *et al.*, "ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," *BMC Bioinformatics*, March 2006.

[12] C. Olsen, P. E. Meyer, and G. Bontempi, "On the impact of entropy estimation on transcriptional regulatory network inference based on mutual information," *EURASIP Journal on Bioinformatics and System Biology*, 2009.

[13] G. Altay and F. E-Streib, "Inferring the conservative causal core of gene regulatory networks," *BMC System Biology*, vol. 4, no. 132, 2010.

[14] [Online]. Available: http://www.ncbi.nlm.nih.gov/sites

[15] K. D. Pruitt and D. R. Maglott, "RefSeq and locuslink: NCBI gene-centered resources," *Nucleic Acids Res* vol. 29, no. 1, pp. 137-140, Jan 2001.

[16] G. Liu, *et al.*, "NetAffx: Affymatrix probesets and annotations," *Nucleic Acids Res*, vol. 31, no. 1, pp. 82-86, 2003.

[17] M. Hopkin, "Bonobos face extinction," *The Chimpanzee Sequencing Consortium: Nature*, December 2004.

[18] I. Tirosh, A. Weinberger, D. Bezalel, M. Kaganovich, and N. Barkal, "On the relation between promoter divergence and gene expression evolution," *Molecular System Biology*, vol. 4, no. 159, 2008.

[19] J. J. Faraway, *Practical Regression and ANOVA using R*, cran.r-project, July 2002.

[20] S. Das, A. Abraham, and A. Konar, *Metaheuristic Clustering 'Studies in Computational Intelligence*, 1st ed., vol. 178, Springer, ch.1,pp. 12-21.

[21] O. Troyanskaya, *et al.*, "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520-525, February 2001.

[22] Gene Trail. [Online]. Available: http://genetrail.bioinf.uni-sb.de/

[23] L. Kenyon and C. T. Moreas, "Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids," in *Proc. Natl. Acad. Sci. USA*, vol. 94, no. 17, June 1997, pp. 9131-9135.
[24] J. A. Calarco, *et al.*, "Global analysis of alternative splicing differences between humans and chimpanzees," *Genes Dev*, *vol.* 21, no. 22, pp. 2963-2975, November 2007.
[25] E. D. Taranawa, M. D. Baker, G. Aloisio, B. R. Carr, and D. H. Castrillion, "Gonadal expression of Foxo1, but not Foxo3, is conserved in diverse mammalian species: Society for the study of reproduction," *BOR Papers in Press*, March 2013.
[26] P. D. C. Mail, R. Blekhman, A. E. Chabot, A. Oshlack, and Y. Gilad, "A combination of genomic approaches reveals the role of foxo1a in regulating an oxidative stress response pathway," *PLoS ONE*, vol. 3, no. 2, e1670, February 2008.
[27] H. Michael, J. Behr, A. Harrenga, and A. Kannt, "Cytochrome C oxidase: Structure and spectroscopy," *Annual Review of Biomol. Struct*, vol. 27, pp. 329-356, 1998.
[28] L. Bojarski , J. Herms, and J. Kuznicki, "Calcium dysregulation in alzheimer's disease," presented in Neurochem. Int., Mar-Apr 52(4, 5), 2008, pp. 621-633.
[29] M. Warner, P. Thuriaux, and J. Soutourina, "Structure-function analysis of RNA polymerases I and III," *Curr. Opin. Struct. Biol*., vol. 19, no. 6, pp. 740-45, December 2009.
[30] G. Scita and P. P. D. Fiore, "The endocytic matrix," *Nature* 463, pp. 464-73, January 2010.
[31] [Online]. Available: http://www.genome.jp/dbget-bin
[32] B. M. Tesson, R. Breitling, and R. C. Jasncen, "DiffCoEx: A simple and sensitive method to find differentially coexpressed gene modules," *BMC Bioinformatics*, vol. 11, no. 497, 2010.
[33] T. F. Fuller, A. Ghazalpour, J. E.Aten, T. A. Drake, A. J. Lusis, and S. Horvath, "Weighted gene co expression network analysis strategies applied to mouse weight," *Mamm Genome*, vol. 18, pp. 463-462, 2007.

Aurpan Majumder had obtained his B.Tech in 2002 in Electronics & Telecommunication Engg. from Haldia Institute of Technology, India and M.Tech in 2005 in Computer Science from Jadavpur University, India. Currently he is working as an Assistant Professor at National Institute of Technology, Durgapur, India. He is pursuing PhD in the domain of gene differential coexpression analysis.

Mrityunjay Sarkar had obtained his B.Tech from Asansol Engg. College, INDIA in 2009 and M.Tech from N.I.T durgapur, INDIA in Telecommunication Engg.

