ISSN 1750-9823 (print)
International Journal of Sports Science and Engineering
Vol. 01 (2007) No. 03, pp. 147-156

Monitoring Athletes' Physiological Responses to Endurance Training with Genomic-wide Expression Data

Ilene Chen [+ 1, 3, *], Ling-Hong Tseng [2,*], Hiroto Homma [3], Hong Yan [1, 4], L. Lloyd Keith [5]

[1] School of Electrical and Information Engineering, University of Sydney, Australia
[2] Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital and University of Chang Gung School of Medicine, Taiwan
[3] North Sydney Aussie Swimming Inc., Australia
[4] Department of Electronic Engineering, City University of Hong Kong, Hong Kong
[5] Division of Urology, University of Alabama School of Medicine, Birmingham, Alabama, USA

(*Received August 12 2007, accepted September 2 2007*)

Abstract. A system of fixed effect regression modeling for genome-wide expression data from DNA microarray hybridization is described that uses statistical methods in longitudinal or matched case-control data analysis to monitor athletes' physiological responses to endurance training. We first identify significantly differential expressed genes with endurance training-induced muscle contraction using fixed effect regression modeling, which effectively adapts the confounding effects arising from the interaction between genes. Next, we map key genes onto acknowledged KEGG pathways to attain a linkage between key molecules and biochemical pathways with endurance training-induced muscle contraction in a cause-effect format. To demonstrate this approach, we have used fixed effect logistic regression modeling to study a gene expression model relating to endurance training-induced vastus lateralis muscle contraction. We have found the development of carbohydrate, lipid and energy metabolisms, respectively, the transcriptional regulations of endurance training-induced vastus lateralis muscle contraction status, and the presence of the deleterious effects of oxygen from the metabolic reduction of the reactive oxygen species. The approach described here can supply general tools to monitor athletes' physiological responses to endurance training on the genomic scale.

Key words: fixed effect regression modeling; muscle contraction; metabolism; transcriptional regulation; physiological response

1. Background

Endurance training was initially developed by Dr. Alois Mader in the 1980s [1-4], and it has had substantial influence on the training of athletes in all endurance sports [5]. Coaches design training programs around the idea of pushing athletes to the limit of their pain tolerance and then motivating them to go beyond it [5]. The rapid advance of genome-scale sequencing has led us to investigate mechanisms which determine athletes' physiological nature of stress to endurance training. Scientists have worked closely to observe gene expression patterns relating to endurance training-induced muscle contraction with the aim of monitoring athletes' physiological conditions with genome-wide expression data from DNA microarray hybridization. But their attempts were unsatisfactory as the results they presented were far more disconnected with any formal theory of the behavior of lively organisms and their parts. We need to develop some fundamental tools to monitor athletes' physiological responses to endurance training on this genomic scale.

Because genes and gene products, proteins, do not function independently, but participate in complex, interconnected pathways, regulatory networks and molecular systems, taken together, give rise to the working of lively processes [6], we need to adapt the confounding effects arising from the interaction between genes. With this fundamental theoretical demand in mind, we have recently been led to a system of fixed effect regression modeling, compatible with all known longitudinal or matched case-control data

[+] Ilene CHEN and Ling-Hong TSENG contributed equally in this paper.
Correspondence should be addressed to Ilene Chen ilene@ee.usyd.edu.au TEL: +61 2 93517221 FAX: +61 2 93513847

analysis, and free of arbitrary effects due to the role of shifts in the expression of genes.

In seeking the mechanisms of the physiological nature of stress, a natural basis is to identify significantly differential expressed genes in response to endurance training-induced muscle contraction by adapting the confounding effects arising from the interaction between genes that all muscle cells have in common. To do this, the first step to this end is to adopt a statistical modeling technique to find significantly differential expressed genes that result in or turn off the expression that are actively producing messenger RNA for biosynthesis of their particular proteins. For any gene expression array in multiple samples or in a time series after a given endurance training course, several statistical modeling techniques can be used such as the fixed effect logistic regression, fixed effect probit regression, or fixed effect complementary log-log regression. We have found here that fixed effect logistic regression conforms well to the insightful mechanisms into the physical nature of stress. This may be so since this technique quantifies the expression patterns within each comparable criterion primarily, and judges the effects on expression patterns among various comparable criteria subsequently.

The purpose of this paper is not to survey various types of muscle cells available to compare the resultant physiological states on the basis of their expression patterns, but rather to show how such methods can usefully monitor the physiological conditions in response to endurance training-induced muscle contraction. While several statistical modeling techniques can be used to identify significantly differential expressed genes in response to endurance training-induced muscle contraction, the resultant massive collection of genes remains hard to assimilate. Consequently, we always map them onto acknowledged KEGG pathways representing our current knowledge in metabolisms, genetic information processing, environment information processing, cellular processes and human diseases [7]. Particularly, the pathways of signal transduction represent the processes of converting external signals such as neuropeptides, hormones, cytokines and growth factors to specific internal processes such as gene expression, cell division, or even cell suicide [8]. Since all types of muscle contraction require the transmission of information from the outer membrane of a muscle cell into its nucleus, it would meet with a great deal of enthusiasm to recognize the physiological nature of stress by mapping significantly differential expressed genes onto acknowledged KEGG pathways based on the knowledge of molecular components.

To demonstrate this approach, we have used fixed effect logistic regression modeling to re-evaluate the gene expression model relating to endurance training-induced vastus lateralis muscle contraction by Fluck et al. [9], and have found the underlying physiological responses with endurance training. Due to the role of shifts in the expression of genes, Fluck et al. [9] were unable to derive any insight into the underlying physiological processes in response to stress, and in their conclusion doubted the usefulness of DNA microarray experiments. But the nature of the difficulty appears to lie in the inadequacy of available statistical machinery to “see” the information in gene expression data. The following analysis aims to add particular worth in fixed effect regression modeling with the confidence that greater generality can be achieved in the analysis of endurance training-induced physiological nature of stress at the molecular level.

2. Methods

Source of Experimental Data. The data used here was collected from Gene Expression Omnibus under platform accession no GDS1432; its web address is http://www.ncbi.nlm.nih.gov/geo/gds/gds_browse.cgi?gds=1432. In this analysis, we re-evaluate the gene expression based RNA/sported DNA/cDNA array by Fluck et al. [9] on vastus lateralis muscles before and after a 6-week endurance training course, respectively. Expression patterns of vastus lateralis muscles were examined repeatedly 0, 1, 8 and 24 hours following a 30-minute endurance training course. A detailed description relating to its experimental procedure can be found from Fluck et al. [9].

Identifying significantly differential expressed genes with muscle contraction using fixed effect logistic regression.

Providing expression of genes from a case-control design, where the observations on expression of genes were recorded repeatedly within and among various comparable criteria, respectively, such as, time, place, blood pressure, genotype/variation, etc. Let D and $\bar{D}$ stand for the presence and absence of muscle contraction, respectively. Suppose that the probability that the observations corresponding to the *i*-th subject (e.g. a mouse, or a person) were recorded t times among t comparable criteria repeatedly. Suppose that the presence of muscle contraction when comparing with control depends on expression $X_{1it}, \mathrm{K}, X_{pit}$ of p

genes, where $i = 1,2,\mathrm{K}\ ,M$ and $t = 1,2,\mathrm{K}\ ,T$. Without loss of generality, for the *i*-th subject, the structure of the gene expression has the following form:

Levels of gene expression	D t -th measurement	$\bar{D}$ t -th measurement
X_{1it}	$X_{1it} \mid D$	$X_{1it} \mid \bar{D}$
M	M	M
X_{pit}	$X_{pit} \mid D$	$X_{pit} \mid \bar{D}$

Let $\boldsymbol{Y}$ be the vector of joint expression, that is, $\boldsymbol{Y}=\left(X_{1it},\mathrm{K}\ ,X_{pit}\right)$. Let $Pr\left(Y_{it} \mid D\right)$ denote the probability that the *i*-th subject has the muscle contraction at the *t*-th measurement has the vector of joint expression , Y_{it} , where $i = 1,2,\mathrm{K}\ ,M$ and $t = 1,2,\mathrm{K}\ ,T$. Similarly, let $Pr\left(Y_{it} \mid \bar{D}\right)$ denote the probability that the the *i*-th subject is a control at the *t*-th measurement has the vector of joint expression, Y_{it} . It follows that the joint probability that the Y_{it} corresponds to a case and Y_{it} corresponds to a control is

$$Pr\left(Y_{it} \mid D\right)Pr\left(Y_{it} \mid \bar{D}\right) \qquad [1]$$

The total probability for all likely assignments for the subjects with and without the presence of muscle contraction to the vector of gene expression is

$$\sum_t Pr\left(Y_{it} \mid D\right)Pr\left(Y_{it} \mid \bar{D}\right) \qquad [2]$$

It follows that the required condition likelihood for the *i*-th subject at the *t*-th measurement has joint expression is the ratio of the probability given by [1] and [2], namely

$$L = \frac{Pr\left(Y_{it} \mid D\right)Pr\left(Y_{it} \mid \bar{D}\right)}{\sum_t Pr\left(Y_{it} \mid D\right)Pr\left(Y_{it} \mid \bar{D}\right)} \qquad [3]$$

Recall from the Bayes Theorem, $P\left(y \mid d\right) = {P\left(d \mid y\right)P\left(y\right)}\big/{P\left(d\right)}$. It follows that the conditional likelihood in [3], the terms $Pr\left(D\right)$, $Pr\left(\bar{D}\right)$ and $Pr\left(Y_{it}\right)$ cancel out and we are left with

$$L = \frac{Pr\left(D \mid Y_{it}\right)Pr\left(Y_{it}\right)Pr\left(\bar{D} \mid Y_{it}\right)Pr\left(Y_{it}\right)}{\sum_t Pr\left(D \mid Y_{it}\right)Pr\left(Y_{it}\right)Pr\left(\bar{D} \mid Y_{it}\right)Pr\left(Y_{it}\right)} \qquad [4]$$

So the probability that the presence of muscle contraction for the *i*-th subject at the *t*-th measurement with expression $X_{1it},\mathrm{K}\ ,X_{pit}$ of p genes can be expressed as a linear regression model

$$Pr\left(D \mid Y_{it}\right) = \alpha_i + \beta_1 X_{1it} + \mathrm{K} + \beta_p X_{pit} \qquad [5]$$

Supposedly, [5] can be used directly, but the fitted regression coefficients in [5] may run the fitted values in negative values or values greater than unity that would be difficult to assimilate. To tackle this difficulty,

we therefore transform the probability from the range $(0,1)$ to $(-\infty,\infty)$ by fitting a linear logistic regression model where

$$log\frac{Pr(D \mid Y_{it})}{1-Pr(D \mid Y_{it})} = \alpha_i + \beta_1 X_{1it} + \mathrm{K} + \beta_p X_{pit} \quad [6]$$

and

$$Pr(D \mid Y_{it}) = \frac{exp(\alpha_i + \beta_1 X_{1it} + \mathrm{K} + \beta_p X_{pit})}{1+exp(\alpha_i + \beta_1 X_{1it} + \mathrm{K} + \beta_p X_{pit})}$$

Since a subject can be either a case, with the muscle contraction, or a control, without the muscle contraction, the conditional likelihood of α_i, β_1, ⋏ , β_p and $Y_{it} \mid D$ is given by

$$\begin{aligned} L &= L(\alpha_i, \beta_1, \mathrm{K}, \beta_p; Y_{it} \mid D) \\ &= \prod_M \binom{M}{Y_{it} \mid D} Pr(D \mid Y_{it})^{Y_{it} \mid D} (1 - Pr(D \mid Y_{it}))^{M - Y_{it} \mid D} \\ &= \prod_M \binom{M}{Y_{it} \mid D} \frac{(exp(\alpha_i + \beta_1 X_{1it} + \mathrm{K} + \beta_p X_{pit}))^{Y_{it} \mid D}}{(1+exp(\alpha_i + \beta_1 X_{1it} + \mathrm{K} + \beta_p X_{pit}))^{M}} \end{aligned} \quad [7]$$

The conditional log-likelihood is therefore

$$\begin{aligned} l &= log\, L \\ &= \sum_M \{(Y_{it} \mid D)(\alpha_i + \beta_1 X_{1it} + \mathrm{K} + \beta_p X_{pit}) - \\ &\quad M\, log(1 + \alpha_i + \beta_1 X_{1it} + \mathrm{K} + \beta_p X_{pit}) + log\binom{M}{Y_{it} \mid D}\} \end{aligned}$$

In consequence, the maximum likelihood estimates for α_i, β_1, ⋏ , β_p can be worked out using the Newton-Raphson algorithm

$$\boldsymbol{[Br+1]=[Br]-H^{-1}U}$$

where $\boldsymbol{B'}=(\alpha_i, \beta_1, \mathrm{K}, \beta_p)^T$, $\boldsymbol{U}$ is the first derivative vector of the log-likelihood function; i.e. $\boldsymbol{U}=[U_i]=\left[\partial l / \partial \beta_i\right]$; $\boldsymbol{H}$ is Hessian matrix; i.e.

$$\mathrm{H} = \text{[illegible]}_{ij} = \text{[illegible]}\,\partial^2 l / \partial \beta_i \partial \beta_j \,.$$

The proof of the model is now given to verify that the effects on expression of genes are measured within and among various comparable criteria, respectively. For this, let us consider the conditional probability given by [7]. Let $X^*_{1it}, \mathrm{K}, X^*_{pit}$ and $X^{**}_{1it}, \mathrm{K}, X^{**}_{pit}$ refer to the expression of genes in the case that the *i*-th subject with and without the presence of muscle contraction, respectively. While the *i*-th subject at the *t*-th measurement has the presence of muscle contraction, we have

$$\frac{Pr(D \mid Y_{it})}{Pr(\bar{D} \mid Y_{it})} = \alpha_i + \beta_1 X^*_{1it} + \mathrm{K} + \beta_p X^*_{pit}$$

While the *i*-th subject at the *t*-th measurement does not have the presence of muscle contraction, we have

$$\frac{Pr(D \mid Y_{it})}{Pr(\bar{D} \mid Y_{it})} = \alpha_i + \beta_1 X^{**}_{1it} + \mathrm{K} + \beta_p X^{**}_{pit}$$

Therefore, the conditional likelihood in [7] becomes

$$\left(1+\sum_{M} exp\left(\beta_1\left(X^{*}_{1it}-X^{**}_{1it}\right)+\mathrm{K}+\beta_p\left(X^{*}_{pit}-X^{**}_{pit}\right)\right)\right)^{-1}$$

Specifically, the values of $X^{*}_{1it}-X^{**}_{1it}$, ⋯, $X^{*}_{pit}-X^{**}_{pit}$ quantifies the effects on expression recorded within each comparable criteria. Consequently, this conditional likelihood considers the effects on expression within each of the comparable criteria primarily; subsequently, it judges the effects on expression measured among various comparable criteria. In such a method, the process of estimating the parameter ($i=1,\mathrm{K},p$) is termed as fixed effect logistic regression modeling, also yields standard errors of the estimates, which can be standardized and normalized parameters that reflect levels of statistical significance. To identify significantly differential expressed genes associated with the presence of muscle contraction when comparing with the control, separate univariate fixed effect logistic regressions are fitted for each gene corresponding to the *i*-th subject at the *t*-th measurement. There is a coefficient estimator, $\hat{\beta}_i$ ($i=1,\mathrm{K},p$) representing biological importance, its associated standard errors, $\sigma(\hat{\beta}_i)$ ($i=1,\mathrm{K},p$), and its normalized coefficient estimator $\left[\beta_i-\mu(\hat{\beta})\right]\Big/\sigma(\hat{\beta})$ ($i=1,\mathrm{K},p$) representing statistical significance, for each of the p genes. To account the tradeoff between the biological importance and statistical significance, significantly differential expressed genes are prioritized relating to selected combinations of these two criteria.

Mapping significantly differential expressed genes onto biochemical pathways. Based on the resulting significantly differential expressed genes associated with vastus lateralis muscle contraction after a 6-week endurance training course when comparing with its control, before a 6-week endurance training course, identified using fixed effect logistic regression, their associated proteins, conserved domains, biological processes and components were examined using existing NCBI database http://www.ncbi.nlm.nih.gov/. The balancing of biochemical pathways are examined using the public KEGG pathway database http://www.genome.jp/kegg/pathway.html representing our current knowledge in metabolisms, genetic information processing, environmental information processing, cellular processes and human diseases [8].

3. Results

We have used fixed effect logistic regression modeling to re-evaluate the time course gene expression model relating to endurance training-induced vastus lateralis muscle contraction by Fluck et al. [9]. Providing the expression patterns of vastus lateralis muscle cells happened within 0, 1, 8 and 24 hours, respectively, following a 30-minute endurance training course, the case and control were characterized at 6-week before and after undergoing the endurance training course, respectively.

The results were very interesting. First, complex metabolic reactions, including carbohydrate, lipid and energy metabolisms, were found to have strong relation with vastus lateralis muscle contraction. This muscle contraction requires energy, which is made available from two energy rich molecules: ribosomes and transfer-RNA. The ribosomes have sites that adapt transfer-RNA, which recognize molecular components that occur between macromolecular entities in vastus lateralis muscle cells. Next, on the basis of the mapping of the resultant significantly differential expressed genes associated with endurance training-induced vastus lateralis muscle contraction onto the pathways of axon guidance such as the regulation of actin cytoskeleton, Ca2+ signaling transduction, mitogen-activated protein kinase (MAPK) signaling transduction and peroxisome-activated receptor (PPAR) signaling transduction, we are able to identify the transcriptional regulations of endurance training-induced vastus lateralis muscle contraction status including the level of contractile force, tissue strength, plasticity and energy homeostasis. Thus, the mapping of a gene onto the KEGG pathways mimics its actual transcription. A more insightful explanation is provided as follows.

Regulation of the level of contractile force. The contraction of vastus lateralis muscle is supported by the nervous systems. Examining the mapping on the pathways representing the regulations of axon guidance (Fig. 1) and actin cytoskeleton (Fig. 2), respectively, indicates that developing motor neurons extend axons

were found to be guided by Ephrin-B along specific pathways to regulate cytoskeleton dynamics in axonal growth cones in response to contraction stimulated by the assembly of focal adhesions and actin stress fibers in vastus lateralis muscle cells.

Ras proteins of small GTPases domains are strong candidates to transmit guidance signals in the growth cone in response to stress. They regulate muscular cellular processes where filaments actin plays a central role. Both Ras proteins and ATP motor proteins are identified as responsible for the generation of mechanical forces upon the hydrolysis, consisting with their structural similarity of the catalytic region of GTPases and ATPases domains [10]. The force transmission mechanism is based on an irreversible structural change, produced by the hydrolysis, which trigger thermal switching between force-generating substates through change in the configurational space of the proteins [10]. Studies of vertebrate growth cone suggest common mechanisms that regulate growth cone behaviors and axon branching [11]. These include reorganization of actin and microtubules, effect of axon guidance factors, actions of actin regulatory proteins and dynamic changes in intracellular Ca2+ signaling. These guidance systems are contrived to encourage as astonishing varied set of neuronal circuits underlying axon guidance in response to the contraction stimulated by the assembly of focal adhesions and actin stress fibers.

The concentration of cytoplasmic Ca2+ has had been known to regulate the contractile state of skeletal muscle cells and tissues [12]. According to the mapping on the Ca2+ signaling pathway, high-voltage activated Ca2+ channels are found to be regulated by the cyclic AMP second messenger system that ultimately serve to regulate the level of contractile force in the contraction stimulated by the assembly of focal adhesions and actin stress fibers. Besides, Ca2+ is exhibited to assemble various signaling components, including G proteins, thyroid hormones, growth factors and fibroblast growth factors through vastus lateralis muscle contraction status. These signaling components, produced by pituitary gland, implicated that the central nervous system regulates the process of molecular recognition through vastus lateralis muscle contraction status.

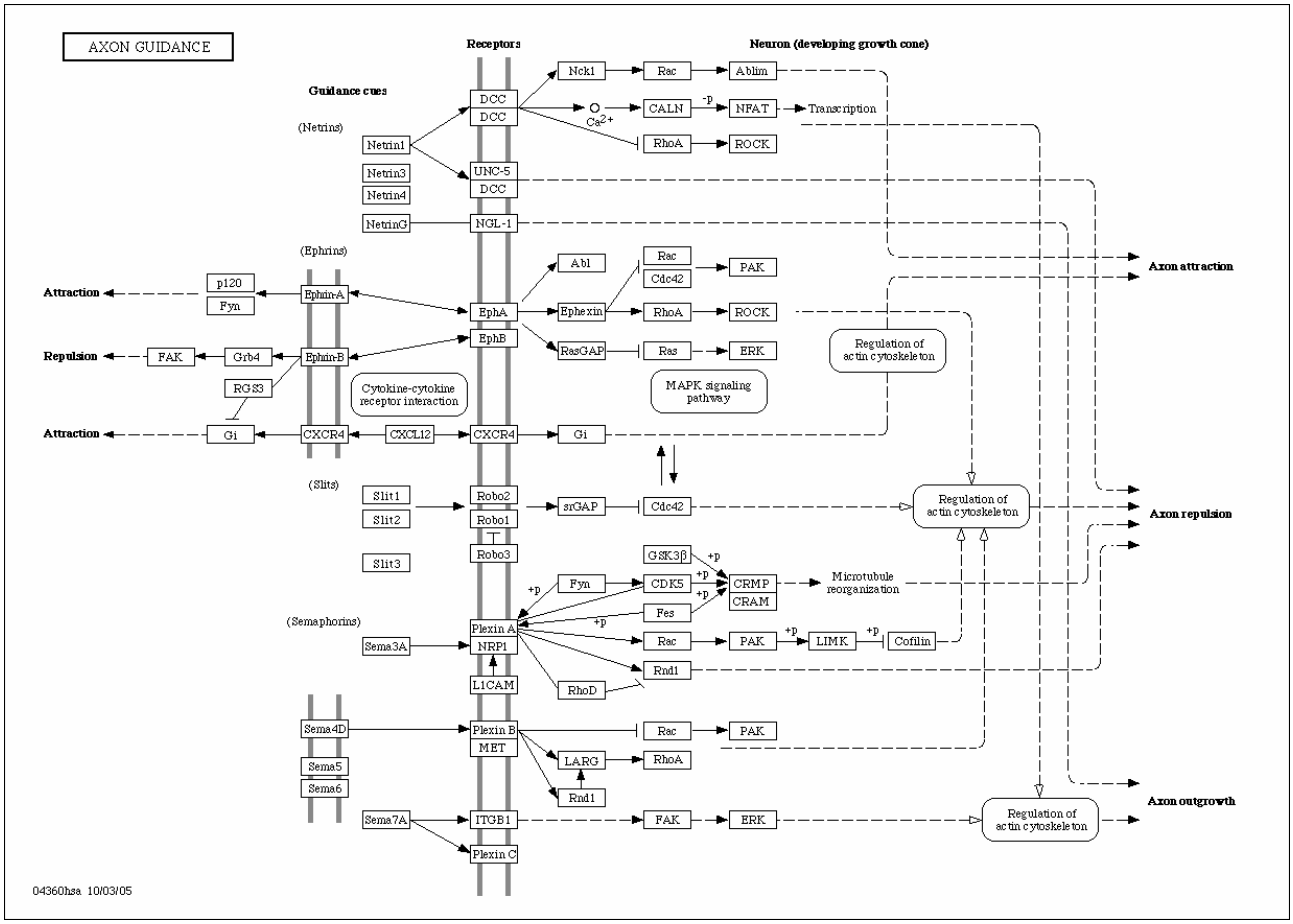

Fig. 1. Reference KEGG pathway representing the regulation of axon guidance.

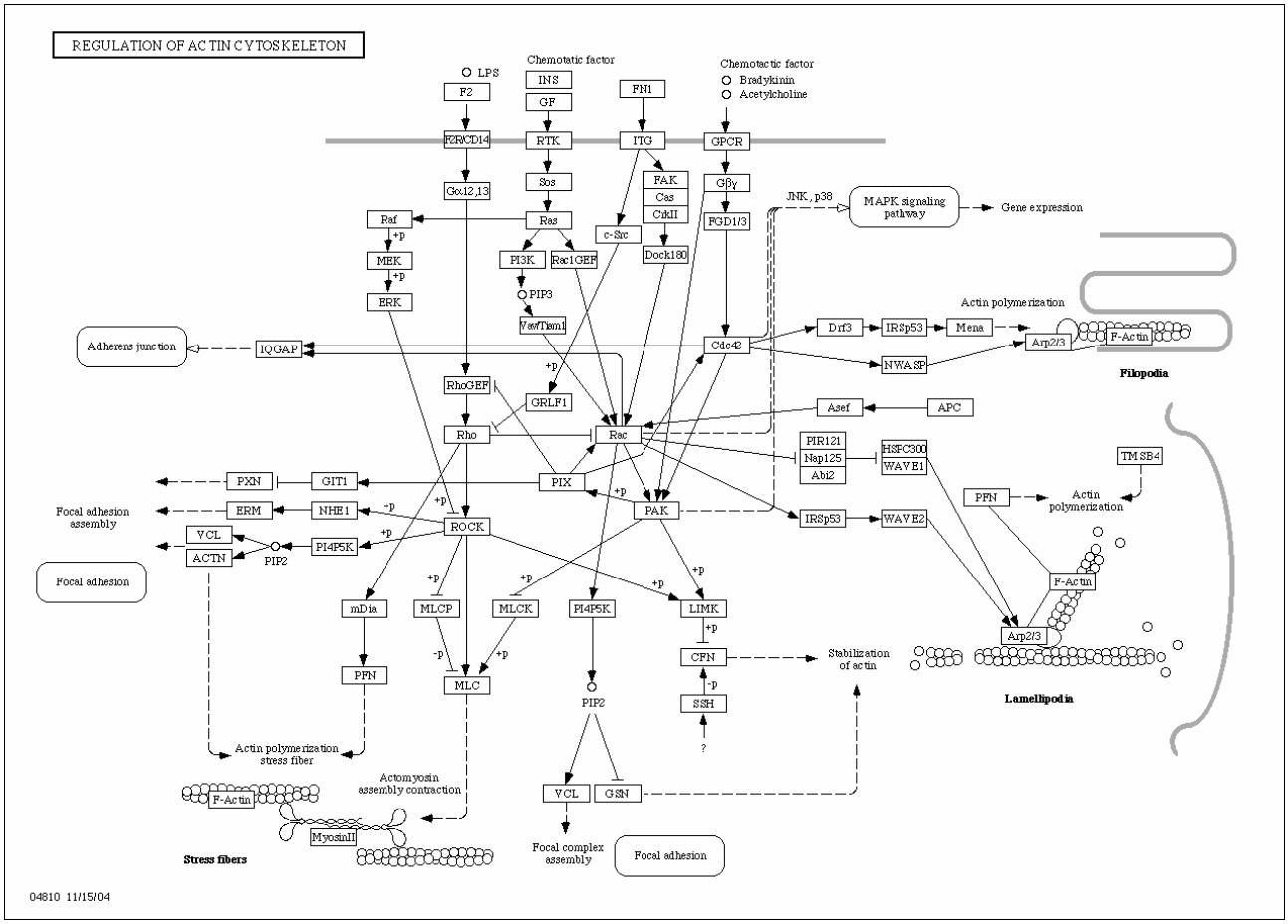

Fig. 2. Reference KEGG pathway representing the regulation of actin cytoskeleton.

Regulation of tissue strength and plasticity. The mitogen receptors are known to regulate synaptic strength and plasticity in adult nervous system via a cell-dependent manner [13]. Inspecting the mapping on the MAPK signaling pathway indicates that neutrophins signal through Trk receptor domains of platelet derived growth factor receptors and fibroblast growth factor receptors to regulate cell survival and proliferation. The whole process involved the fate of neural precursors, axon and dendrite growth and patterning, the expression and activity of functionally important proteins through vastus lateralis muscle contraction process [13]. Thus, the classical MAPK signaling transduction system regulates the strength and plasticity of vastus lateralis muscle tissues.

Regulation of energy homeostasis. PPAR is a member of the nuclear receptor superfamily, which is activated by various hydrophobic compounds [14]. PPAR has three isoforms, PPARalpha, PPARgamma and PPARdelta, which regulate homeostasis, cell proliferation, cell differentiation and associated hypolipidemia, atherosclerosis, diabetes and obesity [15]. According to the mapping on the PPAR signaling pathway (Fig. 3), PPARalpha and PPARgamma are activated in response to the stimulus of vastus lateralis muscle contraction. PPARalpha plays a role in the clearance of circulating in lipid metabolism in liver and skeletal muscle [15]. PPARgamma promotes adipocyte differentiation to enhance blood glucose uptake [15]. Thus, PPARalpha and PPARgamma act at crucial nodes of the regulatory network, which regulate energy homeostasis in response to the contraction stimulated by the assembly of focal adhesions and actin stress fibers.

Reactive oxygen species excess. Finally, according to the mapping on the mapping on the pathways of angiogenesis, leukocyte transendothelial migration, renin-angiotensin, complement and coagulation cascades, respectively, there is an indication of the presence of reactive oxygen species (ROS) excess (oxidative stress) during endurance training-induced vastus lateralis muscle contraction status. To date, it is well known that ROS are a family of molecules and its derivatives produced in all aerobic cells, which are yielded from the metabolisms of molecular oxygen [16]. On the basis of our present observations this endurance training course has led to the deleterious effects of oxygen from the metabolic reduction of the highly reactive and toxic species, promoting endothelial damage or dysfunction and atherosclerosis [16].

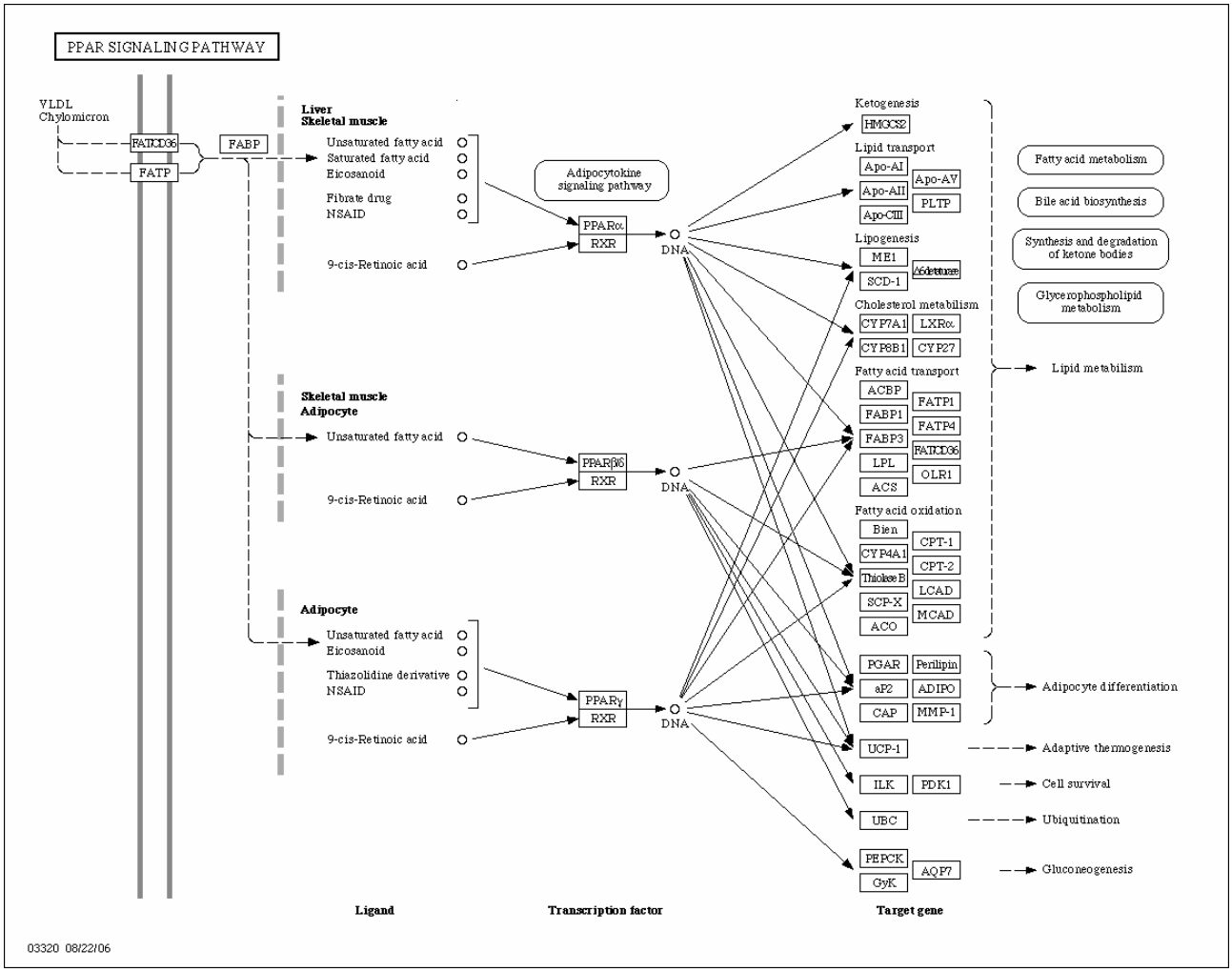

Fig. 3. The KEGG reference pathway representing the regulation of PPAR signaling transduction.

4. Discussions

We very often need to evaluate athletes' physiological conditions to endurance training from an initial state to a final state through a succession of many intermediate states. Unfortunately, none of the physiological assessment tools in current use, including the measurements of heart rate, oxygen consumption, and blood lactate concentration [5], can precisely monitor the changes in athletes' physiological conditions to endurance training. Particularly remarkable among these assessment tools is the measurement of blood lactate concentration. To do this, for example, a sample of blood is collected from the ear or fingertip from each elite swimmer after each swim in order to measure the amount of lactate acid [5]. Such blood testing involve swimming a series of repeated at progressively faster speeds during the training of first-rate elite swimmers [5]. Indeed, athletes may not prefer such blood testing since it can increase the possibility carrying the AIDS, hepatitis B, hepatitis C and many other viruses from one person to another [5]. Moreover, it would be unnecessary to do so since the quantities that are dealt with are merely the amount of lactate acid; consequently, an integrated knowledge of the athletes' physiological conditions being studied would be quite superficial.

Nonetheless, if the state of athletes' physiological conditions can be monitored using microarray-based genomic surveys, our approach described here will make available the recognition of athletes' physiological and pathological responses on the basis of the underlying expression patterns of genes. To do this, a natural way is to first scan and inspect significantly differential expressed genes associated with endurance training-induced muscle contraction and then to address the details of interest. In this paper, we first identify significantly differential expressed genes with endurance training-induced muscle contraction using fixed effect regression modeling, which effectively adapts the confounding effects arising from the association between genes. Next, we map key genes onto acknowledged biochemical pathways to attain a linkage between key molecules and biochemical pathways with endurance training-induced muscle contraction in a

cause-effect format. A study of the gene expression model relating to endurance training-induced vastus lateralis muscle contraction exhibits the occurrence of complex metabolisms, the transcriptional regulations of endurance training-induced vastus lateralis muscle contraction status, and the presence of oxidative stress causing endothelial damage or dysfunction and atherosclerosis. This example has demonstrated a feature of gene expression that makes fixed effect regression modeling particular valuable, namely the tendency of gene expression to derive insight into the underlying physiological and pathological conditions in connection with endurance training-induced vastus lateralis muscle contraction.

What we have found to be the most valuable feature of the approach described here is that it allows the recognition of the biological processes, giving analyses of the transcriptional regulations of biological interesting matters, which may not be observed easily otherwise. It is, of course, not very surprising that this approach as presented has resolved the problems of time shifts and space alterations in the expression patterns of genes, and that is has made possible a solid philosophy of space, time and gene expression data. The success of this computational approach has given us confidence to extract the information in gene expression data based on biological necessities.

Finally, we have seen that this approach described here can supply general tools to monitor athletes' physiological responses to endurance training, and that is has made possible a revolution in the outlook of athletic training.

5. Supplementary Information

Supplementary Table 1: significantly differential expressed genes associated with endurance training-induced vastus lateralis muscle contraction identified using fixed effect logistic regression modeling

6. Acknowledgment

This work is in part supported by a grant from Hong Kong Research Grant Council (Project CityU 122005).

7. References

[1] J. Olbrecht, O. Madsen, A. Mader, H. Liesen, and W. Hollmann, Relationship between swimming velocity and lactic concentration during continuous and intermittent training exercises. *International Journal of Sports Medicine.* 1985, **7**: 74-77.

[2] H. Heck, A. Mader, G. Hess, S. Mucke, R. Muller, and W. Hollman, Justification of the 4-mmol/1 lactate threshold. *International Journal of Sports Medicine.* 1985, **6**: 117-130.

[3] A. Mader, and H. Heck, A theory of the metabolic origin of anaerobic threshold. *International Journal of Sports Medicine Supplement.* 1986, Supplement **7**: 45-65.

[4] W. Hollman, R. Rost, K. De Meirleir, H. Liesen, H. Heck, and A. Mader. Cardiovascular effects of extreme physical training. *ACTA MEDICA SCANDINAVICA. SUPPLEMENTUM*. 1986, **11**:193-203.

[5] Maglischo, E.W. *Swimming Fastest*. Illinois: Human Kinetics Press. 2002.

[6] T. S. Collins, E. D. Green, A. E. Guttmacher, and M. S. Guyer. A vision for the future of genomics research. *Nature*. 2003, **422**: 835-847.

[7] Nature Biotechnology. Signal transduction as a drug discovery platform. *Nat. Biotech.* 2000, **18**: IT37-IT39.

[8] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Kirakawa. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research.* 2006, **34**: D354-D357.

[9] M. Fluck, C. Dapp, S. Schmutz, E. Wit, and H. Hoppeler. Transcriptional profiling of tissue plasticity: role of shifts in genes expression and technical limitations. *Journal of Applied Physiology*. 2005, **99**: 397-413.

[10] I. Kosztin, R. Bruinsam, P. O'Lague, and K. Schulten. Mechanical force generation by G proteins. *Proceeding to National Academic of Scence USA*. 2002, **99**: 3575-3580.

[11] D. J. Dickson. Molecular mechanisms of axon guidance. *Science*. 2002, **298**: 1959-1964.

[12] H. H. Williams, C. W Ward, E. E. Spangenburg, and R.M. Nelson. Functional aspects of skeletal muscle contractile apparatus and sarcoplasmic reticulum after fatigue. *Journal of Applied Physiology*. 1998, **85:** 619-626.

[13] P. Dent, A. Yacoub, P. B. Fisher, M. P. Hagan, and S. Grant. pathways in radiation responses. *Oncogene*. 2003, **22**: 5885-5896.

[14] I. Takada, and S. Koto. PPARs target genes. *Nippon Rinsho.* 2005, **63**: 573-577.

[15] N. Takahashi, T. Goto, T. Kusudo, T. Moriyama, and T. Kawada. The structures and functions of peroxisome

proliferator-activated receptors (PPARs). *Nippon Rinsho*. 2005, **63**: 557-564.

[16] W. Waris, and H. Ahsah. Reactive oxygen species: role in the development of cancer and various chronic conditions. *Carcinogenesis*. 2006, **5**: 1186-1192.