Genomics & Informatics Vol. 5(3) 102-106, September 2007

Prediction of the Exposure to 1763MHz Radiofrequency Radiation Based on Gene Expression Patterns

Min Su Lee[1], Tai-Qin Huang[2], Jeong-Sun Seo[2] and Woong-Yang Park[2]*

[1]Biointelligence Laboratory, School of Computer Science and Engineering, [2]ILCHUN Genomic Medicine Institute, MRC and Department of Biochemistry and Molecular Biology, College of Medicine, Seoul National University, Seoul 110-799, Korea

Abstract

Radiofrequency (RF) radiation at the frequency of mobile phones has been not reported to induce cellular responses in *in vitro* and *in vivo* models. We exposed HEI-OC1, conditionally-immortalized mouse auditory cells, to RF radiation to characterize cellular responses to 1763 MHz RF radiation. While we could not detect any differences upon RF exposure, whole-genome expression profiling might provide the most sensitive method to find the molecular responses to RF radiation. HEI-OC1 cells were exposed to 1763 MHz RF radiation at an average specific absorption rate (SAR) of 20 W/kg for 24 hr and harvested after 5 hr of recovery (R5), alongside sham-exposed samples (S5). From the whole-genome profiles of mouse neurons, we selected 9 differentially-expressed genes between the S5 and R5 groups using information gain-based recursive feature elimination procedure. Based on support vector machine (SVM), we designed a prediction model using the 9 genes to discriminate the two groups. Our prediction model could predict the target class without any error. From these results, we developed a prediction model using biomarkers to determine the RF radiation exposure in mouse auditory cells with perfect accuracy, which may need validation in *in vivo* RF-exposure models.

Keywords: support vector machine, prediction, microarray, radiofrequency radiation, auditory cell

Introduction

RF radiation does not transfer high enough energy to break the covalent bonds of macromolecules, but these low energy stimuli might be enough to induce molecular responses such as cell proliferation or cell death (Moulder *et al.*, 1999). While RF radiation itself can not mediate direct effects on DNA and proteins, it can trigger signaling pathways through changes in ionic distribution or membrane fluidity. In addition, the interactions between genes and RF radiation can lower the threshold of physiological changes, including various pathological conditions.

The brain is the most important target tissue to study the biological effect of RF radiation in mobile phone users (Hardell *et al.*, 1999). There was a report on theincrease in tumors ipsilateral to the side of the head on which subjects recalled phone use (Inskip *et al.*, 1999), but this was not substantiated by other studies (Muscat *et al.*, 2000). RF radiation exposure to rat headsdid not affect the incidence, malignancy, volume, multiplicity, latency, or fatality associated with any kind of neurogenic tumor (Zook and Simmens, 2006).

Alteration of cognitive and physiological functions of brains upon exposure to mobile phone-frequency RF radiation has been reported by several electrophysiological studies (Curcio *et al.*, 2005). In summary, RF exposure can induce measurable changes in human brain electrical activity, particularly in the alpha frequency band (8-13 Hz) over posterior regions of the scalp. In addition, rats exposed to RF showed neuronal damage in the cortex, hippocampus, and basal ganglia (Salford *et al.*, 2003). However, this work was not reproduced by others (Joubert *et al.*, 2007), and there are a number of points to consider regarding whether RF radiation can affect the human brain and its subsequent output in the form of cognition and behavior.

Gene expression profiling with microarrays can provide valuable information on the characteristics of certain physiological and pathological conditions. For example, gene expression profiles of -irradiated Jurkat cells showed p53-independent activation of the NF-B pathway (Park *et al.*, 2002). In the case of RF radiation, the expression patterns of HL-60 cells exposed to 2.45 GHz RF radiation were examined to select genes related to RF radiation (Lee *et al.*, 2005b). We have used HEI-OC1 mouse auditory cells to find biomarkers related to RF radiation-exposure, and have developed prediction models of RF-exposure using support vector machine (SVM) algorithms (Vapnik, 1998) with selected biomarkers based on the Weka environment (Witten and Frank, 2005).

*Corresponding author: E-mail wypark@snu.ac.kr
Tel +82-2-740-8241, Fax +82-2-744-4534
Accepted 15 August 2007

Materials and Methods

Cell culture

Conditionally-immortalized HEI-OC1 auditory cells were kindly provided by Dr. Kalinec (House Ear Institute, Los Angeles, CA, USA). Cells were cultured under permissive conditions (33°C 10% CO2) in high-glucose Dulbecco's Modified Eagle's Medium (DMEM; GIBCO/BRL, Gaithersburg, Md., USA) containing 10% heat-inactivated fetal bovine serum (FBS, GIBCO/BRL) and 50 U/mL gamma-interferon (Gemzyme, Cambridge, MA, USA) as previously described (Jat, 1991).

RF exposure system

The exposure system specifically designed for this study was reported previously by Lee *et al.* in our laboratory (Lee *et al.*, 2006). For the PCS exposure system, a real CDMA signal at 1762.5 MHz was applied.

Cells were exposed to 1763 MHz RF radiation in 100 mm Petri dishes containing 18 mL of medium. The exposure system was then warmed up for 30 min to equilibrate it before exposure of RF. RF radiation exposure was conducted at SAR values of 20 W/kg. During exposure, the temperature inside the chamber was maintained at 33±0.2°C by circulating water within the cavity. After an exposure of 24 or 48 hr, the cells were transferred to a cell culture incubator for 5 hr and then harvested. For sham exposure, cells were incubated in the RF radiation device, but were not exposed to RF radiation.

RF-exposure design

HEI-OC1 mouse auditory cells were exposed to 20 W/kg of 1763 MHz RF radiation for 24 hr, and then cells were incubated for 5 hr to recover from acute responses (R5 group). Sham-exposed cells were harvestedin parallel (S5 group). We repeated these sets of experiment threetimes to collect biological triplicates for every sample.

RNA extraction and microarray

Samples in each group were harvested, and their total RNA was extracted by dissolving cells in Trizol (Sigma). After fractions containing the RNA were collected, total RNA was purified using Qiagen RNeasy columns. The array used in this experiment was the Applied Biosystems 1700 full genome expression mouse microarray,which includes 32,000 mouse genes from the public and Celera databases (http://www.pantherdb.org/).

Data analysis

Fluorescence intensity was processed and measured using the Applied Biosystems 1700 Chemiluminescent Microarray Analyzer. Intensity data were imported to an in-house microarray database. The overall procedure for constructing a prediction model to discriminate the exposure of RF radiation using the microarray data is shown in Fig. 1.

To perform reliable microarray data analysis, some probes were filtered out based on the following 3-step gene-filtering process. Firstly, control genes were excluded after checking their expression level. Then, unreliable intensities of probes were excluded based on flag information. Finally, reliable probes, whose ratio of signal-to-noise intensity (S/N) was more than 3 in 50% or more of the total samples, were used to construct a prediction model. The threshold of the S/N ratio was derived from a rigorous quality control test by Applied Biosystems that showed that if the S/N of a probe is more than 3, then the signal intensity of the probe is reliable with 99.9% confidence. As a result, 15,040 probes were left.

Quantile normalization was applied to remove systematic variance. Quantile normalization is a simple yet powerful method to normalize across arrays by making the distribution of probe intensities for each array in a set of arrays the same, and it reduces the variance slightly better than Lowess normalization (Bolstad1 *et al.*, 2003).

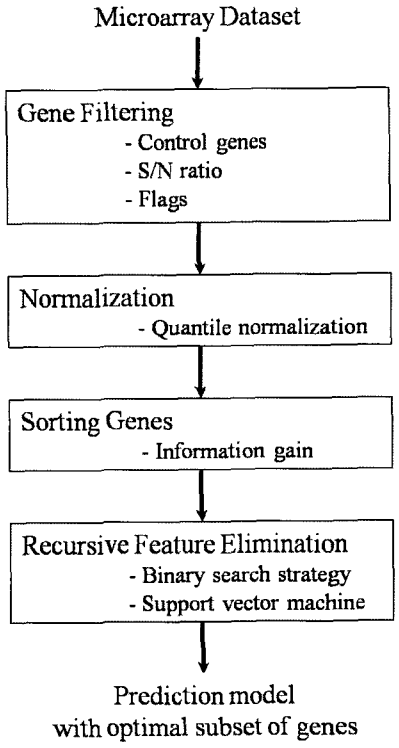

Fig. 1. Analysis procedure for constructing a prediction model

We sorted genes according to the significance of changes in gene expression levels between the two groups using information gain. Information gain is a measure of the effectiveness of a feature in classifying the training data (Mitchell, 1997). Given entropy $Entropy(S)$ as a measure of the impurity in a training data, information gain $IG(S, F)$ of a feature F relative to a collection of samples S is the expected reduction in entropy caused by partitioning the samples according to this feature. Hence, the information gain can be defined as the difference between the entropy of the original collection S and the expected value of the entropy after S is partitioned using feature F as follows:

$$IG(S,F) \equiv Entropy(S) - \sum_{v \in Value(F)} \frac{|S_v|}{|S|} Entropy(S_v),$$

where $Value(F)$ is the set of all possible values for feature F, and S_v is the subset of S for which feature F has value v.

To construct an optimal classification model to predict the exposure to RF radiation with a minimal set of biomarkers, we adopted a recursive feature elimination (RFE) procedure (Furlanello, *et al.* 2003, Guyon, *et al.* 2002) based on information gain and a support vector machine (SVM) algorithm (Vapnik, V.N., 1998). The RFE procedure recursively removes the features that cause the minimum variation with support vector machines (Guyon *et al.* 2002). SVMs have exhibited superb performance in binary classification tasks that search for a hyperplane that separates two classes of data with the largest margin between the hyperplane and the point closest to it. At each model building step in RFE, a classifier and a ranked gene set are constructed and evaluated, and a single feature is eliminated at each step. Hence,the basic RFE procedure with SVM has high computational cost. To speed up the RFE-SVM procedure, we used the information gain-based RFE procedure to obtain an optimal gene subset for predicting the exposure of RF radiation. Gene selection with RFE procedures based on information gaineliminates gene redundancy effectively and yields better and more compact gene subsets. We constructed prediction models with selected gene subsets using a sequential minimal optimization

```
F[0..N-1]: a feature set with N features that is sorted by information gain in decreasing order
accuracy(i): accuracy of a prediction model based on SVM with F[0...i] gene set

low = 0
high = N-1
value = accuracy(N-1)

IG_RFE_SVM(F[0...N-1], value, low, high) {
    if (high ≤ low)
        return F[0...N-1] and value
    mid = (low + high) / 2
    value_2 = accuracy(mid)
    if (value_2 ≥ value)
        return IG_RFE_SVM(F[0...mid], value_2, low, mid)
    else (value_2 < value)
        return IG_REF_SVM(F[0...high], value, mid, high)
}
```

Fig. 2. Pseudo-code of information gain-based recursive feature elimination procedure with SVM

Table 1. List of selected genes derived by information gain-based RFE with SVM

UniGene ID	Gene_Symbol	Gene_Name
Mm.298	Slc25a3	solute carrier family 25 (mitochondrial carrier, phosphate carrier), member 3
Mm.161149	Zswim1	zinc finger, SWIM domain containing 1
Mm.246003	Fzd1	frizzled homolog 1 (Drosophila)
Mm.44228	D6Wsu163e	DNA segment, Chr 6, Wayne State University 163, expressed
Mm.17993	1700027N10Rik	RIKEN cDNA 1700027N10 gene
Mm.316306 Mm.379375 Mm.371779	LOC630401 Hnrpa3 LOC545592 LOC629025 LOC620454 LOC547321	heterogeneous nuclear ribonucleoprotein A3
Mm.12882	Hsd17b7	hydroxysteroid (17-beta) dehydrogenase 7
Mm.348054	Pcdhb22	protocadherin beta 22
Mm.246563	Tgoln2 Tgoln1	trans-golgi network protein 2|trans-golgi network protein

algorithm with a logistic regression model and RBF kernel for training a support vector classifier (Platt, J., 1998) based on the Weka environment (Witten and Frank, 2005).

We estimated the prediction error of the model with leave-one-out cross-validation (LOOCV) (Tan, *et al.*, 2005). Specifically, if the number of samples is *n*, LOOCV means that the cross-validation procedure is run *n* times, each time using one of the samples as the test set and the others as training sets. LOOCV had the advantage of utilizing as much data as possible for training. In addition, the test sets were mutually exclusive, and they effectively covered the entire dataset. Also, LOOCV seems to offer a chance of squeezing the maximum out of a small dataset and obtaining as accurate an estimate as possible (Tan, *et al.*, 2005).

Results and Discussion

RF exposure to HEI-OC1 mouse auditory cells

Upon exposure to 1763 MHz RF radiation, HEI-OC1 mouse auditory cells did not show any change in cell morphology or growth (unpublished data). We kept the cells in a CO_2 incubator while cells were exposed to RF radiation. RF radiation can produce heat in the culture media, but isothermal water at 37°C circulated continuously at the bottom of the chamber throughout the experiment. Because we kept the chamber inside a CO_2 incubator, cells were completely shielded from any electromagnetic field radiation generated from other electronic sources. For the sham group, we assembled the exact same experimental setup as in the RF exposure without applying the current.

Information gain-based recursive feature elimination with SVM

We identified a discriminative and compact gene subset for predicting the exposure of RF radiation according to the following procedure. Firstly, genes were sorted by their information gain score. Then using RFE with SVM based on a binary search algorithm, we searched for the gene subset with the highest accuracy and minimal size. Specifically, a binary search algorithm finds the median element g_n in a sorted gene set, compares the prediction accuracy of the top *n* genes to the one that was previously computed based on the SVM algorithm, and determines if the prediction accuracy is greater than, less than, or equal to the previously-computed one, iteratively. The pseudo-code of the information gain-based recursive feature elimination procedure with SVM is shown in Fig.2.

By adopting information gain-based RFE with SVM in a divide-and-conquer approach, we could effectively eliminate chunks of uninteresting genes and identify 9 genes with perfect classification accuracy. We briefly summarize information about the 9 genes in Table 1.

Acknowledgements

This work was supported by the IT R&D program of MIC/IITA [2007-F-043-01, Study on Diagnosis and Protection Technology based on EM].

References

Bolstad, B.M., Irizarry R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. *Bioinformatics* 19, 185-193.

Curcio, G., Ferrara, M., Moroni, F., D'Inzeo, G., Bertini, M., and De Gennaro, L. (2005). Is the brain influenced by a phone call? An EEG study of resting wakefulness. *Neurosci Res.* 53, 265–270.

Furlanello, C., Serafini, M., Merler, S., and Jurman, G. (2003). Entropy-based gene ranking without selection bias for the predictive classification of microarray data. *BMC Bioinformatics*, 4, 54.

Guyon, I., Weston, J., Barnhill, B., and Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 46, 389-422.

Hardell, L., Nasman, A., Pahlson, A., Hallquist, A., and Mild, K.H. (1999).Use of cellular telephones and the risk of brain tumours: a case control study. *Int J Oncology* 15, 113-116.

Inskip, P.D., Tarone, R.E., Hatch, E.E., Wilcosky, T.C., Shapiro, W.R., Selker, R.G., Fine, H.A., Black, P.M., Loeffler, J.S., and Linet, M.S. (2001). Cellular-telephone use and brain tumors. *New England Journal of Medicine* 344, 79-86.

Jat, P.S., Noble, M.D., Ataliotis, P., Tanaka, Y., Yannoutsos, N., Larsen, L., Kioussis, D. (1991). Direct derivation of conditionally immortal cell lines from an H-2Kb-tsA58 transgenic mouse. *Proc. Nati. Acad. Sci. USA* 88, 5096-5100.

Joubert, V., Leveque, P., Cueille, M., Bourthoumieu, S., and Yardin, C. (2007). No apoptosis is induced in rat cortical neurons exposed to GSM phone fields. *Bioelectromagnetics* 28, 115-121.

Lee, J.S., Huang, T.Q., Kim, T.H., Kim, J.Y., Kim, H.J., Pack, J.K., and Seo, J.S. (2006). Radiofrequency radiation does not induce stress response in human T-lymphocytes and rat primary astrocytes. *Bioelectromagnetics*. 27, 578-88.

Lee, S., Johnson, D., Dunbar, K., Dong, H., Ge, X., Kim, Y.C., Wing, C., Jayathilaka, N., Emmanuel, N., Zhou, C.Q., Gerber, H.L., Tseng, C.C., and Wang, S.M. (2005). 2.45 GHz radiofrequency fields alter gene expression in cultured human cells. *FEBS Lett*. 579, 4829-4836.

Mitchell, T.M. (1997). Machine Learning. McGraw-Hill, Singapore.

Moulder, J.E., Erdreich, L.S., Malyapa, R.S., Merritt, J., Pickard, W.F., and Vijayalaxmi. (1999). Cell phones and cancer: what is the evidence for a connection? *Radiat Res.* 151, 513-531.

Muscat, J.E., Malkin, M.G., Thompson, S., Shore, R.E., Stellman, S.D., McRee, D., Neugut, A.I., and Wynder, E.L. (2000). Handheld cellular telephone use and risk of brain cancer. *JAMA* 284, 3001–3007.

Park, W.Y., Hwang, C.I., Im, C.N., Kang, M.J., Woo, J.H., Kim, J.H., Kim, Y.S., Kim, J.H., Kim, H., Kim, K.A., Yu, H.J., Lee, S.J., Lee, Y.S., and Seo, J.S. (2002). Identification of radiation-specific responses from gene expression profile. *Oncogene*21, 8521-8528.

Platt, J. (1998). Fast training of support vector machines using sequential minimal optimization. Advances in kernel methods - support vector learning. MIT Press.

Salford, L.G., Brun, A.E., Eberhardt, J.L., Malmgren, L., and Persson, B.R. (2003). Nerve cell damage in mammalian brain after exposure to microwaves from GSM mobile phones. *Environ Health Perspect*. 111, 881-883.

Tan, T.N., Stenbach, M., and Kumar, V. (2005). Introduction to Data Mining, Addison Wesley.

Vapnik, V.N. (1998). Statistical Learning Theory. Wiley, New York, NY.

Witten, I.J. and Frank, E. (2005). Data mining: practical machine learning tools with java implementations. 2nd ed., Morgan Kaufmann, San Francisco, CA.

Zook, B.C. and Simmens, S.J. (2006). The effects of pulsed 860 MHz radiofrequency radiation on the promotion of neurogenic tumors in rats. *Radiat Res*. 165, 608-615.