Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/100402

Bioinformatic Analysis of Human OPN3 Alternative Promoters Associated with Cancer

Mohammed A. Ibrahim Al-Obaide[1], *Abdel-Salam G. Abdel-Salam* [2,*], *Hytham Alobydi*[3] *and Kalkunte S. Srivenugopal*[1].

[1] Department of Biomedical Sciences, School of Pharmacy, Texas Tech University Health Sciences Center, Amarillo, TX 79106, USA.
[2] Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, Qatar
[3] Biomedica, LLC; Sterling Heights, MI, USA.

Received: 16 Mar. 2016, Revised: 2 May 2016, Accepted: 3 May 2016
Published online: 1 Jul. 2016

Abstract: In this bioinformatics study we analyzed the functional motifs and structural features of regulatory regions of OPN3, a member of the guanine nucleotide-binding protein (G protein)-coupled receptor super-family, which is involved in cancer and drug resistance. Our analyses showed major differences in the two OPN3 promoters located at the 5'-side of the locus. The analyzed parameter set, structural differences of regulatory sequences dignified by dinucleotide base stacking energy and GC-skew, CpG islands (CGI) and transcription factors binding sites (TFBS) of alternative promoters, showed differences that can be exploited to modulate and/or enhance the expression of OPN3 in human tissues. Further, the statistical cluster analysis of OPN3 mRNA expression measurements in various human tissues indicated significant variation that could be associated with SOX5 and EBF3, transcription factors highly associated with cancer. Our findings strengthen a role for OPN3 regulatory sequences in oncogenesis and drug resistance.

Keywords: Bioinformatics, Biostatistics, Cancer, Drug resistance, Genomics, GC skew, Stacking energy.

1 Introduction

Bioinformatics has evolved through the convergence of genomics and informatics to give deeper insight into the complex sets of genomics and biological data. The bioinformatic tools have become indispensable in the cancer genomics research and are used to give additional insight into complex multifactorial genetic characteristics and regulatory genomic sequences associated with cancer and personalized treatment.

Opsin 3, OPN3, previously known as encephalopsin is a member of the guanine nucleotide-binding protein (G protein)-coupled receptor superfamily expressed in brain, liver, placenta, heart and pancreatic tissues [1, 2]. The human OPN3 protein consists of single polypeptide of 403 amino acids showing 85.5% identity with the mouse protein [1]. It is implicated in encephalic photoreception, G-protein coupled receptor signaling pathway, photo-transduction and regulation of circadian rhythm [1, 2]. Recently, OPN3 was found associated with resistance to chemotherapeutic agents in various cancers. It is differentially regulated by histone deacetylase inhibitors (HDACi) in human myeloma cell lines (HMCL) [3]. OPN3 is also associated with cisplatin resistance in various cancer cells [4, 5]. Further, OPN3 depletion is involved in fluorouracil (5-FU) resistance, suggesting that therapeutic strategies targeting OPN3 may improve HCC sensitivity to chemotherapy [6]. Knockdown of OPN3 within 200 kb up and downstream of SNPs that were associated with SCLC overall survival significantly desensitized H196 cells to paclitaxel [7]. In addition, cancer databases showed that variants of this gene has been implicated in various types of cancers, breast, kidney, salivary glands, and intestine. Moreover, OPN3 has been reported to be involved in the pathogenesis and susceptibility to asthma [8]. Thus, the OPN3 plays important roles in numerous physiological and pathological processes.

Although the cDNA sequence encoding OPN3 was reported in 2001 [2], little is known on the regulation of OPN3 transcription and involvement of regulatory elements in its transcription. A single promoter is believed to mediate the varied and tissue-specific expression of the OPN3 [4, 5]. However, we observed several promoters in the databases for this gene. Granting the gene undergoes promoter methylation in several cancer types and consequent silencing of the gene [4, 5],

* Corresponding author e-mail: abdo@qu.edu.qa

© 2016 NSP
Natural Sciences Publishing Cor.

there are no reports in the literature about the roles played by TFBSs and structural features of the OPN3 alternative promoters in gene expression and anticancer resistance. Resistance to anticancer drugs greatly impedes successful therapy, and in a novel concept we hypothesized that cancer cells utilize differentially the regulatory motifs and structural features of alternative promoters for acquisition of drug resistance. Here, we report the bioinformatic analysis of transcription factors binding sites (TFBS) and structural characteristics of two unique OPN3 alternative promoters, which revealed novel specific regulatory sequences for each promoter that can be exploited for differential expression by OPN3.

2 Methods

2.1 Genomic databases

The genomic criteria of OPN3 gene and the alternative promoters of the gene were searched during January-June 2015 in genomics databases Table 1. Several tools in the databases were used to retrieve and analyze the sequences, identify the strand (forward or reverse), flip the strand and in search for a specific sequence.

2.2 Verification of map locations

The precise genomic map locations of the identified sequences of the OPN3 and alternative promoters were verified and updated to hg38 version of human genome sequence by using the BLAT tool, Gene Sorter and Table Browser tools of UCSC Genome Bioinformatics database.

2.3 Search for regulatory elements in the alternative promoters

The identified and verified sequences of the OPN3 alternative promoters were analyzed for search of TFBSs, namely, TATA-8 (TATAWAWR) and TATA-532 (HWHWWWWR, excluding: HTYTTTWR, CAYTTTWR, MAMAAAAR, CTYAAAAR), INR (YYANWYY), BRE (SSRCGCC), DPE (RGWCGTG), EBF3 and SOX5 binding sites [9–13]. Transcriptional Regulatory Element Database (TRED) and Biomedicallc tools were used to search for the TFBSs.

2.4 Search for CpG islands

Identification of CGIs in the alternative promoters was carried out according to the parameter set and equation reported by Gardiner-Garden and Frommer 1987 [14]. The parameter sets used to search for CGIs in the alternative promoters: Obs/Exp CpG $\geq$ 0.6 and $\%G+C > \%55$. The ratio observed/expected (O/E) CpG was calculated according to the reported equation:

$$Obs/ExpCpG = \left[\frac{Number\,of\,CpG}{Number\,of\,C \times Number\,of\,G} \times N\right]$$

Where N is the total number of nucleotides in the sequence being analyzed. The DNA sequences were searched at 100 bps window (N= 100) moving across the sequence at 1 bps intervals.

2.5 Analysis of structural features of the OPN3 regulatory sequences

The dinucleotide base-stacking energy values were derived from values of dinucleotide base-stacking energy provided by Ornstein et al. [15]. According to the scale, which is in kcal/mol, the range of values from -3.82 kcal/mol (unstack easily) to -14.59 kcal/mol (difficult to unstack), thus the obtained values show the relative dissociation stability of the double helix structure. Whereas, GC skew is calculated as ((G-C)/(G+C)), where C and G denote the numbers of cytosine and guanine [16, 17]. GC skew is useful for predicting the R loop formation and the origin and terminus of replication. Programs written in R were used to analyze the sequences and to plot the data of dinucleotide base-stacking energy values (Kcal/mol/dimer) and GC skew along the length of the promoter sequence.

2.6 Blast tree map

NCBI standard nucleotide BLAST tool was used to search by pairwise alignments for similar sequences to the OPN3 and the two promoters sequences investigated in the study. The distance tree of the obtained pairwise comparisons was produced to show evolutionary relatedness of regulatory sequences among species.

2.7 Statistical analysis

The regulatory sequences were analyzed and achieved using Excel software and programs written in R. The independence of each promoter element was examined using Fisher's exact probability test. Cluster analysis of OPN3 expression measurements were analyzed by cluster analysis using SPSS software [18, 19].

© 2016 NSP
Natural Sciences Publishing Cor.

Table 1: Web-based resources used in this study.

Database/ Description	URL
NCBI-GenBank	http://www.ncbi.nlm.nih.gov/genbank/
TRED (Transcriptional Regulatory Element Database)	https://cb.utdallas.edu/cgibin/TRED/tred.cgi?process=home
PEDB (Mammalian Promoter/Enhancer Database)	http://promoter.cdb.riken.jp/
EPD (Eukaryotic promoter database)	http://epd.vital-it.ch/
MPromDb (Mammalian Promoter Database)	http://mpromdb.wistar.upenn.edu/
Ensembl	http://useast.ensembl.org/info/genome/funcgen/index.html
UCSC Genome Bioinformatics	https://genome.ucsc.edu/index.html

Table 2: Map locations of alternative promoters of OPN3 (chr1:241593150-241640399) reported in the Transcriptional Regulatory Element Database (TRED). Plus and star signs indicate the presence of identified of CGI and overlapping promoters respectively.

Alternative promoter source	Map locations	CGI
TRED- 2156*	chr1:241640049-241641048	+
TRED-2155*	chr1:241640054-241641053	+
TRED-118881	chr1: 241593313 -241594312	-
TRED-114168	chr1: 241633328 -241634327	-
TRED-2157	Chr1:241635630-241636629	+

3 Results and discussion

3.1 Genomic context of OPN3 alternative promoters

The human OPN3 locus is located in the negative strand of the long arm of chromosome one at 1q43, and occupies genomic space of 47250 nucleotides at chr1: 241593150-241640399. NCBI-Gene database showed this region is mapped in the chromosome 1 - NC_000001.11 region that also contained three more protein coding genes, CHML, FH and KMO. Careful analysis of map locations for genes in this region showed OPN3 is divergently paired head to head with KMO gene and 2.497 kb is the sequence between two genes (Figure 1 A). KMO encodes kynurenine 3-monooxygenase a mitochondrion outer membrane protein that catalyzes the hydroxylation of L-tryptophan metabolite, L-kynurenine. The revelation that the OPN3 gene, located on reverse strand, and is paired with the other protein coding gene located on opposite forward strand, might indicate presence of regulatory region taking into account recent report by Ensembl database about presence of overlapping Promoter Flanking Region between two genes. Table 2 shows the five promoters involved in expression of OPN3, four are located at the 5'-side of OPN3, two are overlapping sequences, TRED- 2156 and TRED-2155, TRED-2157 is located between TRED-2156 and TRED-114168, and TRED-118881 is located at 3'-side of OPN3 (Figure 1 B). It is possible to assume that TRED-118881 might be associated with the Promoter Flanking Region and involved in expression of KMO and OPN3, this remains for further investigation in future studies. The alternative promoters of OPN3 could be divided into two groups according to %G+C and Obs/Exp CpG values, which are indicators of CGIs [14]. Accordingly, alternative promoters with CGIs are: TRED-2157, TRED- 2156 and TRED-2155 (Table 2); whereas the rest of alternative promoters, TRED-118881 and TRED- 114168 lack CGIs. For that reason we selected two promoters at 5'-side one with CGI and the second without for further analysis, namely, TRED-2156 and TRED-114168.

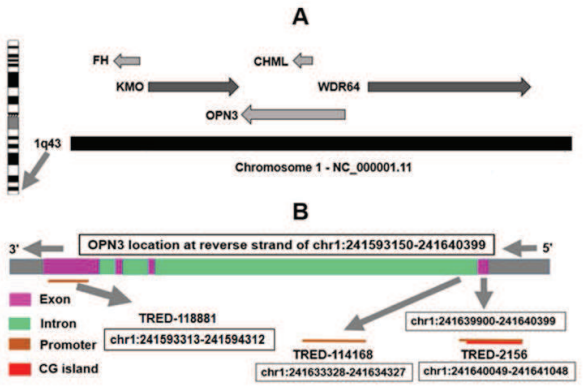

Fig. 1: Map locations of OPN3 locus and its alternative promoters on chromosome 1 - NC_000001.11 region. (A) Chromosomal locations of CHML, FH and KMO in comparison with OPN3. (B) Intergenic and intragenic locations of alternative promoters along OPN3 space.

© 2016 NSP
Natural Sciences Publishing Cor.

Table 3: Transcription factors binding sites (TFBSs) in the two alternative promoters of OPN3 associated with cancer.

Gene	Promoters ID	INR	TATA-8	TATA-532	CCAAT	TAACC	EBF3	BRE	DPE	SOX
OPN3	TRED-2156	8	0	2	1	0	1	5	0	1
	TRED-114168	14	2	18	2	1	0	0	0	2

3.2 Molecular and structural characteristics of TRED-2156 and TRED-114168

The bioinformatics analysis revealed that the two selected promoters, TRED-2156 and TRED-114168, have unique TFBSs in the OPN3 regulatory region located at reverse strand (Table 3). The TRED-2156 was found rich with BRE and INR sequences, and the clustering pattern of INR and BRE was unique, seven out of eight INR sequences were clustered in the 5'- side of TRED-2156 sequence and the five identified BRE sequences were located in the 3'- side. It is interesting to note, TRED-2156 sequence does not contain any of TATA-8. On the other hand, the TRED-114168 sequence harbored TATA-8 in the downstream sequence, whereas the INR motifs were found along upstream and downstream promoter sequence. Other regulatory TFBS motif, EBF3 binding site, was observed in the TRED-2156 but was not found in the TRED-114168. However, the SOX5 TFBSs were identified in both promoters (Table 3).

Considering the importance of EBF3 and SOX5 transcription factors, because of their involvement in cancer [20, 21]. We analyzed the occurrence of their binding sites in both strands of two OPN3 promoters. Our analysis showed that the identified TFBS motifs for SOX5 and EBF3 are also abundant in the forward strand of TRED-2156 and TRED-114168 promoters.

Table 4 shows the specific sequences of EBF3 and SOX5 TFBSs and their map locations in TRED-2156 and TRED-114168. This observation could be correlated with the status of OPN3 expression in the human tissues. The type of tissue-specific pattern of mRNA expression can show important signs about gene function and its potential role in physiological and pathological processes. Accordingly, we used cluster analysis, a well-known and a widely used technique, to partition a set of OPN3 mRNA expression patterns into disjoint and homogeneous clusters or groups. Further justification for use of cluster analysis in such analysis, it is a data reduction tool that creates subgroups that are more manageable than individual datum [18]. Therefore, cluster analysis technique was applied to the OPN3 177 measurements of 79 tissues reported in BIOGPS database. We observed, OPN3 mRNA expression measurements of the same tissue are distributed in different classes. For example, our cluster analysis of OPN3 mRNA expression measurements of 40 samples showed the gene expression data for prefrontal cortex, temporal lobe, thalamus, thyroid, whole blood, and whole brain were distributed in different clusters (Figure 2).

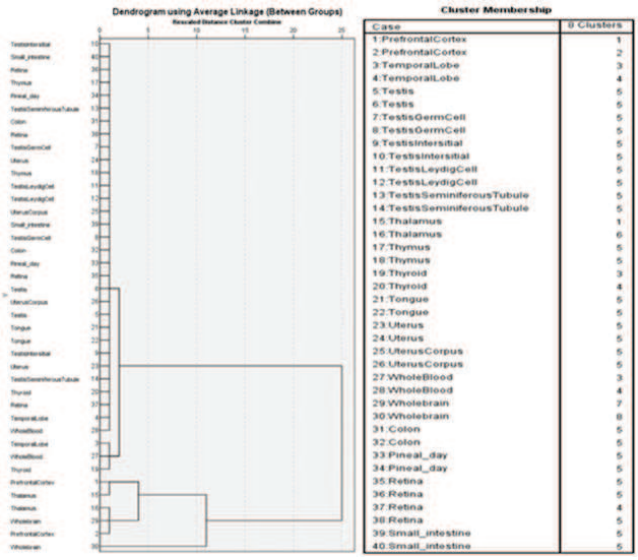

Fig. 2: Cluster analysis of OPN3 mRNA expression measurements of 40 tissues samples.

Together, these presented and other analyses not presented here give an indication of the differential usage of TBFSs in OPN3 expression.

3.3 Structural characteristics of two promoters

We investigated possible correlation between structural characteristics and distribution of TFBSs along two promoters' sequences, TRED-2156 and TRED-114168. In this context, we investigated the dinucleotide base-stacking energy and GC skew values for two regulatory sequences. The INR and TATA sites were found mainly along sequences easily unstack, whereas BRE sequences were observed at sequences comparatively difficult to unstack (Figures 3 and 4). TRED-2156 and TRED-114168 regions with calculated stacking energy values of -6.5 to -7 kcal/mol harbored INR and TATA motifs. On the other hand, BRE sequences were found clustered within regions that showed values between -9 and -9.75 kcal/mol. Also we observed GC skew in the TRED-2156 CGI region at TSS vicinity (Figure 5).

GC skew is a result of strand asymmetry down-stream TSSs and therefore it is an indication of possible formation of R-loops that are correlated with

© 2016 NSP
Natural Sciences Publishing Cor.

Table 4: TFBSs for SOX5 (SBS) and EBF3 (EBS) identified in the sequences of two OPN3 promoters in the reverse and forward strands.

Promoter	TFBSs	Start	End	Strand	Sequence
TRED-2156	SBS	181	187	-	TTCAATG
	EBS	533	540	-	CCCGCGGG
	EBS	460	467	+	CCCGCGGG
TRED-114168	SBS	189	195	-	TACAAAG
	SBS	702	708	-	ATCAATG
	SBS	304	310	+	AACAAAG
	SBS	314	320	+	TTCAATG
	SBS	383	389	+	ATCAAAG

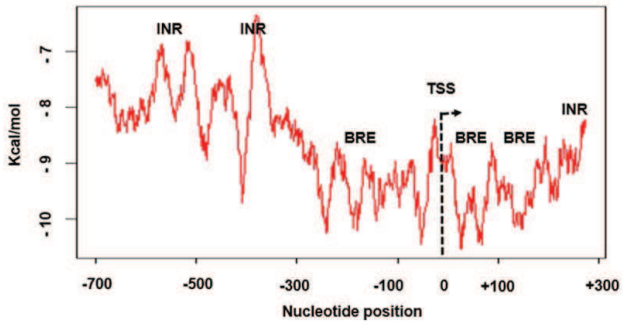

Fig. 3: Plot of the dinucleotide base-stacking energy values (Y axis) along the TRED-2156 regulatory sequence. The locations of INR, BRE, and TSS are shown.

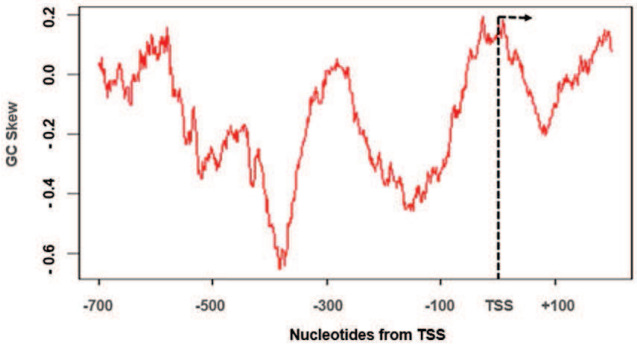

Fig. 5: The GC skew (Y axis) along the TRED-2156 promoter sequence (X axis). The bent arrow represent transcription start site (TSS).

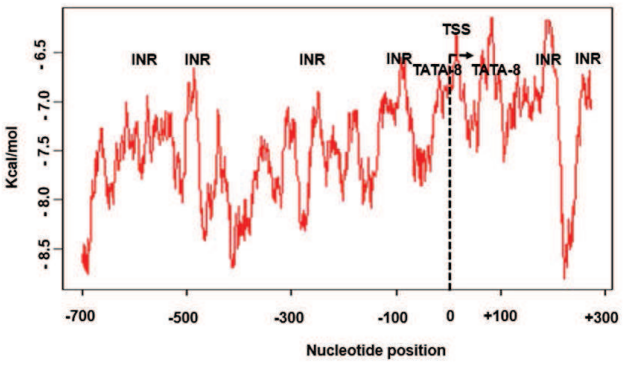

Fig. 4: Plot of the dinucleotide base-stacking energy values (Y axis) along the TRED-114168 regulatory sequence. The locations of INR, TATA-8, and TSS are shown.

un-methylated status of CGI [17]. Accordingly, dinucleotide base-stacking energy and GC skew values of OPN3 alternative promoters sequences could be considered other important independent variables besides TFBS and CGI for OPN3 expression.

3.4 Blast tree map of OPN3 and regulatory sequences

The genomic evolutionary relationships of OPN3 and the two regulatory sequences, TRED-2156 and TRED-114168, were explored by blast tree map. The generated blast tree maps showed that OPN3 was found mainly in primates indicating its recent evolution. On the other hand, the TRED-2156 regulatory sequence with CGI was found widespread and conserved across the animal kingdom as compared with the blast tree map of TRED-114168, which suggested more recent involvement of TRED-114168 in OPN3 gene expression.

4 Conclusions

The bioinformatic tools have become indispensable in the cancer genomics research and are used to give additional insight into complex multifactorial genetic characteristics and genomic sequences associated with cancer and personalized treatment. Before this study, there was no definitive mapping of the TFBSs for OPN3, which is involved in physiological and pathological settings including cancer and drug resistance. Our extensive and

© 2016 NSP
Natural Sciences Publishing Cor.

careful analysis revealed specific TFBSs for each promoter, of these two types SOX5 and EBF3 binding sites apparently have unique roles in OPN3 expression and are associated with oncogenesis. The data reported here highlights the roles of structural features of regulatory sequences as given by CGI, GC-skew, R-loop formation and dinucleotide base stacking energy in the complexity and sophistication underlying the transcriptional and possible posttranscriptional regulation of the OPN3 gene and consequent differential expression in human normal and cancer cells. We conclude that bioinformatics analysis can give further insight into the mechanisms of acquisition of specific molecular hallmarks in the regulatory regions of OPN3 associated with cancer development and drug resistance.

Conflict of Interest

The authors confirm that this article content has no conflict of interest.

Acknowledgements

This work was supported in part by a grant from the Cancer Prevention and Research Institute of Texas (RP130266) to KSS.

References

[1] S. Blackshaw, S.H. Snyder, J. Neurosci. 19, 3681-3690 (1999).
[2] S. Halford, M. S. Freedman, J. Bellingham, S.L. Inglis, S. Poopalasundaram, B.G. Soni, R.G. Foster, D.M. Hunt. Genomics 72, 203-208 (2001).
[3] S. Mithraprabhu, T. Khong, A. Spencer, Cell Death and Disease 5, e1134 (2014).
[4] X. Chang, C. Monitto, S. Demokan, M. Kim, S. Chang, X. Zhong, J.A. Califano, D. Sidransky, Cancer Res. 70, 2870-2879 (2010).
[5] C.T. Viet, D. Dang, S. Achdjian, Y. Ye, S.G. Katz, B.L. Schmidt, PLoS ONE 9(11), e112880 (2014).
[6] J. Jiao, S. Hong, J. Zhang, L. Ma, Y. Sun, D. Zhang, B. Shen, C. Zhu, Cancer Letters 320, 96-103 (2012).
[7] N. Niu, D.J. Schaid, R.P. Abo, K. Kalari, B. L. Fridley, Q. Feng, G. Jenkins, A. Batzler, Abra G Brisbin, J. M Cunningham, L. Li, Z. Sun, P. Yang, L. Wang, BMC Cancer, 12, 422 (2012).
[8] D. K. Agrawal, Z. Shao, Curr Allergy Asthma Rep 10(1), 39-48 (2010).
[9] C. Yang, E. Bolotin, T. Jiang, F. Sladek, E. Martinez, Gene 389, 52-65 (2007).
[10] P. Hsu, C. Chao, C. Yang, Y. Yea, F. Liub, Y. Chuangb, C.Lana, Eukaryot Cell 12, 804-815 (2013).
[11] L.J. Ko, J.D. Engel, Mol Cell Biol 13(7), 4011-4022 (1993).
[12] T.A. Blauwkamp, M.V. Chang, K.M. Cadigan, EMBO J 27(10), 1436-1446 (2008).
[13] T.P. Shkurat, N.S. Ponomareva, A.A. Aleksandrova, M.A. Shkurat, A.I. Butenko, A.E. Panich, Open Journal of Genetics 2, 1-4 (2012).
[14] M. Gardiner-Garden, M. Frommer, J. Mol. Biol. 196, 261-282 (1987).
[15] R. Ornstein, R. Rein, D. Breen, R. Macelroy, Biopolymers 17, 2341-2360 (1978).
[16] T. Vesth, K. Lagesen, O. Acar, D. Ussery, PLoS ONE 8(4), e60120 (2013).
[17] P.A. Ginno, P.L. Lott, H.C. Christensen, I. Korf, F. Chdin, Mol Cell 45(6), 814-825 (2012).
[18] W.K. Hardle and L. Simar. Applied Multivariate Statistical Analysis, 3rd edition, Springer, (2012)
[19] IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY, IBM Corp, (2013).
[20] D. Liao, Mol Cancer Res. 7(12),1893-901 (2009).
[21] E. Tchougounova, Y. Jiang, D. Brster, N Lindberg, M. Kastemar, A. Asplund, B. Westermark, L. Uhrbom, Oncogene 28, 1537-1548 (2009.)

Mohammed A. Ibrahim Al-Obaide is a Professor of Molecular Biology and Biotechnology, obtained his Ph.D. in Molecular Biology from the University of East Anglia/England-U.K. in July 1976. Before joining the TTUHSC in September 2013, Prof. Al-Obaide worked for seven years as Professor, Consultant and Researcher of molecular biology and biotechnology within Royal Scientific Society and Princess Sumaya University for Technology in Jordan, 21 years in the Departments of Biology and Biotechnology within Al-Nahrain, Baghdad and Sulaimanya Universities and nine years within Scientific Research Council in Iraq. Currently, his research interest is focused on genomics, epigenomics and bioinformatics of cancer and biological systems. Al-Obaide supervised eight Ph.D. and 43 M.Sc. students, published 86 research papers and seven books and granted four patents.

Abdel-Salam Gomaa Abdel-Salam holds BS and MS (2004) degrees in Statistics from Cairo University and MS (2006) and PhD (2009) degrees in Statistics from Virginia Polytechnic Institute and State University (Virginia Tech, USA). Dr. Abdel-Salam has published several research papers and delivered numerous talks and workshops. He has awarded couples of the highest prestige awards such as Teaching

© 2016 NSP
Natural Sciences Publishing Cor.

Excellence from Virginia Tech, Academic Excellence Award, Freud International Award, and Mary G. Natrella Scholarship from American Statistical Association (ASA) and American Society for Quality (ASQ), for outstanding graduate study of the theory and application of Quality Control, Quality Assurance, Quality Improvement, and Total Quality Management. He is a member of the ASQ and ASA. Also, He has received several funds such as UREP, CWSP, and the start-up grants from Qatar University (2014/15) and The Cairo University Award for international publication in 2014. His research interests include all aspects of industrial statistics and Economic Capital Models, including statistical process control, multivariate analysis, Regression Analysis, Exploratory and Robust Data Analysis, mixed models, Nonparametric and Semiparametric Profile Monitoring, Health-related monitoring and prospective public health surveillance.

Hytham W. Alobydi completed his B.Sc. and M.Sc. degrees in Medical Engineering in ALNAHRAIN UNIVERSITY, Baghdad, Iraq. Moved to Amman Jordan 2007 and worked as an IT Consultant on Matlab and Web Technologies. Worked with FORD-USA from 2010 to 2012 as a Matlab Specialist, in 2012 he joined SPAR INC as a Web Developer. Also he is lead developer in BIOMEDICA LLC to provide state of the art programs and software that assist scientists, biologists, physicians and pharmacists in their research work. He is Co-Founder of Iraq Association of Medical Engineering Technology (IAMET). His interest is focused on applying computational analysis to the fields of bioinformatics and biomedical research.

Kalkunte S. Srivenugopal, PhD is a tenured full Professor in the Department of Biomedical Sciences at Texas Tech University Health Sciences Center School of Pharmacy, Amarillo, Texas. Dr. Srivenugopal received his Ph.D. in Biochemistry from the Indian Institute of Science, Bangalore, India in 1981. Following postdoctoral training at the University of Washington, he held faculty positions at the Rosalind Franklin University of Medicine and science, University of Medicine and Dentistry of New Jersey and the University of Texas M.D. Anderson Cancer Center before joining the Texas Tech HSC. Dr. Srivenugopal's research interests include anticancer drug discovery, tumor resistance focusing on the DNA-repair protein O6-methylguanine-DNA methyltransferase (MGMT), redox stress in cancers and regulation of p53 function. He has published over 75 peer-reviewed articles and has been funded in the past by the American Cancer Society, US National Institutes of Health (R29, RO1 and RO3).

© 2016 NSP
Natural Sciences Publishing Cor.