



ANALYSIS OF THE MONO-, DI-, AND TRINUCLEOTIDE TANDEM REPEATS IN INF2 OF GLOMERULAR PODOCYTE

SIM-HUI TEE

Multimedia University, Cyberjaya, Malaysia

E-mail: shtee@mmu.edu.my

ABSTRACT

Inverted formin 2 (INF2) is an essential gene that is expressed in glomerular podocyte in kidney. Malfunction in INF2 will result in kidney diseases, which could be avoided if the mechanisms of its transcription and translation are well-studied. This research aims to use a computational approach to understand the tandem repeats, which is a phenomenon closely linked with diseases, in this gene. A genome database, web server, and a computational motif model were primarily used in this research. The results of this study provide insights that will guide biologists and biomedical practitioners in molecular studies and drug designs for the INF2-implicated diseases.

Keywords: *Bioinformatics, Java Programming, Database, Tandem Repeats, INF2, Glomerular Podocyte, Gene Analysis*

1. INTRODUCTION

Glomerular podocytes are unique renal cell with a complex cytoarchitecture. Coupled with glomerular basement membrane, they form the glomerular filtration barrier which prevents protein leakage into urine filtrates [1]. Though it was postulated that glomerular podocytes play an essential role in establishing the specific permeability properties and cleaning of the glomerular filter, the detailed mechanisms are remained unknown [2]. However, there is a common understanding that the aberration and failure of glomerular podocytes may culminate in the renal diseases, such as collapsing glomerulopathy [3] and focal-segmental glomerulosclerosis [4]. The common characteristics of the pathobiology of glomerular podocytes include the alteration in the molecular composition of the slit diaphragm and the reorganization of the foot processes [5]. Understanding of the molecular mechanisms of the pathogenesis of glomerular podocytes is important for the prevention and diagnosis of the renal diseases.

In this study, a bioinformatics approach has been carried out to investigate the genetic characteristics of glomerular podocyte. Computational tools that have been applied include database, web server, and a computational motif model which was executed in Java programming.

Inverted formin 2 (INF2) gene, which is located on chromosome 14 of human genome, is the most crucial gene that is expressed in glomerular podocyte. The tandem repeats of mono-, di-, and trinucleotide of INF2 will be analyzed to study the potential aberration of the glomerular podocyte. Tandem repeats are gene motifs that occur in a repeating fashion, which is ubiquitous in organisms and they can occur at any location on a gene [6]. Computational methods represent an efficient approach in the analysis of tandem repeats because of the enormous volume of genetic data. It is not surprising that computational tools such as genome software [7-10], computational simulation [11-18], graphical representation [19-23], databases [24-28], web servers [29-32] and computer modeling [33-35] have been used extensively by computer scientists and geneticists in the studies of genetic data. These computational tools allow scientists to understand dynamic biological processes [36], and have practical applications in the biomedical sectors, such as microarray analysis of genes in medical informatics [37] and the identification of gene expression signatures in clinical settings [38-41].

2. METHODS

The nucleotide sequences of the studied gene, INF2, were retrieved in the FASTA format from the GenBank of NCBI. The UCSC Genome Browser database [47] was used to visualize the loci of INF2 gene on the chromosome. KOBAS 2.0 web server [48] was used to search for the structural motifs of INF2 gene and the annotated

disease pathway. A combinatorial-based *mreps* algorithm [42] was used to identify the repeated nucleotides. A motif model [43] was used to identify the repeating motif (Θ^M) and non-motif (Θ^B). This computational motif model was executed using Java Programming.

$$\Theta = \{\Theta^B, \Theta^M\} = \begin{bmatrix} \theta_{-0}^B \theta_{-1}^M \theta_{-2}^M \dots \theta_{-w}^M \\ \theta_{a1,0}^B \theta_{a1,1}^M \theta_{a1,2}^M \dots \theta_{a1,w}^M \\ \theta_{a2,0}^B \theta_{a2,1}^M \theta_{a2,2}^M \dots \theta_{a2,w}^M \\ \dots \dots \dots \\ \theta_{aj,0}^B \theta_{aj,1}^M \theta_{aj,2}^M \dots \theta_{aj,w}^M \end{bmatrix} \quad (1)$$

Let S^{ij} be the subsequence of length W at position j in a sequence i . Let a be the symbol that occurs at a position k of either Θ^M or Θ^B ; let the position k be $1 \leq k \leq W$, and L be the set for the length of nucleotide sequence. The conditional probabilities that S^{ij} is found using the motif model are computed as such [43]:

$$P_M(S^{ij}) = \prod_{k=1}^W \prod_{a=1}^L (\theta_{ak}^M)^{I(S_{i,j+k-1}=a)} \quad (2)$$

Whereas the conditional probabilities that S^{ij} is found using the non-motif model are computed by [43]:

$$P_B(S^{ij}) = \prod_{k=1}^W \prod_{a=1}^L (\theta_{a0}^B)^{I(S_{i,j+k-1}=a)} \quad (3)$$

Let λ be the prior probability of motif occurrence in the gene sequences. The motif occurrence probability Z at position j in sequence i is derived from [43]:

$$Z_{ij} = \frac{\lambda P_M S^{ij}}{(\lambda P_M S^{ij}) + (1 - \lambda) P_B S^{ij}} \quad (4)$$

In this research, S^{ij} is taken to be a genuine motif hit when the following is fulfilled [43]:

$$\log\left(\frac{P_M(S^{ij})}{P_B(S^{ij})}\right) \geq \log[(1 - \lambda) / \lambda] \quad (5)$$

Because of the complexity of motif, the pseudo-count of motif is likely to happen. We identified pseudo-count as [43]:

$$p_{sc_{ak}} = \sum_{b=1}^L \theta_{b,k+1}^M P_{a/b} \quad (6)$$

Where $P_{a/b}$ is the BLOSUM substitution probability for amino acid a (triplet) from the observed amino acid b (triplet). Mono-, di-, and trinucleotide tandem repeat motifs with the minimum repeat number of 6, 2, 2, respectively, was analyzed. Relative frequency was used to analyze the total repeat per kilobase (kb) in the nucleotide sequence of INF2 gene. A triplet classification system [44] was used to categorize the trinucleotide tandem repeats. Besides, DNA Analysis Server [46] was used to validate the structural parameters of INF2 gene.

3. RESULTS AND DISCUSSION

INF2 gene is constituted by 11112 base pairs (bp) of nucleotides with 19.51% Adenine, 15.35% Thymine, 30.84% Guanine, and 34.30% Cytosine. It is apparently a GC-rich gene. However, in general there is no correlation between the gene motif's GC nucleotide percentage and its length [45].

Using the UCSC Genome Browser database [47], INF2 gene was visualized on the locus of Chromosome 14. Figure 1 illustrates the graphical position of INF2 gene.

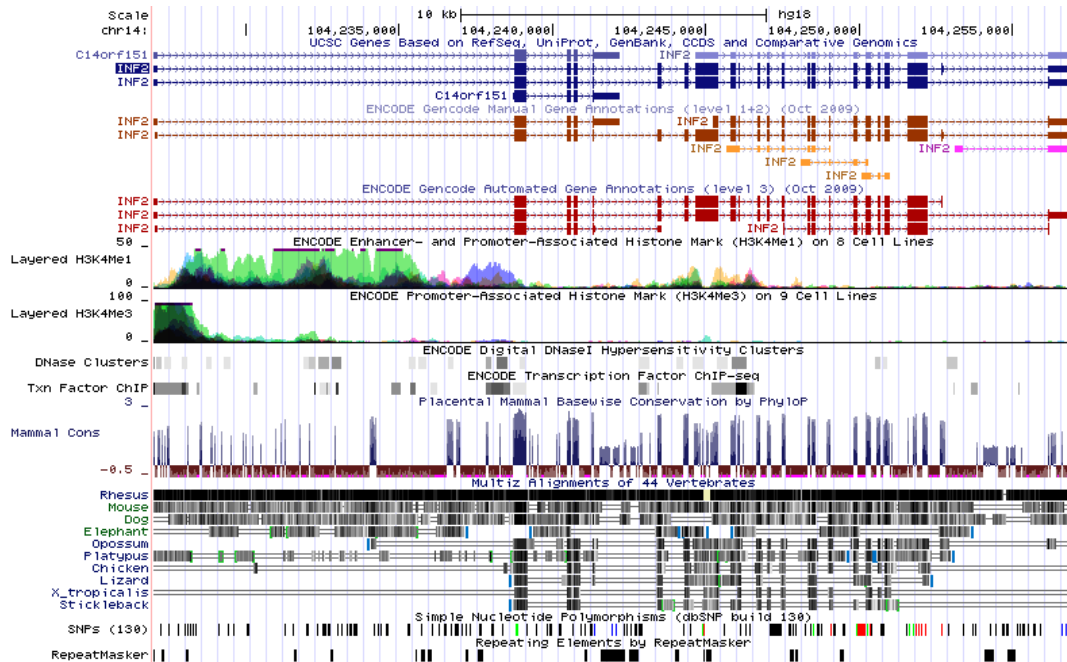


Figure 1. The Position Of INF2 On Chromosome 14.

We queried INF2 gene using KOBAS 2.0 web server [48] for the structural motifs and the disease pathways. Figure 2 illustrates the sample query result for 3 transcript variants of INF2. The results

show that INF2 is implicated in the urinary and reproductive diseases. Glomerulosclerosis, which is a kidney disease, is INF2-implicated.

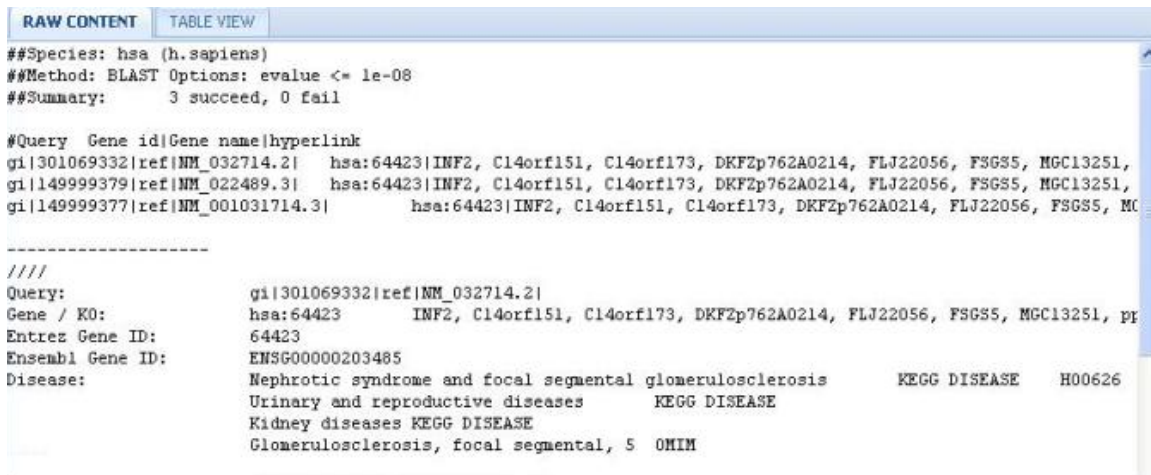


Figure 2. Query Result For INF2 Gene

Besides, the structural motifs (which are also functional motifs) of INF2 were identified using KOBAS 2.0. It was found that the four motifs of INF2 are located at the vicinity of 5' region of the gene. These motifs are listed in Table 1.



Table 1. Motifs Of INF2

Motif ID	From	To	Definition	E-value
pf:Drf_GBD	23	152	Diaphanous GTPase-binding Domain	5.2e-17
pf:Drf_FH3	156	343	Diaphanous FH3 Domain	1.2e-38
pf:FH2	556	920	Formin Homology 2 Domain	7.9e-75
pf:WH2	974	989	WH2 motif	8e-5

A total of 482 tandem repeats were identified for mono-, di-, and trinucleotide motifs. The interspersed repeats span a length of 657 bp, which is equivalent to 5.91% of the total length of INF2 gene. The length of simple repeat is 54 bp, which is equivalent to 0.49% of the total length of INF2 gene. The distribution of tandem repeats of INF2 gene is given in Table 2.

Table 2. Distribution Of Tandem Repeats Of INF2

Repeat motif	No. of occurrence	Relative frequency
Mononucleotide:	16	1.44
Dinucleotide:	280	25.20
Trinucleotide	186	16.74

As shown in Table 2, the relative frequency of mononucleotide tandem repeats is quite low as compared to dinucleotide and trinucleotide tandem repeats. None of the 16 mononucleotide tandem repeats is falling within the functional motifs which are listed in Table 1, neither does any one of the mononucleotide tandem repeat is found in the vicinity of the functional motifs. The vicinity is defined as 100 bp upstream and downstream of the functional motifs. Of 280 dinucleotide tandem repeats there are 14 repeats in vicinity, whereas 19 repeats are falling within one of the functional motifs listed in Table 1. Of 186 trinucleotide tandem repeats there are 5 repeats found in the vicinity of the functional motifs, whereas 8 repeats are found within the functional motifs. Because the number of tandem repeats that is falling within the functional motifs is scarce, the impact of the polymorphism of these tandem repeat motifs on the functional motifs may not be significant. As it is known that tandem repeat regions are highly polymorphic [49], the less abundance of mono-, di-, and trinucleotide tandem repeats in the functional motifs suggests that the mutation of these sequences is less likely to contribute to the aberration of INF2 gene. It also implies that the

mutation of these tandem repeats is not closely associated to the diseases of glomerular podocytes in kidney.

Because trinucleotides constitute amino acids in the coding region, it is important to further examine their distribution pattern in INF2 gene. Table 3 displays the trinucleotide tandem repeats categorized according to a triplet classification system [44].

Table 3. Distribution Of Trinucleotide Tandem Repeat

Class	Repeat motifs (with frequency of occurrence)	Total	Relative Freq
T1	AAT(0) ATA(0) TAA(0) ATT(0) TTA(0) TAT(0)	0	0
T2	AAG(2) AGA(0) GAA(7) CTT(2) TTC(2) TCT(2)	15	1.35
T3	AAC(2) ACA(2) CAA(0) GTT(0) TTG(0) TGT(0)	4	0.36
T4	ATG(0) TGA(2) GAT(0) CAT(0) ATC(0) TCA(0)	2	0.18
T5	AGT(0) GTA(0) TAG(0) ACT(5) CTA(0) TAC(0)	5	0.45

T6	AGG(7) GGA(16) GAG(8) CCT(13) CTC(4) TCC(4)	52	4.68
T7	AGC(11) GCA(8) CAG(10) GCT(14) CTG(14) TGC(10)	67	6.03
T8	ACG(0) CGA(0) GAC(0) CGT(0) GTC(0) TCG(0)	0	0
T9	ACC(4) CCA(6) CAC(9) GGT(3) GTG(2) TGG(2)	26	2.34
T10	GGC(6) GCG(0) CGG(5) GCC(4) CCG(0) CGC(0)	15	1.35

It is striking that quite a number of trinucleotide classes are either exhibiting low abundance of tandem repeat motifs (e.g., class T2, T3, T4, T5 and T10) or do not have motifs presented in INF2 gene (e.g., T1 and T8). As each class of tandem repeat consists of six member trinucleotide tandem repeat motifs, T1 and T8 which do not have any occurrence of motifs implies that all of their member repeat motifs do not present in the gene. Across 60 repeat motifs from class T1 to T10, there are 30 repeat motifs (50%) do not occur in INF2 gene. This explains the observed low relative frequency (the last column in Table 3) of trinucleotide tandem repeats in INF2 gene. Class T7, which has the highest value of relative frequency, constitutes amino acid Ser (AGC), Ala (GCA and GCT), Gln (CAG), Leu (CTG), and Cys (TGC) in the coding region. Mutations of these codons in the coding region may have impacts on the protein expression of INF2 gene.

DNA Analysis Server [46] was used to validate the structural parameters of INF2 gene. We used *plot.it* server of the DNA Analysis Server to

visualize the parametric plot of physicochemical and statistical parameters. The structural properties of trinucleotide and dinucleotide have been verified. The verified parameters (DNA rigidity and bendability) for trinucleotide is shown in Figure 3.

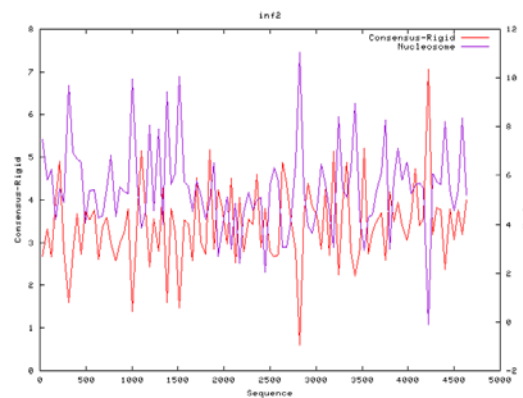


Figure 3. Trinucleotide Rigidity And Bendability

Figure 3 shows a normal distribution of the extent of nucleotide rigidity and bendability across INF2 gene. In general, the diverging values are not very large to the extent that the interaction between proteins and gene would be prevented. Similar distribution of the parameter values was observed for the roll angle and tilt angle of dinucleotide, as displayed in Figure 4.

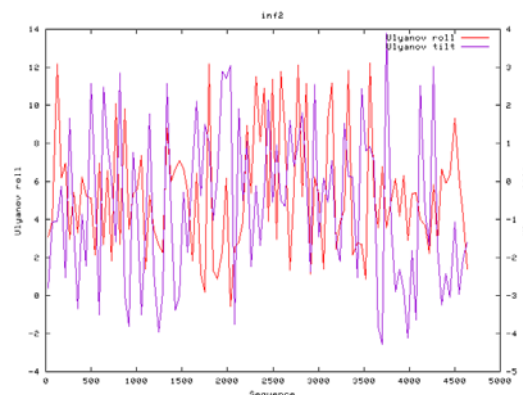


Figure 4. Roll And Tilt Angle Of Dinucleotide

Although Figure 4 shows that the roll angle and tilt angle of dinucleotide are normal, it was observed that the largest diverging value of tilt angle occurs at 3500-4000bp. However, we did not observe the aberrant patterns of tandem repeat in this region.

4. CONCLUSION

This study analyzes the mono-, di-, and trinucleotide tandem repeats in INF2 gene of



glomerular podocyte. It was found that the relative frequency of the mononucleotide tandem repeats is very much lower than that of dinucleotide and trinucleotide tandem repeats. In addition, it was found that the number of tandem repeats that is falling within the functional motifs is scarce, implying that the impact of the polymorphism of these tandem repeat motifs may not be significant. The mutation of these tandem repeats is not closely associated to the diseases of glomerular podocytes in kidney. However, the analysis on the distribution of trinucleotide tandem repeats demonstrates that the repeat motifs in class T7 may have implications in the protein expression of INF2 gene in their mutated forms.

REFERENCES

- [1] S. Somlo and P. Mundel, "Getting a foothold in nephrotic syndrome", *Nature Genetics*, Vol. 24, 2000, pp. 333-335.
- [2] H. Pavenstädt, W. Kriz, and M. Kretzler, "Cell biology of the glomerular podocyte", *Physiological Reviews*, Vol. 83, 2003, pp. 253-307.
- [3] M. Nagata, M. Hattori, Y. Hamano, K. Ito, K. Saitoh, and T. Watanabe, "Origin and phenotypic features of hyperplastic epithelial cells in collapsing glomerulopathy", *American Journal of Kidney Diseases*, Vol. 32, 1998, pp. 962-969.
- [4] V. D'agati, "The many masks of focal segmental glomerulosclerosis", *Kidney International*, Vol. 46, 1994, pp. 1223-1241.
- [5] K. Asanuma and P. Mundel, "The role of podocytes in glomerular pathobiology", *Clinical and Experimental Nephrology*, Vol. 7, 2003, pp. 255-259.
- [6] A.J. Hannan, "Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'", *Trends in Genetics*, Vol. 26, No. 2, 2009, pp. 59-65.
- [7] S. Hoehme and D. Drasdo, "A cell-based simulation software for multi-cellular systems", *Bioinformatics*, Vol. 26, No. 20, 2010, pp. 2641-2642.
- [8] N. Goto, P. Prins, M. Nakao, R. Bonnal, J. Aerts, and T. Katayama, "BioRuby: bioinformatics software for the Ruby programming language", *Bioinformatics*, Vol. 26, No. 20, 2010, pp. 2617-2619.
- [9] S. Costantini, G. Colonna, and A.M. Facchiano, "PreSSAPro: A software for the prediction of secondary structure by amino acid properties", *Computational Biology and Chemistry*, Vol. 31, 2007, pp. 389-392.
- [10] A.G. Gonzalez, A. Naldi, L. Sánchez, D. Thieffry, and C. Chaouiya, "GINsim: A software suite for the qualitative modelling, simulation and analysis of regulatory networks", *BioSystems*, Vol. 84, 2006, pp. 91-100.
- [11] Y. Cheng, W. Cui, Q. Chen, C-H. Tung, M. Ji, and F. Zhang, "The molecular mechanism studies of chirality effect of PHA-739358 on Aurora kinase A by molecular dynamics simulation and free energy calculations", *Journal of Computer-Aided Molecular Design*, Vol. 25, No. 2, 2011, pp. 171-180.
- [12] C. Hallmen and M. Wiese, "Molecular dynamics simulation of the human adenosine A₃ receptor: agonist induced conformational changes of Trp243", *Journal of Computer-Aided Molecular Design*, Vol. 20, 2006, pp. 673-684.
- [13] D. Lama and R. Sankararamakrishnan, "Molecular dynamics simulations of pro-apoptotic BH3 peptide helices in aqueous medium: relationship between helix stability and their binding affinities to the anti-apoptotic protein Bcl-X_L", *Journal of Computer-Aided Molecular Design*, Vol. 25, 2011, pp. 413-426.
- [14] I. Eberini, A. Emerson, C. Sensi, L. Ragona, P. Ricchiuto, A. Pedretti, E. Gianazza, and A. Tramontano, "Simulation of urea-induced protein unfolding: A lesson from bovine β -lactoglobulin", *Journal of Molecular Graphics and Modelling*, Vol. 30, 2011, pp. 24-30.
- [15] H. Park, S. Ko, and Y.H. Jeon, "Force field design and molecular dynamics simulations of factor-inhibiting HIF-1 and its complex with known inhibitors: Implications for rational inhibitor design", *Journal of Molecular Graphics and Modelling*, Vol. 29, 2010, pp. 221-228.
- [16] N.S. Gandhi and R.L. Mancera, "Molecular dynamics simulations of CXCL-8 and its interactions with a receptor peptide, heparin fragments, and sulfated linked cyclitols", *Journal of Chemical Information and Modeling*, Vol. 51, 2011, pp.335-358.
- [17] M. Münz and P.C. Biggin, "JGromacs: A Java package for analyzing protein simulations", *Journal of Chemical Information and Modeling*, Vol. 52, 2012, pp. 255-259.
- [18] J.S. Mitchell, C.A. Loughton, and S.A. Harris, "Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA", *Nucleic*



- Acids Research*, Vol. 39, No. 9, 2011, pp. 3928-3938.
- [19] F. Chevenet, O. Croce, M. Hebrard, R. Christen, and V. Berry, "ScripTree: scripting Phylogenetic graphics", *Bioinformatics*, Vol. 26, No. 8, 2010, pp. 1125-1126.
- [20] K.R. Brown, D. Otasek, M. Ali, M.J. McGuffin, W. Xie, B. Devani, I.L. van Toch, and I. Jurisica, "NAViGaTOR: Network analysis, visualization and graphing Toronto", *Bioinformatics*, Vol. 25, No. 24, 2009, pp. 3327-3329.
- [21] G.H. Weber, O. Rübél, M-Y. Huang, A.H. DePace, C.C.Fowlkes, S.V.E. Keränen, C.L.L. Hendriks, H. Hagen, D.W. Knowles, J. Malik, M.D. Biggin, and B. Hamann, "Visual exploration of three-dimensional gene expression using physical views and linked abstract views", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 6, No. 2, 2009, pp. 296-309.
- [22] J. Kim, H-J. Chung, C.H. Park, W-Y. Park, and J.H. Kim, "ChromoViz: multimodal visualization of gene expression data onto chromosomes using scalable vector graphics", *Bioinformatics*, Vol. 20, No. 7, 2004, pp. 1191-1192.
- [23] M. Spitzer, G. Fuellen, P. Cullen, and S. Lorkowski, "VisCoSe: visualization and comparison of consensus sequences", *Bioinformatics*, Vol. 20, No. 3, 2004, pp. 433-435.
- [24] J.L. van Hemert and J.A. Dickerson, "PathwayAccess: CellDesigner plugins for pathway databases", *Bioinformatics*, Vol. 26, No. 18, 2010, pp. 2345-2346.
- [25] S. Moon, Y. Byun, and K. Han, "FSDB: A frameshift signal database", *Computational Biology and Chemistry*, Vol. 31, 2007, pp. 298-302.
- [26] M. Carrillo-Tripp, C.M. Shepherd, I.A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J.E. Johnson, C.L. Brooks III, and V.S. Reddy, "VIPERdb²: an enhanced and web API enabled relational database for structural virology", *Nucleic Acids Research*, Vol. 37, 2009, pp. D436-D442.
- [27] A. Bhasi, P. Philip, V. Manikandan, and P. Senapathy, "ExDom: an integrated database for comparative analysis of the exon-intron structures of protein domains in eukaryotes", *Nucleic Acids Research*, Vol. 37, 2009, pp. D703-D711.
- [28] E. Portales-Casamar, D. Arenillas, J. Lim, M.I. Swanson, S. Jiang, A. McCallum, S. Kirov, and W.W. Wasserman, "The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences", *Nucleic Acids Research*, Vol. 37, 2009, pp. D54-D60.
- [29] I. Gabdank, D. Barash, and E.N. Trifonov, "FineStr: a web server for single-base-resolution nucleosome positioning", *Bioinformatics*, Vol. 26, No. 6, 2010, pp. 845-846.
- [30] L.J. McGuffin, "The ModFOLD server for the quality assessment of protein structural models", *Bioinformatics*, Vol. 24, No. 4, 2008, pp. 586-587.
- [31] R. Rawi, L. Whitmore, and M. Topf, "CHOYCE: a web server for constrained homology modelling with cryoEM maps", *Bioinformatics*, Vol. 26, No. 13, 2010, pp. 1673-1674.
- [32] B. Rost and J. Liu, "The PredictProtein server", *Nucleic Acids Research*, Vol. 31, No. 13, 2003, pp. 3300-3304.
- [33] J. Eldstrom and D. Fedida, "Modeling of high-affinity binding of the novel atrial anti-arrhythmic agent, vernakalant, to Kv1.5 channels", *Journal of Molecular Graphics and Modelling*, Vol. 28, 2009, pp. 226-235.
- [34] A. Almond, P.L. DeAngelis, and C.D. Blundell, "Hyaluronan: the local solution conformation determined by NMR and computer modeling is close to a contracted left-handed 4-fold helix", *Journal of Molecular Biology*, Vol. 358, 2006, pp. 1256-1269.
- [35] P. Makkar, R.P.R. Metpally, S. Sangadala, B.V.B. Reddy, "Modeling and analysis of MH1 domain of Smads and their interaction with promoter DNA sequence motif", *Journal of Molecular Graphics and Modelling*, Vol. 27, 2009, pp. 803-812.
- [36] A.H.K. Roeder, P.T. Tarr, C. Tobin, X. Zhang, V. Chickarmane, A. Cunha, and E.M. Meyerowitz, "Computational morphodynamics of plants: integrating development over space and time", *Nature Reviews Molecular Cell Biology*, Vol. 12, 2011, pp. 265-273.
- [37] F. Ferrazzi, P. Magni, L. Sacchi, A. Nuzzo, U. Petrovič, and R. Bellazzi, "Inferring gene regulatory networks by integrating static and dynamic data", *International Journal of Medical Informatics*, Vol. 76S, 2007, pp. S462-S475.
- [38] P.R. Bushel, H.K. Hamadeh, L. Bennett, J. Green, A. Ableson, S. Misener, C.A. Afshari, and R.S. Paules, "Computational selection of



- distinct class- and subclass-specific gene expression signatures”, *Journal of Biomedical Informatics*, Vol. 35, 2003, pp. 160-170.
- [39] R.S. Stearman, L. Dwyer-Nield, M.C. Grady, A.M. Malkinson, and M.W. Geraci, “A macrophage gene expression signature defines a field effect in the lung tumor microenvironment”, *Cancer Research*, Vol. 68, No. 1, 2008, pp. 34-43.
- [40] K. Oshima, Y. Naoi, K. Kishi, Y. Nakamura, T. Iwamoto, K. Shimazu, T. Nakayama, S.J. Kim, Y. Baba, Y. Tamaki, and S. Noguchi, “Gene expression signature of *TP53* but not its mutation status predicts response to sequential paclitaxel and 5-FU/epirubicin/cyclophosphamide in human breast cancer”, *Cancer Letters*, Vol. 307, 2011, pp. 149-157.
- [41] H. Sasaki, S. Shimizu, K. Okuda, O. Kawano, H. Yukiue, M. Yano, and Y. Fujii, “Epidermal growth factor receptor gene amplification in surgical resected Japanese lung cancer”, *Lung Cancer*, Vol. 64, 2009, pp. 295-300.
- [42] R. Kolpakov, G. Bana, and G. Kucherov, “mreps: efficient and flexible detection of tandem repeats in DNA”, *Nucleic Acids Research*, Vol. 31, No. 13, 2003, pp. 3672-3678.
- [43] T. Le, T. Altman, and K. Gardiner, “HIGEDA: a hierarchical gene-set genetics based algorithm for finding subtle motifs in biological sequences”, *Bioinformatics*, Vol. 26, No. 3, 2010, pp. 302-309.
- [44] J. Jurka and C. Pethiyagoda, “Simple repetitive DNA sequences from primates: compilation and analysis”, *Journal of Molecular Evolution*, Vol. 40, 1995, pp. 120-126.
- [45] R. Jaksik and J. Rzeszowska-Wolny, “The distribution of GC nucleotides and regulatory sequence motifs in genes and their adjacent sequences”, *Gene*, Vol. 492, 2012, pp. 375-381.
- [46] K. Vlahoviček, L. Kaján, and S. Pongor, “DNA analysis servers: plot.it, bend.it, model.it and IS”, *Nucleic Acids Research*, Vol. 31, 2003, pp. 3686-3687.
- [47] P.A. Fujita, B. Rhead, A.S. Zweig, A.S. Hinrichs, D. Karolchik, M.S. Cline, M. Goldman, G.P. Barber, H. Clawson, A. Coelho, M. Diekhans, T.R. Dreszer, B.M. Gardine, R.A. Harte, J. Hillman-Jackson, F. Hsu, V. Kirkup, R.M. Kuhn, K. Learned, C.H. Li, L.R. Meyer, A. Pohl, B.J. Raney, K.R. Rosenbloom, K.E. Smith, D. Haussler, and W.J. Kent, “The UCSC Genome Browser database: update 2011”, *Nucleic Acids Research*, Vol. 39, 2011, pp. D876-D882.
- [48] C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C-Y. Li, and L. Wei, “KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases”, *Nucleic Acids Research*, Vol. 39, 2011, pp. W316-W322.
- [49] H. Ellegren, “Microsatellite mutations in the germline: implications for evolutionary inference”, *Trends in Genetics*, Vol. 16, 2000, pp. 551-558.