Communications in Information and Systems
Volume 18, Number 3, 149–171, 2018

Wavelet analysis of candidate genes for diabetic retinopathy

En-Bing Lin

Dedicated to the memory of my parents
Dr. and Mrs. Chen-Chuan Lin

Diabetic retinopathy, a major cause of adult blindness, is a medical condition in which damage occurs to the retina due to diabetes mellitus. Based on the Diaretinopathy database, we perform wavelet analysis on several candidate genes responsible for causing diabetic retinopathy. We obtain approximation and detail information of the numerical representations of candidate genes. We compute discrete and continuous wavelet transforms of each gene. We compare the computational and graphical results of seven candidate genes. We also perform wavelet analysis and wavelet transform on the Fibroblast Growth Factor 21 gene which can be included in the gene therapy. Through this study, it is anticipated to provide better disease management and ensure better prognosis in diabetic retinopathy.

1. Introduction

Diabetic retinopathy is a damaged retina in the eye caused by weak blood vessels as a result of diabetes. It is a progressive microvascular complication associated with both type 1 and type 2 diabetes, with the latter being more prone to disease susceptibility. Diabetic retinopathy is composed of a characteristic group of lesions found in the retina of individuals having had diabetes mellitus for several years. It is a major cause of non-inherited blindness among adults around the world. It affects the retina of almost all patients with type I diabetes and about 60% of those with type II diabetes [11]. It is estimated that in 2002 diabetic retinopathy accounted for about 5% of world blindness, representing almost 5 million blind people [49]. Although many diabetic retinopathy articles have been published, the understanding of the underlying molecular mechanisms of diabetic retinopathy is limited. Many genes play a crucial role in causing the disease. Diaretinopathy database [48],

a gene database for diabetic retinopathy, provides information of 102 potential candidate genes causing diabetic retinopathy at molecular, biomedical and at structural level. For each candidate gene the database is designed by taking 24 parameters into consideration that comprises official symbol, alternative names, description, chromosome map showing the location, number of exons and GT-AG introns, motif, polymorphic variation, Enzyme Commission number, catalytic activity, active site, cofactor, biophysicochemical properties, enzyme regulation, induction, molecular pathway, interactors, post-translational modification, and 3D structure. In addition to the molecular class and function of these genes, this database also provides links to download the corresponding nucleotide and amino acid sequences in FASTA format from NCBI and UNIPROT database respectively [36], which may further be used for their computational approaches. Worldwide, the prevalence of diabetic retinopathy is increasing at an alarming rate, the World Health Organization predicted that the number of adults with diabetes in the world, would be from 171 million in 2000 to 366 million in 2030. Translating basic biomedical research into clinical practice is vital to our health care services. It will not only influence global health but also impact the world economy. This paper is mainly to analyze several candidate genes in the diaretinopathy database. In what follows, we will provide some motivations of our methodologies and approaches of our intended studies.

One of the primary goals of bioinformatics is to extract valuable information from a large amount of biological data. Several clustering and other techniques are applied to DNA (Deoxyribo Nucleic Acid) and protein sequences [6, 9, 16, 37, 44], by which one can correlate the inherent relationships between DNA or protein sequences. In fact, digital signal processing techniques can be used to characterize genomic data more efficiently in comparison to other methods [2–5, 7, 8, 13, 41, 47]. In particular, one of the useful tools in signal processing, wavelet techniques have been used in various applications in bioinformatics and medical areas [1, 14–22, 24–33, 35, 38, 43, 45]. This is mainly due to robustness, efficiency, and flexibility of wavelets' characteristics [34]. We address the DNA code of the human being in the perspective of signal processing. It is motivated by the adoption of the wavelets for the study of the DNA information. DNA is a double helix constituted by two polymers connected by hydrogen atoms. The polymers contain three types of nucleotides, namely deoxyribose, a phosphate group, and a nitrogenous base. There are four distinct nitrogenous bases: thymine, cytosine, adenine, and guanine, denoted by the symbols {T, C, A, G}. To apply wavelet techniques, we need to map DNA sequences into mathematical representations, which include binary coding [40, 50], complex number [7],

integer number [12], EIIP (electron-ion interaction potential) [33], Z-curves [52, 53]. Other models, such as DNA walks [39], are also available. The integer representation appears to be useful and effective [42]. In this paper, we will use the integer representation, namely, we map integer numbers to the four nucleotides as T=0, C=1, A=2, and G=3. Based on the work done by Saini and Dewan [42], we will perform wavelet analysis on several candidate genes by using Daubechies (db1) wavelet at level 1 which has least reconstruction errors in comparison with many other situations.

Our paper is organized as follows. In Section 2, we provide some backgrounds in wavelet analysis. We then perform wavelet analysis on several candidate genes in Section 3. Some comparison results are presented in Section 4. We conclude with several comments in Section 5.

2. Preliminary backgrounds

In this section, we give a brief account of wavelet analysis by recalling multiresolution analysis, scaling functions, wavelet functions, as well as continuous and discrete wavelet transforms.

A multiresolution analysis (MRA) [34] consists of a sequence of successive approximation spaces $\{V_j\}_{j\in Z}$ of $L^2(R)$ with the following properties:

(i): $V_j \subset V_{j+1}$,

(ii): $\lim_{j\to\infty} V_j = \bigcup_{j\in Z} V_j$ is dense in $L^2(R)$,

(iii): $\bigcap_{j\in Z} V_j = \{0\}$,

(iv): $f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j+1}$,

(v): $f(x) \in V_j \Longleftrightarrow f(x + 2^{-j}k) \in V_j, \ \ \forall k \in Z$,

(vi): There exists a function $\phi \in V_0$ so that $\{\phi(x-j)\}_{j\in Z}$ is an orthonormal basis of V_0.

ϕ is called a *scaling function* that generates a MRA with the above properties. Through translation and dilation of ϕ, a Riesz basis $\{\phi_{j,k}(x)\}_{k\in Z}$ is obtained for the subspace $V_j \subset L^2(R)$ by the properties (iv)(v), where

$$\phi_{j,k}(x) = 2^{\frac{j}{2}}\phi(2^j x - k), \quad j,k \in Z. \tag{1}$$

This family can be generally expressed as $\phi_{m,n}(x) = \frac{1}{a^{\frac{m}{2}}}\phi(\frac{x-nb}{a^m})$.

Since $V_0 \subset V_1$, there is a set of coefficients $\{a_k\}_{k\in Z}$, so that ϕ satisfies the two–scale equation or refinement equation

$$\phi(x) = \sum_k a_k \phi(2x-k). \tag{2}$$

For every $j \in Z$, we define W_j to be the orthonormal complement of V_j in V_{j+1}, we then have

$$V_{j+1} = V_j \bigoplus W_j \tag{3}$$

and

$$W_j \perp W_{j'} \quad if \quad j \neq j'. \tag{4}$$

It follows that, for $j > J$

$$V_j = V_J \bigoplus \left(\bigoplus_{k=0}^{J-j+1} W_{J-k} \right). \tag{5}$$

By virtue of (ii) and (iii) above, this implies

$$L^2(R) = \bigoplus_{j\in Z} W_j \tag{6}$$

which is a decomposition of $L^2(R)$ into mutually orthogonal subspaces. It turns out that a basis for W_0 can be obtained by dilating and translating a single function $\psi(x)$ called basic (mother) wavelet which is defined by (wavelet equation)

$$\psi(x) = \sum_k b_k \phi(2x-k) \tag{7}$$

where $b_k = (-1)^k a_{-k+1}$. In fact, $\{\psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k)\}_{k\in Z}$ forms an orthonormal basis for W_j.

Let $P_j,\ Q_j$ denote the orthogonal projection $L^2 \to V_j,\ L^2 \to W_j$, respectively. Then

$$P_j f(x) = \sum_k \alpha_{j,k}\phi_{j,k}(x), \tag{8}$$

$$Q_j f(x) = \sum_k \beta_{j,k}\psi_{j,k}(x), \tag{9}$$

where the coefficients $\alpha_{j,k},\ \beta_{j,k}$ are given by the inner product:

$$\alpha_{j,k} =< f,\ \phi_{j,k} >= \int_{-\infty}^{\infty} f(x)\phi_{j,k}(x)dx, \tag{10}$$

$$\beta_{j,k} =< f,\ \psi_{j,k} >= \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx. \tag{11}$$

$P_j f$ converges to f in the L^2 norm which is the best approximation of f in V_j.

More precisely, the above coefficients can be obtained by applying wavelet transforms which are defined as follows.

The continuous wavelet trsnsform is defined as:

$$[w_\psi x(t)](a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a}\right) dt \quad a > 0, b \in R, \tag{12}$$

where the symbol * represents the complex conjugate, $x(t)$ is the given signal (DNA sequence) and ψ is a wavelet.

The discrete wavelet transform is defined as:

$$[Dw_\psi x(n)](a,b) = \sum_{n \in Z} x(n) g_{j,k}(n), \quad a = 2^j, b = k2^j, j \in N, k \in Z, \tag{13}$$

where g's are the coefficients of the wavelet equation associated with ψ.

3. Candidate genes

Diabetic retinopathy is a common complication of diabetics. It is important to know more about the underlying molecular mechanisms. Recognizing the relevant genetic susceptibility would help in counseling presymptomatic individuals to adopt preventive and control measures to delay the onset of disease. Therefore it is essential to analyze candidate genes which are responsible for causing diabetic retinopathy. Consequently, it will help progress faster the diagnostic treatment. In what follows, we will perform wavelet analysis of several candidate genes. More precisely, we provide approximation and detail information as well as continuous and discrete transforms of individual genes.

3.1. Angiotensin I converting enzyme (ACE)

This gene encodes an enzyme exopeptidase involved in catalyzing the conversion of angiotensin I into a physiologically active peptide angiotensin II, which is a potent vasopressor and aldosterone-stimulating peptide that controls blood pressure and fluid-electrolyte balance. ACE inhibitors are widely used as pharmaceutical drugs in the treatment of conditions such as high blood pressure, heart failure, diabetic nephropathy, and type 2 diabetes mellitus. Wavelet analysis and wavelet transform of ACE are shown in Figures 1 and 2 respectively.

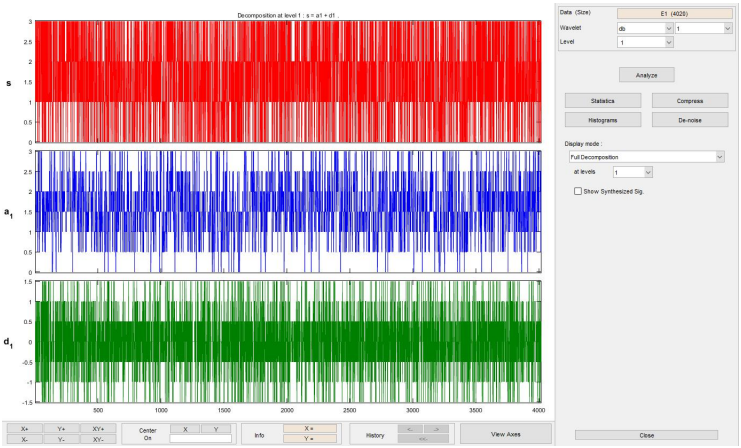

Figure 1: Approximation and Detail of ACE with db1.

3.2. Coiled-coil domain containing 68 (CCDC68)

This gene encodes a member of the hook-related protein family. Members of this family are characterized by an N-terminal potential microtubule binding domain, a central coiled-coiled and a C-terminal Hook-related domain. The encoded protein may be involved in linking organelles to microtubules. Wavelet analysis and wavelet transform of CCDC68 are shown in Figures 3 and 4 respectively.

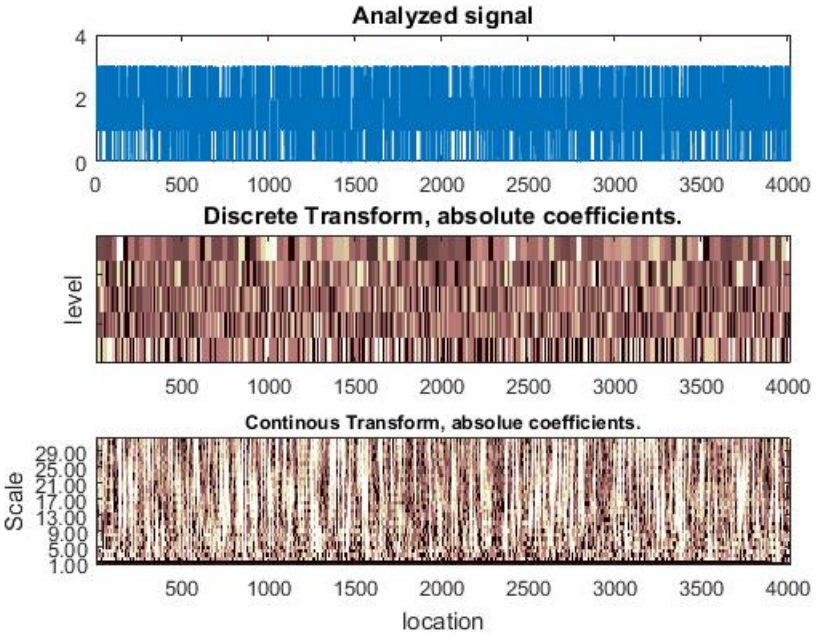

Figure 2: Discrete and Continous Transform of ACE with db1.

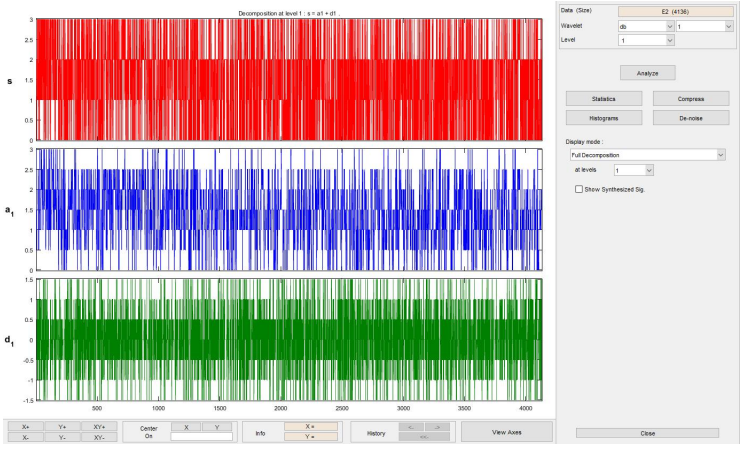

Figure 3: Approximation and Detail of CCDC68 with db1.

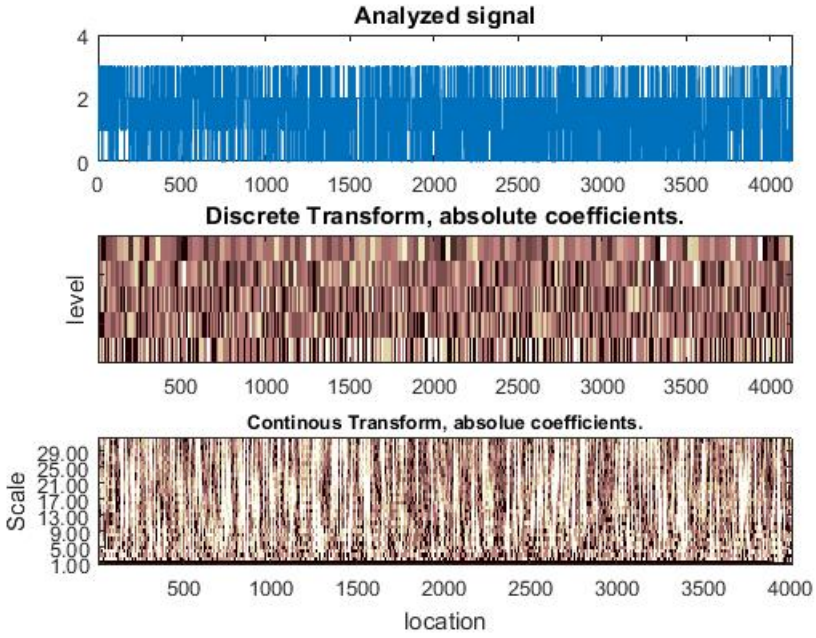

Figure 4: Discrete and Continous Transform of CCDC68 with db1.

3.3. advanced glycosylation end product receptor (AGER)

The advanced glycosylation end product (AGE) receptor encoded by this gene is a member of the immunoglobulin superfamily of cell surface receptors. It is a multiligand receptor interacting with other molecules implicated in homeostasis, development, as well as inflammation, and certain diseases, such as diabetes and Alzheimer's disease. Wavelet analysis and wavelet transform of AGER are shown in Figures 5 and 6 respectively.

3.4. Adiponectin (ADIPOQ)

Adiponectin, an adipose tissue-specific plasma protein, has anti-inflammatory effects on the cellular components of the vascular wall. It is a protein hormone which is involved in regulating glucose levels as well as fatty acid breakdown. Wavelet analysis and wavelet transform of ADIPOQ are shown in Figures 7 and 8 respectively.

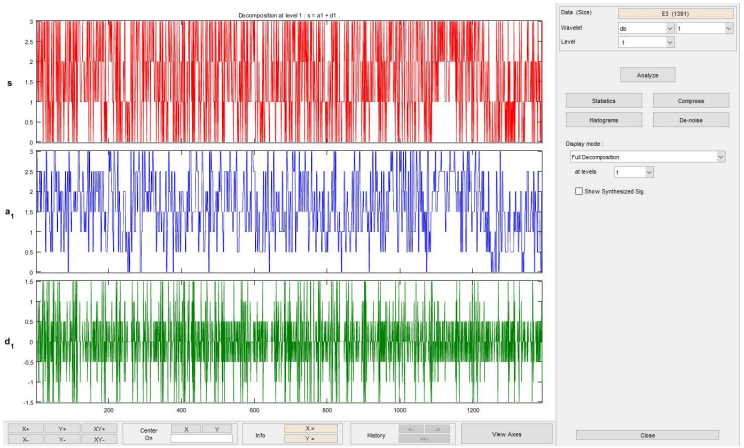

Figure 5: Approximation and Detail of AGER with db1.

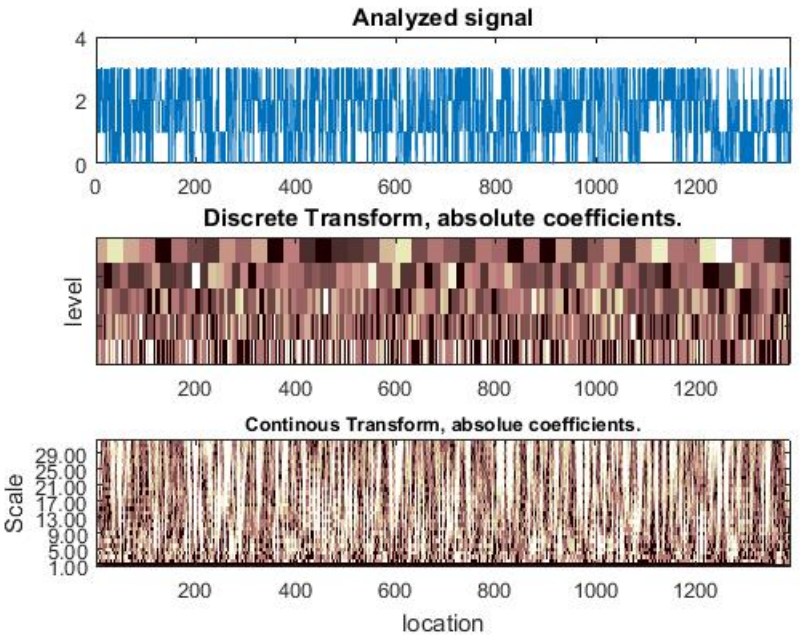

Figure 6: Discrete and Continous Transform of AGER with db1.

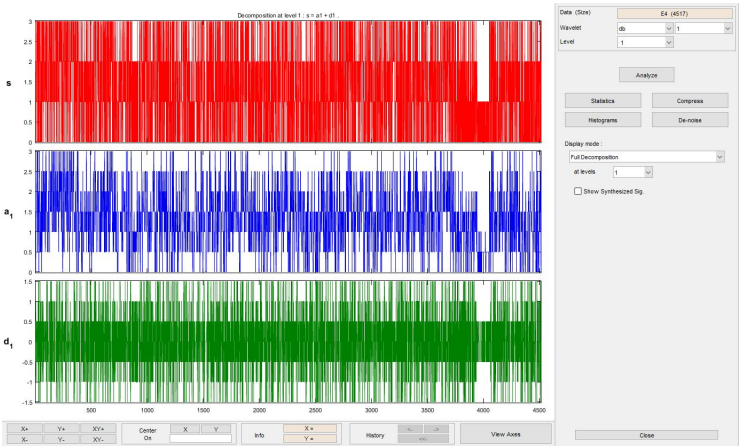

Figure 7: Approximation and Detail of ADIPOQ with db1.

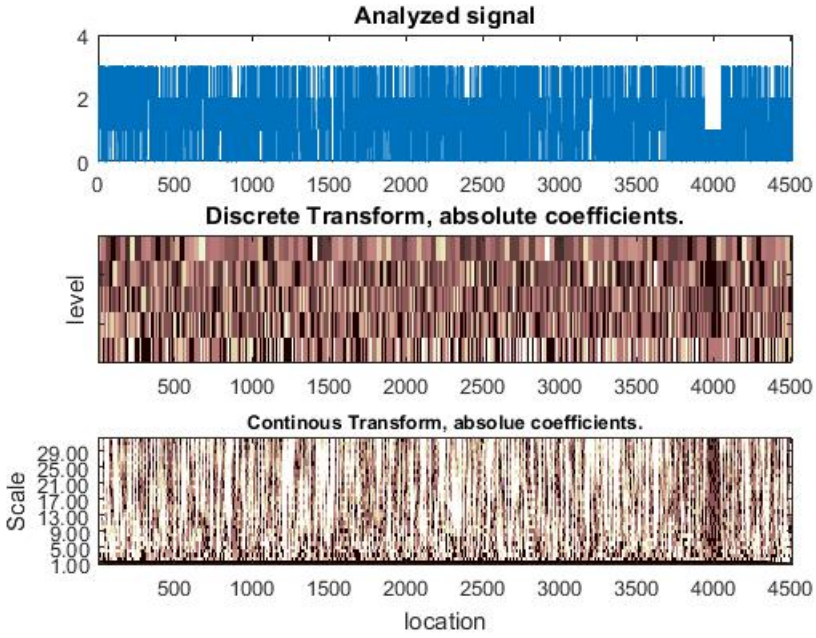

Figure 8: Discrete and Continous Transform of ADIPOQ with db1.

3.5. Angiotensinogen (serpin peptidase inhibitor, clade A, member 8) (AGT)

The angiotensin family of peptides is important in the regulation of blood volume, vascular resistance, and electrolyte balance. The protein encoded by this gene, pre-angiotensinogen or angiotensinogen precursor, is expressed in the liver and is cleaved by the enzyme renin in response to lowered blood pressure. It is produced in a cascade whereby the precursor peptide angiotensinogen is cleaved to produce renin and angiotensin I. Angiotensin-converting enzyme then acts on angiotensin I to yield the octapeptide angiotensin II, and further processing generates angiotensins III and IV. Wavelet analysis and wavelet transform of AGT are shown in Figures 9 and 10 respectively.

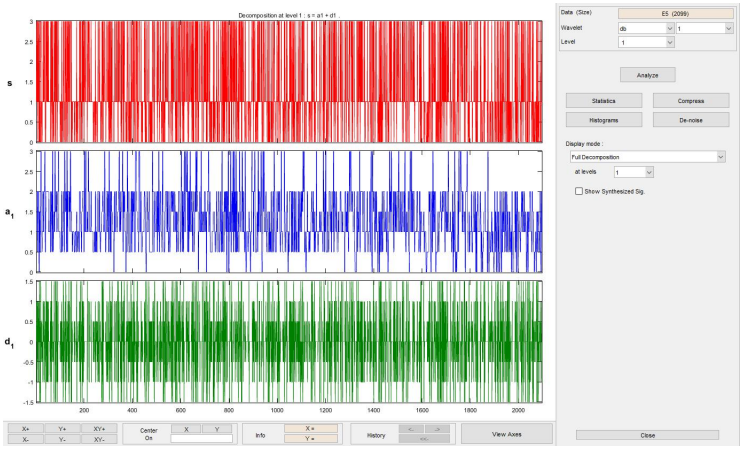

Figure 9: Approximation and Detail of AGER with db1.

3.6. Angiotensin II receptor, type 1 (AGTR1)

Angiotensin II is a potent vasopressor hormone and a primary regulator of aldosterone secretion. It is an important effector controlling blood pressure and volume in the cardiovascular system. It acts through at least two types of receptors. This gene encodes the type 1 receptor which is thought to mediate the major cardiovascular effects of angiotensin II. This gene may play a role in the generation of reperfusion arrhythmias following restoration of blood flow to ischemic or infarcted myocardium. It was previously thought that a related gene, denoted as AGTR1B, existed; however, it is now believed that

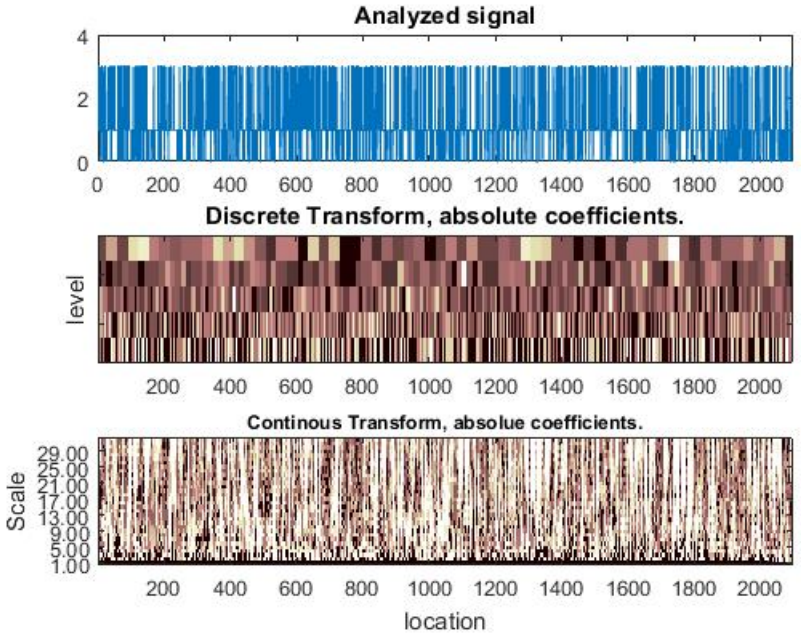

Figure 10: Discrete and Continous Transform of AGT with db1.

there is only one type 1 receptor gene in humans. At least five transcript variants have been described for this gene. Wavelet analysis and wavelet transform of AGTR1 are shown in Figures 11 and 12 respectively.

In this section, we have presented each gene's wavelet analysis, wavelet transforms and coefficient distributions. In what follows, we present some comparisons within each gene and compute some variance and entropy values for each gene. We will show some characteristics of candidate genes and how we express the differences among the genes.

4. Comparisons

To understand each gene's overall wavelet coefficients, we calculate its normalized values by using the following global comparison formula [28]. We plot their figures (Figure 13-18) as follows.

$$N(a) = \frac{w(a,b)}{max(abs(w(a,b)))}, \tag{14}$$

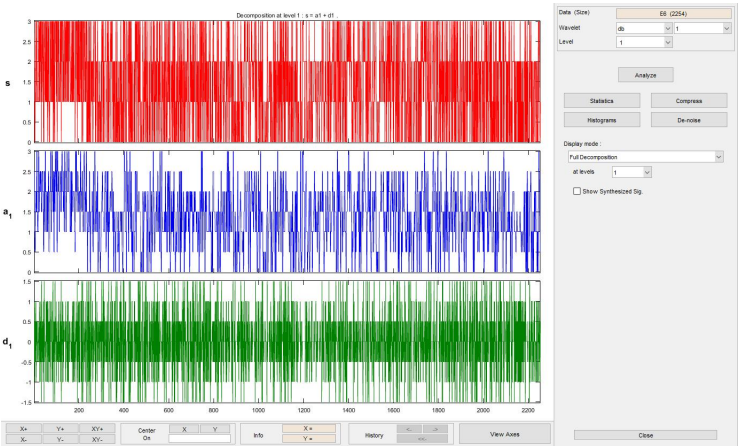

Figure 11: Approximation and Detail of AGTR1 with db1.

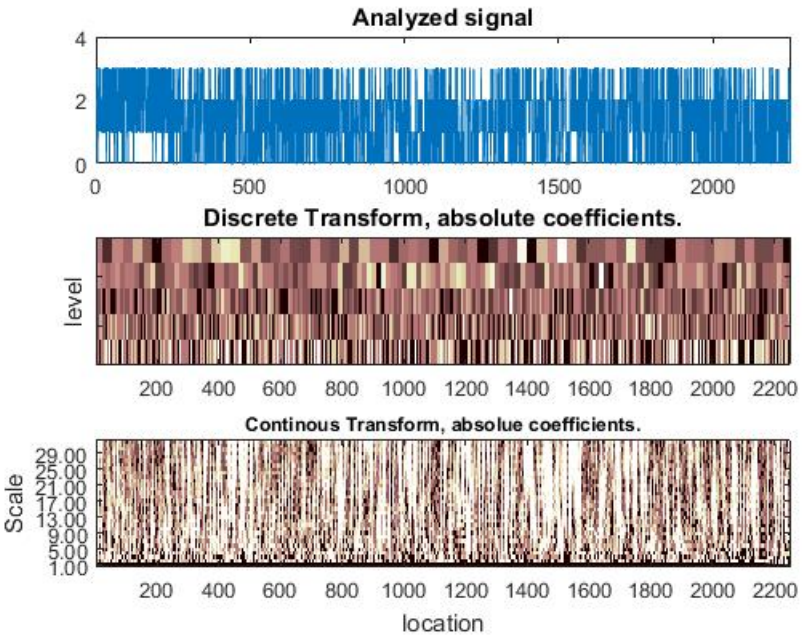

Figure 12: Discrete and Continous Transform of AGTR1 with db1.

where $w(a, b)$ is wavelet transform defined in equation (12).

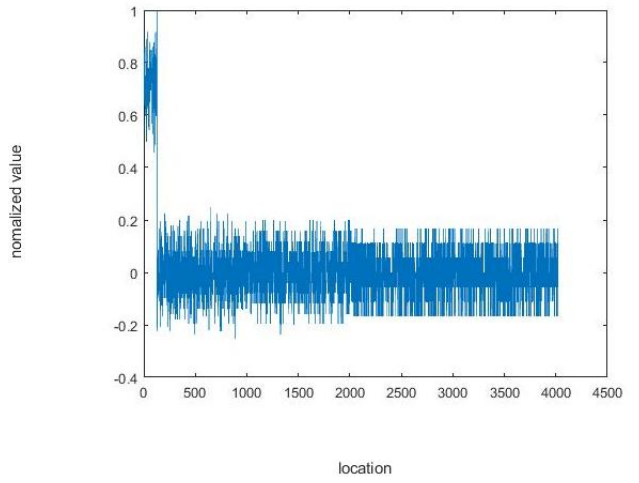

Figure 13: Global Comparisons of ACE.

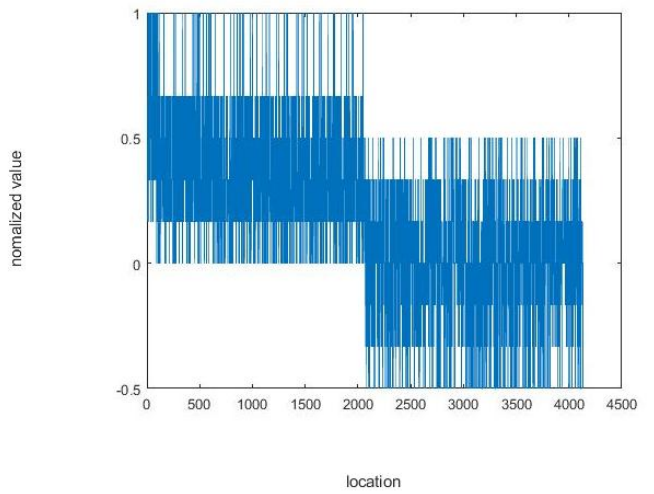

Figure 14: Global Comparisons of CCDC68.

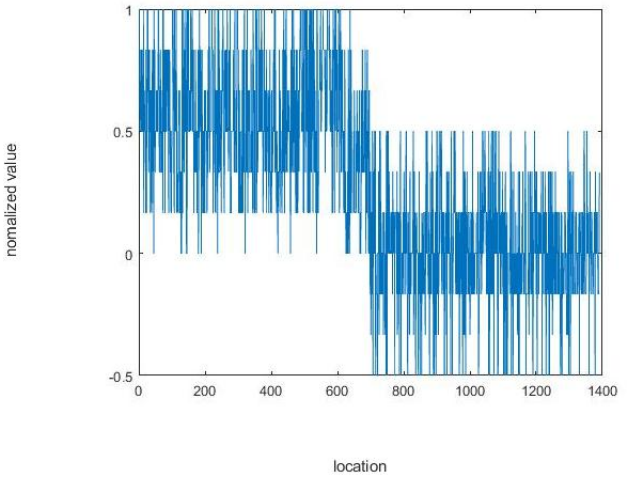

Figure 15: Global Comparisons of AGER.

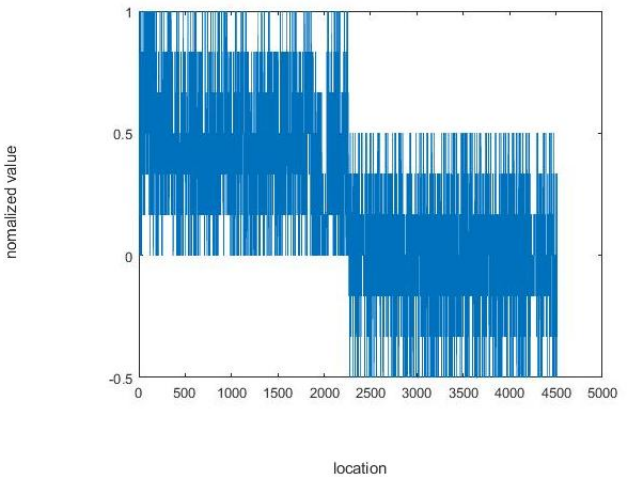

Figure 16: Global Comparisons of ADIPOQ.

Alternatively, we can calculate wavelet variance to understand the comparisons among different genes. The variance is defined as

$$V(a) = \frac{1}{n}\sum_{j=1}^{n} w^2(a, x_j),$$

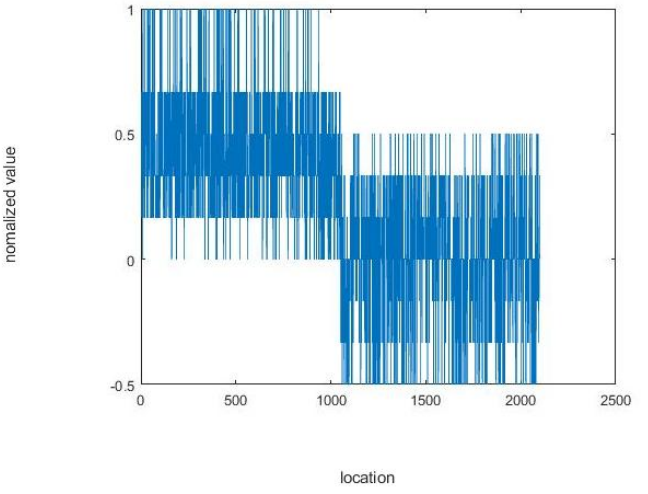

Figure 17: Global Comparisons of AGT.

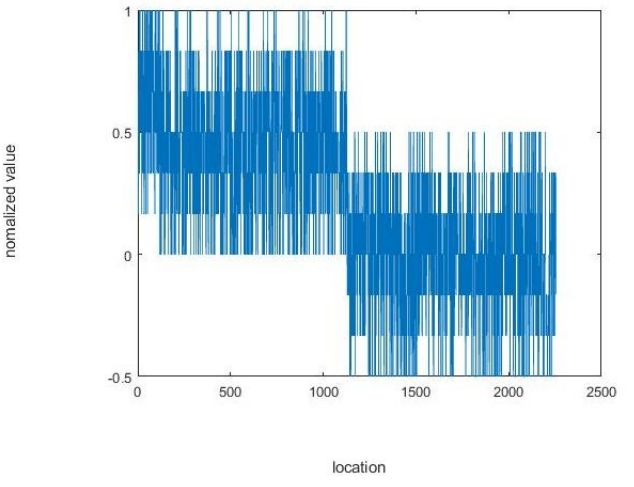

Figure 18: Global Comparisons of AGTR1.

Table 1. Wavelet Coefficients of Candidate Genes						
Genes	ACE	CCDC68	AGGER	ADIPOQ	AGT	AGTR1
Variance	3.8155	3.2372	3.8743	3.1102	2.9081	3.1788

Table 2. Wavelet Entropy of Candidate Genes						
Genes	ACE	CCDC68	AGGER	ADIPOQ	AGT	AGTR1
Entropy	-3536.4	-8695.4	-3390.9	2148.3	-991.3637	-3949.9

where $w(a, x_j)$ are wavelet coefficients [29]. Their values are listed in Table 1.

To measure the disorderliness or randomness in a closed system, we use entropy which is considered as a measure of uncertainty. It is defined as $-\sum w_j log_2 w_j$, where w_j are wavelet coefficients [29]. We present the corresponding entropy for each candidate gene in Table 2.

5. Discussions

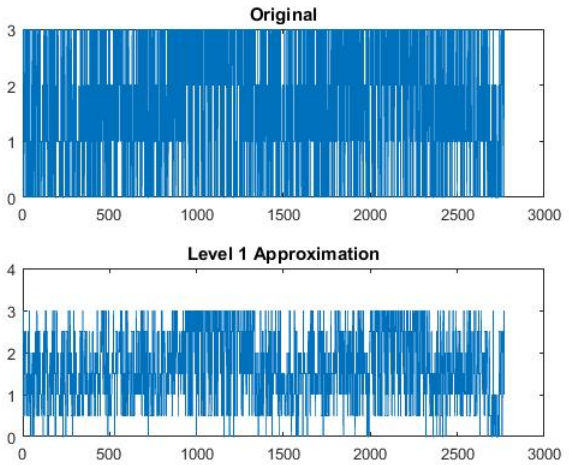

Figure 19: Wavelet Analysis on FGF21.

Wavelet analysis is a useful tool to analyze, decompose and characterize signals. We have presented computational, graphical and conceptional illustrations on several candidate genes. As we compare the results in Table 1 and Table 2 as well as Figures 1-12, every gene is basically unique. They are different from each other. For instance, the values of ACE and AGGER are

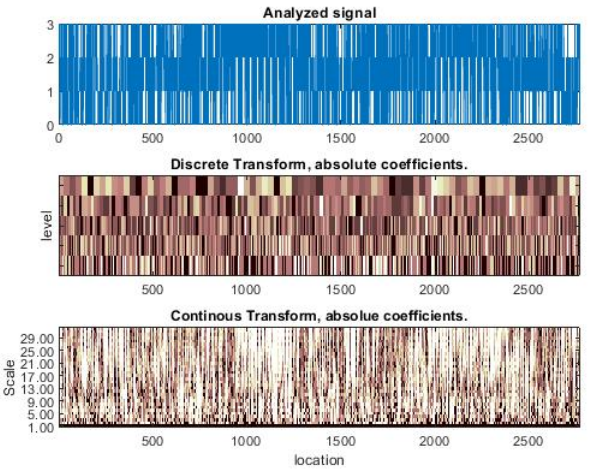

Figure 20: Wavelet Transform on FGF21.

close but their figures are quite different. Observing from the above figures, it is not easy to compare different genes to some extent. It would be more visible or precise if we divide each gene into several segments. Further questions such as prediction and detection problems [10, 46, 51] can be considered by using our methods to investigate new findings. A recent study shows that a single administration of an adeno-associated viral vector (AAV) carrying the FGF21 (Fibroblast Growth Factor 21) gene, resulted in genetic manipulation of the liver, adipose tissue or skeletal muscle to continuously produce the FGF21 protein. Here, we perform wavelet analysis and wavelet transform on FGF21 which are shown in Figure 19 and Figure 20, respectively. This protein is a hormone secreted naturally by several organs that acts on many tissues for the maintenance of correct energy metabolism. By inducing FGF21 production through gene therapy the animal lost weight and decreased insulin resistance, which causes the development of type 2 diabetes [23]. FGF21 is considered a promising therapeutic agent for type 2 diabetics and obesity. In fact, gene therapy includes the approach that involves the introduction of a foreign gene into any cell type in the body, allowing it to produce insulin. The gene introduced could be the insulin gene itself allowing for expression in a gene coding a factor such that in turn activates the insulin gene, thereby allowing for ectopic insulin production. Also, several aspects of developments can be further studied, namely, other methods of numerical representations, other wavelets, coding and non-coding regions of

different genes, sequence alignment, structure analysis, regulation analysis and improvements of diagnosis, treatment, and prevention of the disease. Ultimately, our study would help establish more reliable clinical care for patients.

Acknowledgment. The author would like to thank the anonymous referee's valuable comments.

References

[1] O. Abbasi, A. Rostami, and G. Karimian, *Identification of exonic regions in DNA sequenes using cross-correlation and noise suppression by discrete wavelet transform*, BMC Bioinform **12** (2011), 1–10.

[2] M. Akhtar, J. Epps, and E. Ambikairajah, *On DNA numerical representations for period-3 based exon prediction*, in: Proceedings of IEEE International Workshop on Genomic Signal Processing and Statistics, GENSIPS, Tuusula Finland 1–4, 2007.

[3] M. Akhtar, J. Epps, and E. Ambikairajah, *Signal processing in sequence analysis: advances in eukaryotic gene prediction*, IEEE J Sel. Top. Signal Process **2** (2008), no. 3, 310–321.

[4] D. Anastassiou, *Genomic signal processing*, IEEE Signal Process. Mag. **18** (2001), no. 4, 8–20.

[5] A. Arneodo, Y. D. Aubenton-Carafa, B. Audit, E. Bacry, F. F. Muy, and C. Thermes, *What can we learn with wavelets about DNA sequences*? Phys. A. **249** (1998), 439–448.

[6] J. P. Bao and R. Y. Yuan, *A wavelet-based feature vector model for DNA clustering*, Genetics and Molecular Research **14** (2015), no. 4, 19163–19172.

[7] J. A. Berger, S. K. Mitra, M. Carli, and A. Neri, *New approaches to genome sequence analysis based on digital signal processing*, in: Proceedings of IEEE Workshop on Genomic Signal Processing and Statistics (GENSIPS), Raleigh, NC, 1–4, 2002.

[8] J. A. Berger, S. K. Mitra, M. Carli, and A. Neri, *Visualization and analysis of DNA sequences using DNA walks*, J. Framkl. Inst. **341** (2004), 37–53.

[9] Bharat Bhosale, Bouthina S. Ahmed, and Anjan Biswas, *Wavelet Based Analysis in Bio-informatics*, Life Science Journal **10** (2013), no. 2, 853–859.

[10] C. Burge and S. Karlin, *Prediction of complete gene structures in human genomic DNA*, Journal of molecular biology **26** (1997), no. 1, 78–94.

[11] V. Costa and A. Ciccodicola, *Is PPARG the key gene in diabetic retinopathy*? Britich Journal of Pharmacology **165** (2011), 1–32.

[12] P. D. Cristea, *Genetic signals: an emerging concept*, in: Proceedings of 8th International Workshop on Systems, Signals and Image Processing, Bucharest, Romania, 17–22, 2001.

[13] P. D. Cristea, *Genetic signals representation and analysis*, in: Proceedings of SPIE 4623, Functional Monitoring and Drug-Tissue Interaction, **4623** (2002), 77–84. DOI:10.1117/12.491244.

[14] P. D. Cristea, Genomic Signal Processing and Statistics, Eurasip Book Series on Signal Processing and Communications, Hindawi Publishing Corporation15–65, 2005.

[15] P. D. Cristea, R. Tuduce, C. Banica, and K. Rodewald, *Genomic signals for the study of multiresistance mutations in M Tuberculosis*, in: Proceedings of International Symposium on Signals Circuits and Systems , Romania **1** (2007), 1–4. DOI:10.1109/ISSCS.2007.4292708.

[16] Abdesselem Dakhli Wajdi Belli, and Chokri Ben Amar, *DNA sequence classification using power spectruma and wavelet neural network*, Proceedings of the 16th International Conference on Hybrid Intelligent Systems, 391–402, HIS 2016.

[17] B. Demeler and G. W. Zhou, *Neural network optimiation for E. coli promoter prediction*, Nucl. Acids Res. **19** (1991), no. 7, 1539–1599.

[18] G. Dodin, P. Vandergheynst, P. Levoir C. Cordier, and L. Marcour, *Fourier and wavelet transform analysis a tool for visualising regular patterns in DNA sequences*, J. Theor Biol. **206** (2000), no. 7, 323–326.

[19] T. P.George and T. Thomas, *Discrete wavelet transform de-noising in eukaryotic gene splicing*, BMC Bioinform **11** (2010), Supp 1, S50. DOI:10.1186/1471-2105-11-S1-S50.

[20] A. D. Haimovich, B. Byrne, R. Ramaswamy, and W. J. Welsh, *Wavelet analysis of DNA walks*, J. Comput. Biol. **13** (2006), no. 7, 1289–1298.

[21] Y. Hur and H. Lee, *Wavelet-based identification of DNA focal genomic aberrations from single nucleotide polymorphism arrays*, BMC Bioinform (2011). DOI:10.1186/1471-2105-12-146.

[22] T. M. Inbamalar and R. Sivakumar, *Improved algorithm for analysis of DNA sequences using multiresolution transformation*, Sci. World J. **2015** (2015). Article ID 786497, DOI:10.1155/2015/786497.

[23] V. Jimenez, et al, *FGF21 gene therapy as treatment for obesity and insulin resistance*, Science Daily **9** (2018). DOI:10.15252/emmm.201708791.

[24] A. Krishnan, K. B. Li, and P. Isaac, *Rapid detection of conserved regions in protein sequences using wavelets*, Silico Biol. **4** (2004), no 2, 133–148.

[25] Hon Keung Kwan and S. B. Arniker, *Numerical representation of DNA sequences*, IEEE International Conderence (2009), 307–310.

[26] E. B. Lin and E. Linton, *Wavelet packet analysis of DNA sequences*, Proceedings of the 5th International Conference on Bioinformatics and Biomedical Engineering, 2011.

[27] E. B. Lin, M. Haske, M. Smith, and D. Sowards, *Wavelet analysis of ECG signals, multiscale signal analysis and modeling*, Lecture Notes in Electrical Engineering, Springer, New York, 233–256, 2013.

[28] E. B. Lin and M. Oyapero, *Wavelet analysis of coding and noncoding regions of DNA sequences using integer representation*, Journal of Computational Intelligence in Bioinformatics **5** (2012), no. 2, 149–159.

[29] E. B. Lin and M. Oyapero, *Wavelet variance and entropy of DNA sequences*, International Journal of Computer and Communication Engineering, IACSIT, Singapore, 70–72, 2012.

[30] E. Linton, P. Albee, P. Kinnicutt, and E. B. Lin, *Time frequency analysis of mitochondrial genome*, Proceedings of the 2010 International Conference on Bioinformatics and Computational Biology, 293–297, 2010.

[31] E. Linton, P. Albee, P. Kinnicutt, and E. B. Lin, *Multiresolution analysis of DNA sequences*, Proceedings of Second International Conference on Computer Research and Development, 218–221, 2010.

[32] P. Lio and M. Vannucci, *Finding pathogenicity islands and gene transfer events in genome data*, Bioinformatics **16** (2000), no. 10, 932–940.

[33] J. A. T. Machado, A. C. Costa, and M. D. Quelhas, *Wavelet analysis of human DNA*, Genomics **98** (2011), no. 3, 155–163.

[34] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, 1998.

[35] K. B. Murray, D. Gorse, and J. M. Thornton, *Wavelet transforms for the characterization and detection of repeating motifs*, J. Mol. Biol. **316** (2002), no. 2, 341–363.

[36] National Center for Biotechnology Information (NCBI), `http://www.ncbi.nlm.nih.gov`.

[37] Nha Nguyen, An Vo, Inchan Choi, and Kyoung-Jae Won, *A stationary wavelet entropy-based clustering approach accurately predicts gene expression*, Journal of Computational Biology **22** (2015), no. 3, 236–249. `DOI:10.1089/cmb.2014.0221`.

[38] J. Ning C. N. Moore, and J. C. Nelson, *Preliminary wavelet analysis of genomic sequences*, in: Proceedings of the IEEE Computer Society Conference on Bioinformatics, 509–510, Stanford, CA 2003.

[39] C. K. Peng, S. V. Buldyrev A. V. Goldberger S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, *Long-range correlations in nucleotide sequences*, Nature **356** (1992), 168–170.

[40] R. Ranawana and V. Palade, *A neural network based multi-classifier system for gene identification in DNA sequence*, Neural Comput. Appl. **14** (2005), no. 2, 122–131.

[41] K. D. Rao, and M. n. S. Swamy, *Analysis of genomics and proteomics using DSP techniques*, IEEE Trans. Circuits Syst.-I **55** (2008), no. 1, 370–378.

[42] Shiwani Saini and Lillie Dewan, *Comparison of numerical representations of genomic sequences: choosing the best mapping for wavelet analysis*, International Journal of Applied and Computational Mathematics **3** (2017), no. 4, 2943–2958.

[43] J. Song, A. Ware, and S. Liu, *Wavelet to predict bacterial ori and ter: a tendency towards a physical balance*, BMC Genomics (2003). `DOI:10.1186/1471-2164-4-17`.

[44] Levente Szendrei Tibor Turoczi, Peter Kovacs, Miklos Vecsernyes, Dipak K. Das, and Arpad Tosaki, *Mitochondrial gene expression and ventricular fibrillation in ischemic/reperfused nondiabetic and diabetic myocardium*, Biochemical Pharmacology **63** (2002), 543–552.

[45] A. A. Tsonis P. Kumar, J. B Elsner, and P. A. Tsonis, *Wavelet analysis of DNA sequences*, Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. **53** (1996), no. 2, 1828–1834.

[46] P. P. Vaidyanathan and B. J. Yoon, *Digital filters for gene prediction applications*, in: Proceedings of IEEE asilomar conference on signals systems and computers, Monterey, CA, 2002.

[47] P. P. Vaidyanathan and B. J. Yoon, *The role of signal-processing concepts in genomics and proteomics*, J. Framkl. Inst. **341** (2004), 111–135.

[48] G. Vidhya and B. Anusha, *Diaretinopathy, Database - A Gene database for diabetic retinopathy*, Bioinformation **10** (2014), no. 4, 235–240.

[49] K. Vinita, S. Aripriya, K. Prathiba, K. Vaitheeswaran, R. Sathyabaarathi, M. Rajesh, J. Amali, V. Umashankar, and G. Kumaramanickavel, *ICAM-1 K469E polymorphism is a genetic determinant for the clinical risk factors of T2D subjects with retinopathy in Indians: a population-based case-control study*, BMJ Open **2** (2012), e001036.

[50] R. F. Voss, *Evolution of long-range fractal correlations and l/f noise in DNA base sequence*, Phys. Rev. Lett. **68** (1992), 3805–3808.

[51] X. Yu. T. W. Randolph, H. Tank, and L. Hsu, *Detecting genomic aberrations using products in a multiscale analysis*, Biometrics **66** (2010), 684–693.

[52] M. Q. Zhang, *Identification of protein coding regions in the human genome by quadratic discriminant analysis*, Proc. Natl Acad. Sci. USA, 565–568.

[53] R. Zhang and C. T. Zhang, *Identification of replication origins in archaeal genomes based on the curve method*, Archaea **1** (2005), 335–346.

Department of Mathematics, Central Michigan University
Mt. Pleasant, MI 48859, USA
E-mail address: lin1e@cmich.edu