International Journal on
Advanced Science
Engineering
Information Technology

Vol. 2 (2012) No. 5
ISSN: 2088-5334

In Silico Analysis of Oncogenes for Renal Cancer

Sim-Hui Tee[#], Siew-Kien Mah[*]

[#] *Faculty of Creative Multimedia, Multimedia University, 63100 Cyberjaya, Malaysia*
E-mail: shtee@mmu.edu.my

**Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Malaysia*
E-mail: skmah@mmu.edu.my

Abstract— **Computational tools and methods play a vital role in handling and analyzing a large volume of genomic data. In cancer research, *in silico* methods such as computational algorithm and protein databases are indispensable. In this paper, we adopted an *in silico* approach to analyze oncogenes that cause renal cancer. Our objective is to identify and analyze the genes which are over expressed in the renal cancer tissues. The identification of oncogenes for renal cancer could provide directions and insights for molecular cancer treatment.**

Keywords— ***In Silico*, oncogenes, renal cancer**

I. Introduction

Computational tools and methods have assumed greater role in analysing biological and medical data. These data which are generated in the laboratory are enormously large in volume. Genetic sequences, for example, could possess a length of millions of neucleotides. To make genetic data useful for biological research and medical therapy, it is always necessary to use computational tools to analyse and organize them.

In this paper, we took an *in silico* approach to analyze oncogenes that cause renal cancer. It was reported that nearly 50% of renal cancer patients have eventually suffered from degenerated conditions in which metastases have developed [1]. This malignant condition has contributed to high mortal rate in the patients. It was estimated that approximately 78,000 deaths among renal cancer patients have been recorded [2]. The identification of oncogenes could provide directions and insights for molecular cancer therapy.

We made use of Expressed Sequenced Tags (ESTs) to generate the data for both normal and cancerous renal cell line. ESTs were fragments of cDNA that were widely used by bioinformaticians and computer scientists in gene sequencing and analysis. Our objective is to identify and analyze the genes which are over expressed in the renal cancer tissues. Our computational analysis may improve the understanding of the genes which play a regulatory role in renal carcinogenesis. Furthermore, computational analysis of protein expression of the oncogenes will reveal the relation between the renal cancer and genetic content at molecular level.

II. Methods

We used Expressed Sequence Tags (EST) to analyze both normal and cancerous renal tissue. All EST libraries were selected with a minimum of 2000 sequences per library. We used cell line as the tissue preparation type for both normal and cancerous tissue histology. We set the library protocol to non-normalized in order to maximize the number of candidate genes.

We specified the minimum gene expression factor as 20. It implies that the resultant genes are expected to have over-expressed for at least 20 folds. This high threshold enables us to exclude the irrelevant genes from our analysis.

The false discovery rate was set to 0.001, which is a value considerably sensitive to the significant expression of genes. With all 46 chromosomes being included, we obtained 8 EST libraries, as shown in Table 1.

TABLE I
EST Libraries Implicated in Renal Cancer Cell Line

Cancer cell	Normal cell
NIH_MGC_58	NIH_MGC_141
NIH_MGC_89	NIH_MGC_126
NIH_MGC_14	NIH_MGC_142
	NIH_MGC_127
	NIH_MGC_128

Based on the EST libraries as shown in Table 1, we generated the over-expressed genes which are implicated in renal cancer. The gene ontology was identified. From the obtained mRNA sequences, we identified the protein sequences for each oncogene which is correlated to renal carcinogenesis.

III. Results and discussions

There were 3 EST libraries with a total of 28564 sequences implicated as cancer cells; whereas we obtained 5 EST libraries with a total of 37421 sequences as normal cells. After excluding the genes which have invalid sequence odds ratio (OR), we obtained 17 candidate genes for renal cancer, as shown in Table 2.

TABLE 2
Candidate Genes for Renal Cancer

Genes	Sequences		OR	Q
	Cancer	Normal		
IGFBP3	121	2	79.59	2.79e-38
LMNA	43	1	56.42	2.91e-12
TGFBI	293	7	55.39	6.42e-92
MGAT4B	34	1	44.59	3.35e-9
ITGA3	33	1	43.28	7.12e-9
FSCN1	28	1	36.72	3.02e-7
SLC16A3	46	2	30.18	3.23e-12
SEMA4B	20	1	26.22	1.31e-4
AKT1	20	1	26.22	1.31e-4
GAPDH	215	11	25.79	4.43e-60
DDIT4	38	2	24.92	1.61e-9
SLC25A39	19	1	24.91	2.62e-4
ESYT1	18	1	23.60	5.36e-4
MYO1C	17	1	22.28	9.16e-4
FXYD2	17	1	22.28	9.16e-4
TNIP1	17	1	22.28	9.16e-4
HM13	17	1	22.28	9.16e-4

Table 2 demonstrates a list of 17 candidate genes which are responsible for renal carcinogenesis, with significant low false discovery rate (Q). With such a low false discovery rate, the reliability of the genes is warranted. The chromosomal positions of these genes are given in Table 3.

TABLE 3
The Chromosomal Positions of the Candidate Genes

Genes	Chromosomal positions
IGFBP3	7: 45951955 - 45952627
LMNA	1: 156105877 - 156108962
TGFBI	5: 135364679 - 135399506
MGAT4B	5: 179225009 - 179233625
ITGA3	17: 48140901 - 48143396
FSCN1	7: 5632438 - 5646279
SLC16A3	17: 80186928 - 80197355
SEMA4B	15: 90744573 - 90772890
AKT1	14: 105236679 - 105258980
GAPDH	12: 6643706 - 6647537
DDIT4	10: 74033693 - 74035794
SLC25A39	17: 42396999 - 42400891
ESYT1	12: 56522228 - 56538452
MYO1C	17: 1368124 - 1395951
FXYD2	11: 117690789 - 117695434
TNIP1	5: 150410249 – 150444656
HM13	20: 30154005 - 30154652

Table 3 shows that most of the candidate genes are several thousands bases in length. However, some genes are rather short. IGFBP3 which is located at chromosome 7, for example, has only 672 bases. However, the length of genes does not determine the factor for renal carcinogenesis. We identified the gene ontology for these 17 genes in order to analyze their molecular functions and physiological processes. Gene ontology represents the genome using an agreed-upon terms to describe the gene products [3]. Each gene could have more than one gene ontology annotations. We used a standardized shorthand to represent the type of ontology for the 17 candidate genes. Biological process is denoted as [BP], cellular component as [CC], and molecular function as [MF]. We demonstrate the gene ontology of the top 3 over-expressed genes (IGFBP3, LMNA, and TGFBI) in Table 4.

TABLE 4
Gene Ontology of the Top 3 Over-Expressed Genes

Gene ontology	Genes
[BP]negative regulation of cell proliferation	IGFBP3
[BP]negative regulation of protein phosphorylation	IGFBP3
[BP] negative regulation of signal transduction	IGFBP3
[BP] negative regulation of smooth muscle cell migration	IGFBP3
[BP] negative regulation of smooth muscle cell proliferation	IGFBP3
[BP] positive regulation of apoptosis	IGFBP3
[BP] positive regulation of myoblast differentiation	IGFBP3
[BP] protein phosphorylation	IGFBP3
[BP] regulation of cell growth	IGFBP3
[CC] extracellular region	IGFBP3, TGFBI
[CC] extracellular space	IGFBP3, TGFBI
[CC] insulin-like growth factor binding protein complex	IGFBP3
[CC] nucleus	IGFBP3
[MF] insulin-like growth factor I binding	IGFBP3
[MF] insulin-like growth factor binding	IGFBP3
[MF] metal ion binding	IGFBP3
[MF] protein binding	IGFBP3, LMNA
[MF]protein tyrosine phosphatase activator activity	IGFBP3
[BP] apoptosis	LMNA
[BP] cellular component disassembly involved in apoptosis	LMNA
[BP] establishment or maintenance of microtubule cytoskeleton polarity	LMNA
[BP] muscle organ development	LMNA
[BP] nuclear envelope organization	LMNA
[BP] positive regulation of cell aging	LMNA
[BP] protein localization to nucleus	LMNA
[BP] regulation of apoptosis	LMNA
[BP] regulation of cell migration	LMNA
[BP] spermatogenesis	LMNA
[BP] sterol regulatory element binding protein import into nucleus	LMNA
[BP] ventricular cardiac muscle cell development	LMNA
[CC] cytoplasm	LMNA
[CC] insoluble fraction	LMNA

[CC] intermediate filament	LMNA
[CC] lamin filament	LMNA
[CC] nuclear envelope	LMNA
[CC] nuclear lamina	LMNA
[CC] nuclear matrix	LMNA
[CC] nucleus	LMNA
[CC] perinuclear region of cytoplasm	LMNA
[MF] structural molecule activity	LMNA
[BP] angiogenesis	TGFBI
[BP] cell proliferation	TGFBI
[BP] extracellular matrix organization	TGFBI
[BP] negative regulation of cell adhesion	TGFBI
[BP] response to stimulus	TGFBI
[BP] visual perception	TGFBI
[CC] extracellular matrix	TGFBI
[CC] proteinaceous extracellular matrix	TGFBI
[MF] extracellular matrix binding	TGFBI
[MF] integrin binding	TGFBI

Of these 50 molecular properties (as shown in Table 4) which are represented by three main categories of gene ontology (BP, CC, MF), we noticed that the top 3 over-expressed genes (IGFBP3, LMNA, and TGFBI) exemplify a wide range of physiological characteristics. Of 50 molecular properties, only three properties are shared by two genes, which are extracellular region, extracellular space, and protein binding. However, the phenomenon of unique gene ontology does not imply the connectionless between the genes in renal carcinogenesis. The more gene ontology annotation is attributed to a gene, the more physiological role is assumed. TGBFI has been observed to mutate in the renal cells, which contributes to the renal carcinogenesis [4]. Experiment has also shown that the extracellular matrix protein of TGFBI has a wide implication in cancer therapy by inducing chemotherapy resistance via apoptosis [5]. It has been an agreement that TGFβ, a superfamily of TGFBI, plays a pivotal role in cancer progression [6-7]. It was reported that the phenotypic spectrum of diseases caused by LMNA mutations is diverse [8]. The results of real-time qPCR assay showed that the expression level of LMNA was significantly increased in a time-dependent manner in response to low-intensity pulsed ultrasound in human lymphoma U937 cells [9]. IGFBP3 was reported to have implicated in the renal cell carcinoma [10] and drug tolerance [11]. The results of cDNA microarray demonstrated that IGFBP3 assumes a role in cell proliferation [12].

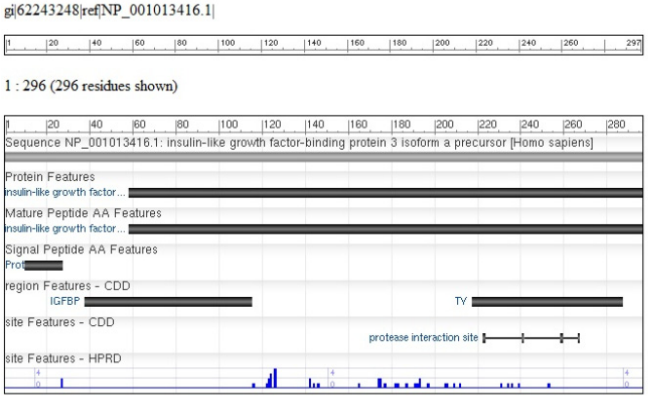

Fig. 1. Amino acid sequences of the IGFBP3 gene product

Fig. 1 provides the amino acid sequences of the IGFBP3 gene product, which shows the characteristic of insulin-like growth factor.

Fig. 2 shows the amino acid sequences of the LMNA gene product. The protein product is prelamin A/C, which catalyzes cascades of phosphorylation in cells.

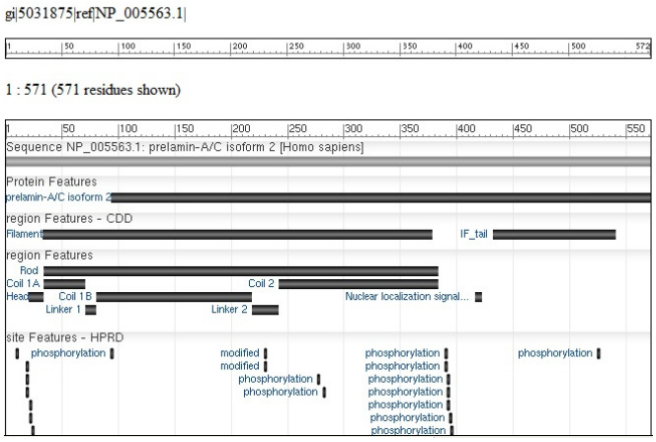

Fig. 2. Amino acid sequences of the LMNA gene product

Fig. 3 depicts the amino acid sequences of the TGFBI gene product, which is a transforming growth factor beta-induced protein.

Fig. 3. Amino acid sequences of the TGFBI gene product

The over-expression of IGFBP3, LMNA, and TGFBI explains the respective high sequence odds ratio in renal cancer cells. Gene ontology also supports the observed expression of these three genes. It further validates that IGFBP3, LMNA, and TGFBI are important oncogenes that account for the renal carcinogenesis.

IV. Conclusions

Practically all genomic data are analysed using computational tools and methods. The gene ontology of the top 3 over-expressed genes that is IGFBP3, LMNA and TGFBI that caused renal cancer are analysed using *in silico* approach in this paper. Factors inducing renal carcinogenesis do not depend on the length of genes. The gene ontology annotation reflects the physiological roles in which mutation in the renal cells contributes to renal carcinogenesis. Examination of amino acid sequences indicated that IGFBP3, LMNA and TGFBI all show the

characteristic of growth factor. We concluded that IGFBP3, LMNA and TGFBI play an important role in renal carcinogenesis

REFERENCES

[1] B. R. Pflug., H. Zheng., M. S. Udan., J. M. D'Antonio., F. F. Marshall., J. D. Brooks., and J. B. Nelson, "Endothelin-1 promotes cell survival in renal cell carcinoma through the ET_A receptor", *Cancer Letters*, vol 246, pp. 139-148, 2007.

[2] L. He., H. Wang., H. Jin., C. Guo., H. Xie., K. Yan., X. Li, Q. Shen., T. Qiao., G. Chen., N. Chai., L. Zhao., Q. Dong., Y. Zheng., J. Liu., and D. Fan, "CIAPIN1 inhibits the growth and proliferation of clear cell renal cell carcinoma", *Cancer Letters*, vol 276, pp. 88-94, 2009.

[3] R. P. Huntley., D. Binns., E. Dimmer., D. Barrell., C. O'Donovan., and R. Apweiler, "QuickGO: a user tutorial for the web-based gene ontology browser", *Database*, 2009.

[4] Y. Zhang., G. Wen., G. Shao., C. Wang., C. Lin., H. Fang., A. S. Balajee., G. Bhagat., T. K. Hei., and Y. Zhao, "TGFBI deficiency predisposes mice to spontaneous tumor development", *Cancer Research*, vol 69 (1), pp 37-44, 2009.

[5] A. A. Ahmed., A. D. Mills., A. E. Ibrahim., J. Temple., C. Blenkiron., M. Vias., C. E. Massie., N. G. Lyer., A. McGeoch., R. Crawford., B. Nicke., J. Downward., C. Swanton., S. D. Bell., H. M. Earl., R. A. Laskey., C. Caldas, and J. D. Brenton, "The extracellular matrix protein TGFBI induces microtubule stabilization and sensitizes ovarian cancers to paclitaxel", *Cancer Cell*, vol 12, pp 514-527, 2007.

[6] C. L. Arteaga, "Inhibition of TGFβ signaling in cancer therapy", *Current Opinion in Genetics & Development*, vol 16, pp 30-37, 2006.

[7] J. J. Worthington., J. E. Klementowicz., and M. A. Travis, "TGFβ: a sleeping giant awoken by integrins", *Trends in Biochemical Sciences,* vol 36, pp 47-54, 2011

[8] S. Rudnik-Schöneborn., E. Botzenhart., T. Eggermann., J. Senderek., B. G. H. Schoser., R. Schröder., M. Wehnert., B. Wirth., K. Zerres., "Mutations of the LMNA gene can mimic autosomal dominant proximal spinal muscular atrophy", *Neurogenetics*, vol 8, pp137-142, 2007.

[9] Y. Tabuchi., I. Takasaki., Q-L. Zhao., S. Wada., T. Hori., L. B. Feril., K. Tachibana., T. Nomura., T. Kondo, "Genetic networks responsive to low-intensity pulsed ultrasound in human lumphoma U937 cells", *Cancer Letters*, vol 270, pp 286-294, 2008.

[10] S-W. Tang., W-H. Chang., Y-C. Su., Y-C. Chen., Y-H. Lai., P-T. Wu., C-I. Hsu., W-C. Lin., M-K. Lai., J-Y. Lin, "MYC pathway is activated in clear cell renal cell carcinoma and essential for proliferation of clear cell renal cell carcinoma cells", *Cancer Letters*, vol 273, pp35-43, 2009.

[11] S.V. Sharma., D. Y. Lee., B. Li., M. P. Quinlan., F. Takahashi., S. Maheswaran., U. McDermott., N. Azizian., L. Zou., M. A. Fischbach., K-K. Wong., K. Brandstetter., B. Wittner., S. Ramaswamy., M. Classon., and J. Settleman, "A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations", *Cell*, vol 141, pp69-80, 2010.

[12] Z. Zhang., L. Cao., J. Li., X. Liang., Y. Liu., H. Liu., J. Du., Z. Qu., M. Cui., S. Liu., L. Gao., C. Ma., L. Zhang., L. Han., W. Sun, "Acquisition of anoikis resistance reveals a synoikis-like survival style in BEL7402 hepatoma cells", *Cancer Letters*, vol 267, pp106-115, 2008.