European Journal of Medical and Health Sciences
www.ejmed.org

Chromosome HeatMap in CDK Patients as Defined by Multiregional Sequencing on Illumina MiSeq Platform

Mohammad F. Fazaludeen, Aymen A. Warille, Mohd Ibrahim Alaraj, Edem Nuglozeh

ABSTRACT

Renal failure and kidney disease are major concerns worldwide and are commonly coupled to diseases like hypertension, diabetes, obesity, and hypercholesterolemia. We undertook this study to explore the scope of genetic spectrum underlying the physiopathology of end-stage renal disease (ESRD) using whole exome sequencing (WES) on genomic DNA (gDNA) from 12 unrelated patients in younger ages. We have performed WES on 12 patients in stage of ESRD and analyze the FASTQ data through GATK pipeline. Here, we report for the first time a novel approach of establishing the severity and the magnitude of a disease on different chromosomes and associated karyotypes using chromosome Heatmap. The chromosome Heat will provide us with a road map to narrow down mutations selection leading us to SNPs characterization. Our preliminary results presented in the form of chromosomes HeatMap prelude our ongoing works which consist in identifying and characterizing new genes involved in the problem of renal diseases, results that depict the magnitude of the uncovered genes mutations and their biological implications related to the genome of these patients.

Keywords: Renal failure, Exome Sequencing, Chromosomes Heatmap.

Published Online: November 16, 2020

ISSN: 2593-8339

DOI: 10.24018/ejmed.2020.2.6.525

Mohammad F. Fazaludeen
A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Finland.
Aymen A. Warille
Department of Histology and Embryology, Faculty of Medicine, Ondokuz Mayis University, Turkey.
Mohd Ibrahim Alaraj
Department of Pharmacology, Faculty of Medicine, University of Hail, Saudi Arabia.
Middle East University, School of Pharmacy, Jordan.
Edem Nuglozeh*
Department of Biochemistry, Faculty of Medicine, University of Hail, Saudi Arabia.
Department of Biochemistry, Faculty of Medicine, Ondokuz Mayis University, Turkey.
(e-mail: edem.nuglozeh@OMU.edu.tr)

****Corresponding Author***

I. Introduction

Renal diseases are a cocktail of vascular diseases and metabolic dysregulations ranging from diabetes, hypertension and obesity. They cover a large spectrum of diseases spanning from relatively common to rare, and from benign disorders to high-morbidity/high-mortality disorders [1]. Chronic kidney disease (CKD) and renal failure are major health concerns worldwide and genetic factors are strongly one of the causes leading to establishment and development of the diseases. CKD is a multifactorial disorder with an important genetic component [2] and a number of monogenic disorders underlying CKD have been identified, but these account only for a small proportion of the total burden of kidney disease. Recent studies have identified common genetic variants at the UMOD, SHROOM3, GATM and MYH9 loci associated with kidney function in European and African American populations. [3], [4].

End-stage renal disease (ESRD) is the ultimate stage complication in CKD culminating with a necessity to undergo dialysis or renal transplantation. The incidence of new cases in the United States is 339 per million population (PMP) [5], or Europe (585 PMP) [6] and in India, it is (163 PMP) [7].

Only 10% of CDK cases are known to be due to Genetic related Renal Disease (GRD) and a big large proportion is yet to be identified as such. For nearly half of those identified as GRD, the gene implicated remains unknown [8, 9]. A typical example of GRD is Autosomal Dominant Polycystic Kidney Disease (ADPKD), a common potentially lethal disease [10].

View the dangerous and endemic slope the proportion of the CDK is undertaking, the need of establishing a genes data base or genes repertoire for CDK is becoming urgent and necessary. Such a gigantesque undertakes aimed at inventory and curate the genes population underlying the pathophysiology of CDK would only be accomplished via powerful and sophisticated methodologies to scan the whole human genome. To this end, we used Exome sequencing on Illumina platform to sequence and to decipher the faulty genes on Illumina MiSeq sequencer.

In pediatric populations' panel, Exome sequencing has been useful to diagnose genetic forms of nephrotic syndrome or congenital kidney defects [11]-[13]. Exome sequencing is a target capture and sequencing of the protein-

coding regions in the genome and the method is increasingly used as the first choice for the diagnostic purpose if it is available.

In the past, Zhang et al. [14] have used gap filling to annotate 31 genes in human MHC haplotype sequence. Subsequently, Sarnataro et al. [15] presented a HeatMap where they show contact domains between chromosomes 19 and 22. To our knowledge, there were no such studies reported yet using chromosomes HeatMap deriving genomic regions and mutations from whole exome sequencing from DNA of patients suffering from renal failure. Here we presented a novel approach to establish the severity or the magnitude of a disease on different chromosomes and associated haplotypes using chromosomes Heatmap.

II. Materials and Methods

A. *Samples Population*

The patients' genomic DNA (gDNA) investigated in this study were subgroup of 135 hemodialysis patients from the outpatient's clinics from King Khaled Hospital. Hemodynamic renal parameters were initially performed and published from these patients [16]. 12 blood samples were offered to our group by Dr. Moh'd Alaraj consent. We limited the number of patients to 12 in part due to the cost of Exome sequencing kits and the flurry of data generated after the sequencing at FASTQ level.

B. *Libraries Preparation*

We have generated human genomic library as reported elsewhere [17]. Brief, high quality gDNA was purified from whole blood using Qiagen kits (QIAamp DNA Blood Mini Kit from QIAGEN, Hilden Germany). The construction of the library was done using the Nextera Transposase Kit and sequenced on MiSeq Illumina sequencing platform.

C. *Bioinformatics Methodology*

Bioinformatics tools and pipeline that we used to analyze the sequence for variants of interest and fishing for the chromosome coordinates are summarized in Fig. 1.

III. Results

The 12 ESRD patients were identified by numbers K2, K3, K11, through numbers K76. Their anthropomorphic, clinical, and biochemical parameters are described in Table 1.

The normal glomerular filtration rate (GFR) as assessed via cystatin (GRFcy) and creatinine clearance (GFRcr) is above 90 mL/min/1.73 m2; in the 12 patients, it ranged from 11 to 30 mL/min/1.73 m2, representing a 60% to 79% decrease in renal function. The patients show an important accumulation of creatinine in the bloodstream (447-1196 µmol/L compared to normal value of 60-110 µmol/L) and most importantly are the case of uric acid which ranges from 154.8 to 1085 mmol/L as compared to the normal range of 2-8 µmol/L). The alteration of the above parameters represents a big deterioration in the renal function characteristic of ESRD. WES was applied on the libraries prepared from genomic DNA of the 12 patients. The number of reads submitted to hg19 allows us to extract chromosome coordinates that were we used subsequently to plot the chromosome HeatMap using Heatmap freely available at [18].

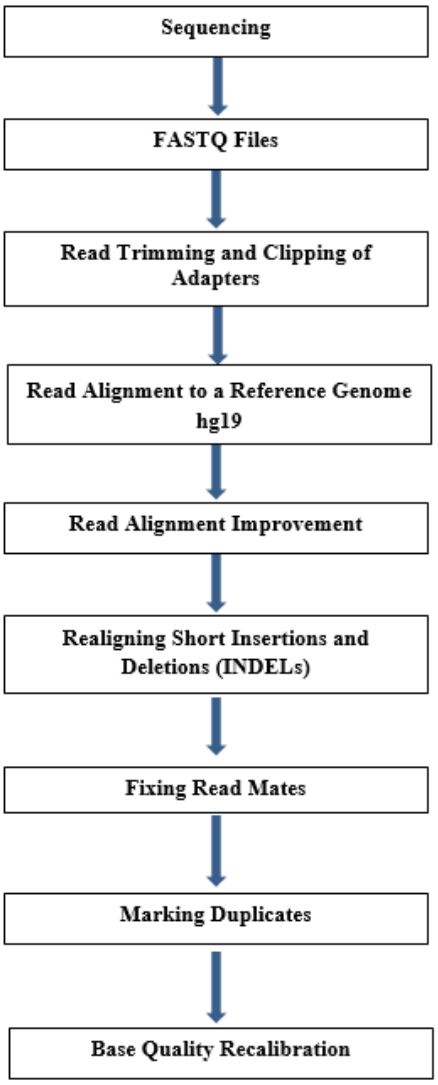

Fig. 1. GDNA library sequencing data flow analysis.

The data on this chart are organized hierarchically by chromosome and haplotypes, by sequenced-clone contigs within each chromosome and by severity of mutations. On the right of the Heat Map, there are 77 elements organized in folders. These include 22 autosomes; X and Y are for the sex chromosomes, 34 unknown chromosomes (ChrUn), sex chromosomes (X and Y), six haplotypes of interest harboring diagnostic variants, 12 haplotypes of minor interest and 22 autosomes.

Figure 2: Heatmap representing the Change rate by sample and Chromosome. HeatMap analysis was performed for the change rate by sample and chromosome, utilizing software Heatmapper (http://www.heatmapper.ca [18]. Rows represent the samples and columns represent the chromosome identified in the heatmap analysis. Color intensity signifies the significance severity of the mutations in the corresponding chromosome, determined by Fisher's Exact Test, with FDR correction for multiple comparisons by the Benjamin-Hochberg method.

TABLE 1: ANTHROPOMORPHIC DATA DEPICTING BIOCHEMICAL PARAMETERS DEDUCED FROM CDK PATIENTS

	Units	*Normal Range*	*Patients*											
			K2	K3	K11	K31	K32	K43	K45	K49	K56	K68	K69	K76
GFRcy	**mL/ min/1.73m²**	**≥ 90**	**11**	**9**	**16**	**30**	**5**	**4**	**12**	**16**	**8**	**7**	**13**	**9**
GFRcr	**mL/ min/1.73m²**	**≥ 90**	**11**	**9**	**15**	**28**	**5**	**4**	**12**	**16**	**8**	**7**	**12**	**9**
Age		Years	26	31	25	33	31	34	32	15	26	37	31	38
Other pathologies		N/A	H	H	HD1	H	XHD	HD1		HD1	H	H	HD2	HD1
Fasting glucose	mmol/L	4.2-6.1	3	4	11	7	4	5	8	7	5	4	7	16
Creatinine	**µmol/L**	**60-110**	**552**	**621**	**414.69**	**234.35**	**1085**	**1196**	**501**	**447**	**701.1**	**744.21**	**490**	**622**
Blood pressure	mmHg	120/80	THT*	THT*	THT*	THT*	THT*	THT*	THT*	THT*	THT*	THT*	THT*	THT*
Uric acid	**mmol/L**	**2.0-8.5**	**225**	**187**	**154.8**	**234.35**	**1085**	**1196**	**501**	**447**	**701.1**	**744.21**	**490**	**622**
Na^+	mmol/L	135-147	140	140	133.4	136.5	133.7	134	145	141	135.3	135.2	140	138
K^+	mmol/L	3.5-5.1	3.87	3.74	4.3	3.9	5.1	5.1	4.2	3.87	4.4	6	3.87	4.09
Cl^-	mmol/L	95-110	106	100	100	97	99	104	105	107	101	102	102	101
Ca^+	mmol/L	2.1-2.8	2.33	2.22	2.96	2.93	1.89	2.51	2.43	2.25	2.37	2.45	2.51	2.08
Urea	mmol/L	2.5-7.1	11.44	9.6	7.73	4.96	22.17	21.28	10.55	15'34	18.96	8.32	24.33	24.33

The bold and shaded areas are parameters most symptomatic of kidney failure. b Abbreviations: GFR, glomerular filtration rate as assessed via cystatin (GFRcy) and creatinine clearance (GFRcr); THT*: Treated Hypertension.

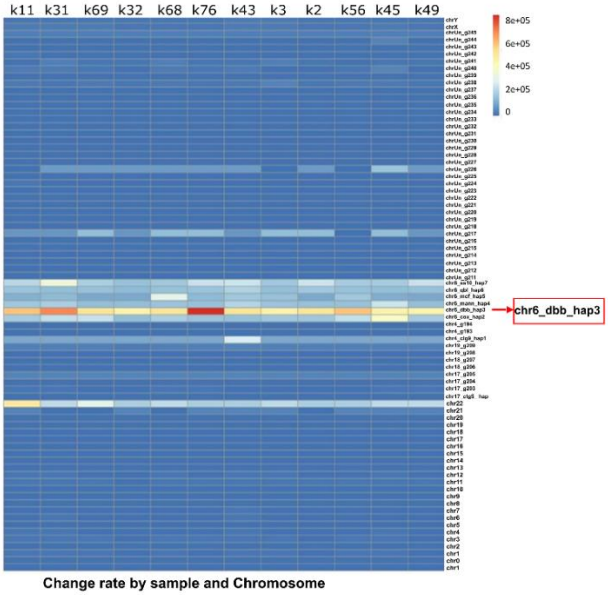

Fig. 2. Change rate by samples and chromosome as deduced from the chromosome coordinates. Patient's identification codes are presented on X axis on the top side of the figure from left to right: K11, K31, K69, K32, K68, K76, K43, K3, K2, K56, K45, and K49. The different karyotypes span the entire figure from the bottom to the top with the diagnostic karyotypes presented in color and scale level on the right top represents log2fold change the different degree of the severity in the mutations. The different chromosomes and karyotypes are presented on the Y axis.

On the top right of the Fig. 2, we presented a scale of values from zero to $8e^{+5}$ ($0\text{-}8e^{+5}$) detailing the levels and severity of the mutations; blue color means absence of mutations, light blue, less mutations, pale orange intermediate mutations, orange more mutations and red extreme mutations. This color code represents an important annotation preluding SNPs identification and sorting. The Fig. 2 also depicts a list of affected chromosomes and haplotypes and the level of the seriousness of mutations from each patient. Affected chromosomes are: one unknown chromosome (ChrUn) existing in all patients and harbors respective positions for the following haplotypes: chr17_ctg5_hap1, chr4_clag9_hap1, chr6_mann_hap4, chr6_mcf_hap5, chr6_qbl_hap6 and chr6_ssto_hap7. These chromosomes have lesser intensity in term of mutation except in patient K11 that presents pronounced level of mutations. The nature and identity of these mutations remained to be elucidated. The haplotype chr6_cox_hap2 presents a small higher level of mutation than aforementioned chromosomes except for the patient K45 that depicts a higher level of mutations compared to the other patients. The haplotype chr6_dbb_hap6 represents the major chromosome of interest as exemplified on the HeatMap. This haplotype presents a higher density of mutation among the 12 patients.

IV. DISCUSSION

There are many ways to process the avalanche of data of SNPs filtration resulting from whole exome sequencing after bioinformatics analysis. Even after running the FASQ data through GATK pipe line followed with variants filtration, diseases association and hard filtration of variants associated with disease [19, 20], we still have at least 250 variants per sample remained that will require further subsequent sorting using Bayesian probability and R-algorithm [21], [22]. Since the faulty or variants of interest are usually few, uncovering these mutations from the big number of SNPs or INDELS will necessitate synthesis of a series number of primers that will be used in subsequent Sanger sequencing for final data confirmation [23]. This increases the cost and the load of the labor. To minimize this load, we envisaged developing a short cut that can allow us to narrow down the number of SNPs. To this end, we plotted the chromosomes HeatMap from chromosomes data coordinates and the HeatMap depicts haplotypes with varieties degree of intensities. Whether, variation in these mutations as depicted by variation in $\log_2$ fold change/color intensity is a result of high number of genes mutated associated with the renal failure or simply due to genetic hierarchy remained to be demonstrated by in-depth molecular works involving transgenesis and gene KO methodologies. Nonetheless, the chromosome HeatMap has provided us with a road map to narrow down our SNPs characterization via primers designing. This approach will guide us to navigate toward

the positional sequencing of specific variants annotation in renal failure Fig 2. The data as it emerged from the HeatMap plot is phenomenal. Indeed, it depicts the haplotype Chr6_dbb_hap3 of biological importance. This haplotype is the site of the major histocompatibility complex (MHC) and is recognized as the most variable region in the human genome and has susceptibility to harbor more than 100 diseases [14], [24]. It is the site of genetic determinants for many inflammatory and autoimmune diseases as well as for some infectious diseases. This genomic region has been recently sequenced to entirety using gap filling methodology and can be used as reference to identify diseases association within this haplotype. Chr6_cox_hap2 is another haplotype of great interest in that, it contains RAGE locus that mediates interactions of advance glycosylation end products (AGE). These are non-enzymatically glycosylated proteins accumulate in vascular bed during aging with subsequent acceleration of diabetes [25], [26]. Indeed, in human diabetes, the kidneys display a marked increase of AGE and RAGE [27] and RAGE and AGE were also increased in human diabetic subject as well as in ESRD patients as compared to non-diabetic human [28]. The other haplotype which merits putting emphasis on, because of its clinical importance is Chr6_ssto_hap7. This haplotype is the site of expression atp6v1g2 which encodes a component of vacuolar ATPase (V-ATPase. In our further Sanger characterization of our SNPs resulting from GATK pipeline, we observed via Sanger sequencing that, this gene was mutated in our 12 patients as well as 100 other End-stage renal diseases (ESRD) patients. We subsequently confirmed the V-ATPase protein mutation with our collaborators on their proteomic suite pipe line at EBI, UK, Saffron Walden (data not shown). Taken together, the chromosomes HeatMap approach represents an excellent strategy for variants annotation and prioritization. The methodology allows us to refine our data stemming from GATK pipe line, an approach that generates a strong lead toward narrowing down the flurry of SNPs stemming from bioinformatics analysis of Exome sequencing data.

In conclusion, profiling genomic gDNA from 12 ESRD unrelated patients using next generation sequencing, we identified more than 250 variants per patients after hard filtration with different algorithms. The plotting of the chromosomal coordinates using HeatMapper revealed genes karyotypes of great clinical significance allowing annotation or mapping of gene sites mutations of specific chromosomes from these 12 ESRD patients. *In-silico* tools remain a valuable approach to predict and discriminate the emergence and severity of mutations inside the genome. Further studies are required to vet these haplotypes against our SNPs data base and ultimately develop genetic modified animals aiming at knock-in specific region of chromosome of interest into mouse genome.

ACKNOWLEDGMENT

This work was funded in part by Research grant offered by Saudi Renal Chair to Dr. Mohd Ibrahim Alaraj. Exome sequencing, bioinformatics analysis and chromosomes heatmap studies are funded Saudi Renal Chair and the Center of Molecular Diagnostic and specialized medicine through a funding Dr. Edem Nuglozeh (PI) of the project.

AUTHOR CONTRIBUTIONS

M.F.F and E.N purified the genomic DNA, prepared the libraries, and performed the exomes sequencing on Illumina Miseq platform. Data were back up and FASTQ data were generated by both M.F.F and E.N. Bioinformatics analysis and SNPs generation were conducted at McGill University, Center of Bioinformatics, Canada. A.A.W and M.I.A performed the anthropomorphic works leading to clinical, and biochemical parameters generation presented in table 1. M.F.F plotted the chromosomes Heatmap and E.N wrote the manuscript. M.F.F and E.N reviewed and finalized the manuscript.

REFERENCES

[1] Duni A, Liakopoulos V, Rapsomanikis KL, Dounousi E. Chronic Kidney Disease and Disproportionally Increased Cardiovascular Damage: Does Oxidative Stress Explain the Burden? Oxid Med Cell Longev. Oxid Med Cell Longev. 9036450, http//doi:org 10.1155/2017/9036450 (2017).

[2] Köttgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M et al. multiple loci associated with indices of renal function and chronic kidney disease. Nat Genet.41 (6): 712–717, http//:doi:10.1038/ng.377. (2009).

[3] Kottgen A, Pattaro C, Carsten A, Böger CA, Fuchsberger C, Olden M et al. New loci associated with kidney function and chronic kidney disease Nat genet 42: (5): 376-384, http//: doi:10.1038/ng.568. (2010).

[4] Chambers C Zhang W, Lord GM, Van der Harst P, Lawlor DA, Sehmi JS et al. Genetic loci influencing kidney function and chronic kidney disease in man. Nat Genet, http//:doi:10.1038/ng.566. (2010).

[5] Foley RN, Collins AJ. J Am Soc Nephrol 18: 2644 –2648, 2007, http//: doi: 10.1681/ASN.2007020220 (2007).

[6] Grassmann A, Gioberge S, Moeller S, Brown G. ESRD Patients in 2004: Global Overview of Patient Numbers, Treatment Modalities and Associated Trends. Nephrol Dial Transplant. Feb;22(2):663-5, http//: doi: 10.1093/ndt/gfl596. (2007).

[7] Modi, GK, Jha, V. The incidence of end-stage renal disease in India A population-based study. Kidney International, (70), 2131–2133. http://www.kidney-international.org/ doi:10.1038/sj.ki.5001958. (2006).

[8] Fletcher J, McDonald S, Alexander SI. Australian, New Zealand Pediatric Nephrology A. Prevalence of genetic renal disease in children. Pediatr Nephrol. 28(2):251–256. http//: doi: 10.1007/s00467-012-2306-6. (2013).

[9] Mallett A, Patel C, Salisbury A, Wang Z, Healy H, Hoy W. The prevalence and epidemiology of genetic renal disease amongst adults with chronic kidney disease in Australia Orphanet Journal of Rare Diseases 9:98 http://www.ojrd.com/content/9/1/98, http//:doi: 10.1186/1750-1172-9-98 (2014).

[10] Levy, Micheline, and Josué Feingold. Estimating prevalence in single-gene kidney diseases progressing to renal failure. Kidney International 58 ;(3): 925-943. https://doi.org/10.1046/j.1523-1755.2000.00250. (2000).

[11] Sadowski CE, Lovric S, Ashraf S, Pabst WL, Gee HY, Kohl S. SRNS Study Group A single-gene cause in 29.5% of cases of steroid-resistant nephrotic syndrome. J Am Soc Nephrol. 2015; 26:1279–89. http//:doi.org/10.1681/ASN.2014050489. (2015).

[12] Lovric S, Fang H, Vega-Warner V, Sadowski CE, Yung Gee HY Halbritter J et al. Nephrotic Syndrome Study Group. Rapid detection of monogenic causes of childhood-onset steroid-resistant nephrotic syndrome. Clin J Am Soc Nephrol; 9:1109–16. http: doi.org/ 10.2215/CJN.09010813. (2014).

[13] Hwang, DY, Dworschak, GC, Kohl, S, Saisawat, P, Vivante A, Hilger, AC, et al. Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary

tract. Kidney International, 85, 1429–1433, http//:doi.org/10.1038/ki.2013.508. (2014).
[14] Zhang Y, Zhang T, Lu Z. Gap Filling for a Human MHC Haplotype Sequence American Journal of Life Sciences, 4, No. 6, pp. 146-151, http: doi.org/ 10.11648/j.ajls.20160406.12 (2016).
[15] Sarnataro S, Chiariello AM, Esposito A, Prisco A, Nicodemi M Structure of the human chromosome interaction network, PLoS ONE 12(11): e0188201, http: doi.org/ 10.1371/.0188201. (2017).
[16] Alaraj, MI, Alaraj, N, Hussein, TD. Early Detection of Renal Impairment by Biomarkers Serum Cystatin C and Creatinine in Saudi Arabia, Journal of Research in Medical and Dental Science 5(1):37/ http//:doi.org/ 10.5455/jrmds.2017518. (2017).
[17] Nuglozeh, E, Fazaludeen, FM, Mbikay, M, Hussain, AG, Al-Hazimi, A, Ashankyty I. Whole-Exome Sequencing Reveals an M268T Mutation in the Angiotensinogen Gene of Four Unrelated Renal Failure Patients from the Hail Region of Saudi Arabia. International Journal of Medical Research & Health Sciences, 6(6): 144-149/ www.ijmrhs.com. (2017).
[18] Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A et al. Heatmapper: web-enabled heat mapping for all Nucleic Acids Research. Vol. 44: W147–W153, http.org/10.1093/nar/gkw419. (2016).
[19] Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetics parameter estimation from sequencing data. Bioinformatics 27(21) 2987–2993, http.org/ doi: 10.1093/nar/gkw419.10.1093/bioinformatics/btr509. (2011).
[20] Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat. Commun. 9: 2337, http.org/10.1038/s41467-018-04737-0 (2018).
[21] Dakal T C, Kala D, Dhiman G, V Yadav V, Krokhotin A. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. Scientific RePorTS | 7: 6525. http.org/ doi: 10.1038/s41598-017-06575-4. (2017).
[22] Mansouri, M, Nounou, H, Nounou, M. Genetic Algorithm-based Adaptive Optimization for Target Tracking in Wireless Sensor Networks. J Sign Process Syst. 74:189–202. http.org/doi/10.1007/s11265-013-0758. (2014).
[23] Nuglozeh, E. Whole-Exomes Sequencing Delineates Gene Variants Profile in a Young Saudi Male with Familial Hypercholesterolemia: Case Report. J Clin Diagn Res. 2017 Jun; 11(6): GD01–GD06, http.org/doi/10.7860/JCDR/2017/28156.10143. (2017a).
[24] Nuglozeh, E. (2017b). Signature of Chromosomes Instability in Different Diseases as Accessed on Illumina Miseq Platform using Depth of Coverage Metrics for Variant Evaluation by GATK, International Journal of Science and Research, 6(2)/ www.IJSR.net ISSN (Online) 2319-7064/, http.org/ doi/ 10.21275/ART2017972. (2017b).
[25] Brownlee, M. (1992). Glycation products and the pathogenesis of diabetic complications Diabetes Care, 15, 1835–1843, http.org/doi/10.2337/diacare.15.12.1835 (1992).
[26] Brownlee, M. Advanced glycation end products in diabetic complications Curr. Opin Endocrinol. Diabetes, 3, 291–297, (1996).
[27] Cipollone F, Iezzi A, Fazia M, Zucchelli M, Pini B, Cuccurullo C et al. The receptor RAGE as a progression factor amplifying arachidonate-dependent inflammatory and proteolytic response in human atherosclerotic plaques role of glycemic control Circulation 108:1070–1077, http.org/doi/ 10.1161/01.CIR.0000086014.80477. (2003).
[28] Tanji N, Markowitz GS, Fu C, Kislinger T, Taguchi A, Pischetsrieder M. et al., Expression of advanced glycation end products and their cellular receptor RAGE in diabetic nephropathy and nondiabetic renal disease. J. Am Soc. Nephrol. 11:1656, (2000).