Exome Sequencing on Illumina Platform revealed Relation between Nucleic Acids Interconversion and the Ratio of Transition/Transversion in Chronic Kidney Disease (CKD) Patients

Edem Nuglozeh[1,2]* and Mohammad F. Fazaludeen[3,4]

[1]Ondokuz Mayis University. Department of Biochemistry, School of Medicine of Medicine Kurupelit Campus, 55139 Atakum Samsun, Turkey

[2]University of Hail, Department of Biochemistry, School of Medicine, Saudi Arabia

[3]A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland

[4]Fin Vector, Kuopio, Finland

[*]Corresponding e-mail: Nuglozeh@gmail.com

ABSTRACT

Aims: Renal Failure and Chronic Kidney Disease (CKD) represent major health problems that affect more than 10% of the general population worldwide amounting to more than 800 million individuals. CKD is a congruence of many diseases like Diabetes, Hypertension, hypercholesterolemia, and different cardiovascular diseases as well as epigenetic factors contributing to the development of this Disease. We undertook this work as a pilot study to establish a database of genetic variants for novel or existing genes involved in the development and progress of CKD in the Saudi population. Methods: Patients' blood samples of 12 donors suffering from CKD were used to perform the Whole Exome Sequencing (WES) on genomic DNA on the Illumina Platform. Bioinformatics algorithms were used to perform different callset to explore different mechanisms such as nucleic acids interconversion, INDELS search as well as Transition/Transversion (Ts/Tv) ratio determination. Results: Patterns of Transition and Transversion Ratio (Ts and Tv) in the Patient samples demonstrated the dominance of transition substitutions over transversion. Our data revealed the mean ratio metrics of (Ts/Tv) is roughly 2.75 which is consistent with former literature findings. Indels with less than 3 bp have a total count of more than 1e+05 whereas those with larger gap are present at the level of trace. Since there is an intrinsic relation between nucleotides interconversions and the transition and transversion mechanisms, on the basis of higher rate of transition over transversion, our study supports this assertion at the levels of the nucleotides, with the state of the nucleotides conversion being as followed: (67% A→G, 64% G→A, 67% T→C, and 67% C→T) and situation which mirrors the dominance of transition effects and lower representation of transversion effects as exemplified by lower state of interconversion of: (13% T→A), (12% A→T), (19% C→G), (17% C→G), (17% G→T) and (20% T→G). Conclusion: CKD is a widespread disease that overlaps the spectrum of many diseases' spectrum in terms of gene variant expressions. The transition and Transversion ratio profile is around 2.75. Nucleotides interconversions rim with the concept of

transversion and INDELS counts and distribution favor the shorter length as compared with larger INDELS.

Keywords: CKD, WES, Nucleotides conversion, Transition/Transversion ratio

INTRODUCTION

The reflection on the structure and origin of the Standard Genetic Code (SGC) has been baffling biologists for a long time since the discovery of the first code [1,2]. This discovery is almost universal with some minor exceptions, with the coding rules being responsible to transmit information stored inside the DNA is translated into the proteins universe. The code uses all 64 possible nucleotide triplets' codons that are possibly digitalized into 20 canonical amino acids as well as three translation stop codons. In this process of digitalization, the total number of codons outnumbers the number the encoded labels amino acids and this represents the concept of degenerescence of the genetic code which means that some amino acids must have more than one codon with the redundant codons also called synonymous which are organized in specific groups. In most situations, codons in such groups differ from one another at their third degenerated nucleotide position, which rim with Francis Crick's concept that, only the first two codon positions were important in primordial code [3]. The genetic code is also known as the energy code and all codons evolve from each other following the laws of thermodynamics via an ATP-centric concept, from energy transformation to informatization by series of transition and transversion. The redundancy in the standard genetic code generates a peculiar relation with the process of single nucleotide mutations. If these changes take place at the third degenerated nucleotide position in the codon or wobble position, the mutated amino acid will be identical to the original one. The mutation of such is called synonymous, whereas those that change the original amino acids or introduce translation stop codon are termed non-synonymous. However, when the mutation changes the original amino acid by a different amino acid, this is called substitution. The progress made in whole human genome sequencing has provided us with unique tools and occasions for an in-depth human genome study, therefore to comprehend the concept of transition and transversion. With the rise of modern sequencing methodologies, so come a flurry of sequencing data encompassing many genetic mutations. In substitution mutation processes, transitions are defined as the interchange of the purine-based A↔G or pyrimidine-based C↔T. Transversions on the other hand are defined as the interchange between two-ring purine nucleobases and one-ring pyrimidine bases. The possible transversions are A↔C, A↔T, C↔G, and G↔T. There are four possible transitions and eight possible transversions. Transitions are observed more often in nucleotide sequence substitution than transversions [4-12]. The transition may be a result of a higher mutation rate than transversion due to similarities in the physicochemical properties of the nucleotides. Besides, transitions are accepted with a greater probability than transversion, because they rarely lead to amino acid substitutions in encoded proteins due to the specific codon degeneracy and transitions are also more frequent during protein synthesis [12]. If substitution mutations occur randomly, then the Ts/Tv ratio should be 0.5, because there are two possible transitions and four possible transversions. However, a transversion is considered to be thermodynamically a more drastic change than a transition, because of the substitution of one-ring to two-ring chemical structure or vice versa and transversions require more energy than transition substitution without

change in the ring structure.

Thus, in the realm of sequencing data, the transition and transversion ratio is often greater than 0.5. The Ts/Tv ratio has been used as an important parameter in many studies such as phylogenetic tree reconstruction and estimation of divergence evolution. Recently, the transition/transversion ratio has also been used as a QC parameter in high-throughput sequencing studies [13-17].

For human-exome sequencing data, the Ts/Tv ratio is generally around 3.0, and about 2.0 outside of exome regions [18]. The Ts/Tv ratio is also different between regions of synonymous and non-synonymous SNPs [19]. The Ts/Tv ratio for the haploid chromosomes (X in males Y, mitochondria) is different compared to the one in diploid chromosomes (chromosomes 1-22). Much stronger bias toward transitions over transversions (Ts/Tv is between 21 and 38) in mitochondria DNA has been observed in multiple studies [15,20]. It has been suggested to consider haploid and diploid chromosomes separately when computing Ts/Tv ratios [17].

The mechanism of gene mutation-induced disease remains challenging. Some mutations can be deleterious causing diseases whereas some will be beneficial. For example, a missense variant mutation in apolipoprotein A1 on chromosome 11 characterized by R[CGC]>C[TGC], i.e., conversion of *arginine* to *cysteine* confers *eucholesterolemic* status to the Milano patients. This mutation is called APOA1 Milano [21-24]. Recent studies have indicated that Single Nucleotide Polymorphisms (SNPs) appear in human DNA approximately every 200 bases (Ensembl release 53.36o) and SNPs which are the most common forms of human genetic variations have correlated to human evolution, drug sensitivity, and disease susceptibility [25-29]. Trivially, non-coding region SNVs are more common than coding region SNVs. However, fewer non-coding variants have thus far been characterized as disease-causing than coding variants and non-synonymous SNVs (nsSNVs) [30]. Many human diseases are monogenic and identifying SNVs causative of monogenic diseases is straightforward [31]. These faulty genes are always functionally disruptive and consistently present in the disease population, but less frequently in healthy control populations [32]. Complex genetic diseases, on the other hand, are generally caused by a combination of moderately deleterious mutations in different genes, often leading to a disruption of the broader functional networks involved Chronic Kidney Disease (CKD) represents a proper model of complex genetic diseases and any one of SNVs uncovered in our study is unlikely to be significantly visible throughout the background of human genetic variation [33,34]. In our search for the underlying causes of Chronic Kidney Disease (CKD), we reasoned that finding a new gene involved in the development and progression of the disease will be the right thing to do and to this end we run exome sequencing on the Illumina platform and uncovered a plethora of Single Nucleotides Polymorphism (SNPs) that we analyzed using various bioinformatics algorithms. These SNPs were filtered against different disease databases for potential genes associated with CKD status and these results are submitted for publication elsewhere. We also inventory by karyotyping different chromosomes linked to this disease and we reported these data in the format of chromosome HeatMap and CKD and this report was published in the European Journal of Medical and Health Sciences [35]. In this work, we are attempting to comprehend and explain the role played by different mechanisms of genetic mutations, *viz*: Nucleotide conversion, Insertions, and Deletions (Indels) as well as the implication of transition and transversion in the development of CKD. The results we are reporting here, support the rationale of this study.

MATERIALS AND METHODS

The patient's blood samples from 12 donors were used for exome sequencing and further processed independently. Genomic DNA specimens are purified from samples as published elsewhere [35]. This study is part of a subgroup of an investigation started a while ago on hemodialysis patients from outpatient clinics of King Khaled Hospital and published with ethical approval issued to Dr. Alaraj [35,36].

Libraries preparation was essentially obtained as we described it elsewhere [35]. In brief, we purified high-quality genomic DNA (gDNA) from whole blood using Qiagen kits (QIAamp DNA Blood Mini Kit from QIAGEN, Hilden Germany). Libraries were built using the Illumina Kits system based on Transposase enzymology. For each of the blood samples, genomic DNA was enriched for the target regions of all human CCDS exons in the genome with Illumina probes included in the sequencing kit (see description below). The enrichment of adapter-modified DNA fragments before sequencing includes an amplification step of 18 cycles with Polymerase Chain Reaction (PCR) in the standard protocol. For one exome, 36 cycles of PCR were run to analyze the effect of the cycle number on the allele frequency distribution. The cluster generation step follows after the library preparation. Its purpose is to increase the fluorescent signal of a fragment on the sequencing flow cell so that it becomes detectable by the camera. The cluster generation includes another 35 PCR cycles in the standard protocol. The raw data of 5 GB per exome was mapped to the haploid human reference sequence hg19.

We later proceeded to QC control of the libraries using Bioanalyzer (Agilent Technology, Santa Clara, CA, USA). The libraries were subsequently sequenced on Illumina MiSeq Platform using MiSeq Reagent Kit optimized v3 Chemistry, 150-cycles to increase cluster density and read length as well as to improve quality (Q) scores. We run the sequencing forward and reverse.

Bioinformatics Methodology

Different bioinformatics pipelines as well as the flowchart we used to conduct the analyses were reported elsewhere [35]. The FASTQC analyze data from CKD patients generated more than 300 SNPs per patient (data not shown).

- FASTQC files were concatenated from multiple runs of the libraries.
- Adapter trimming and base quality scores (less than Q30) were removed using Cut adapt (V1.7.1).
- FASTQC (V0.11.2) was used to check primary and post-trimmed sequences.
- Alignments to the reference human genome (hg19) were conducted using BWA (version 0.7.15).
- The Genome Analysis Tool Kit, GATK (version 3.0.0), was used for base quality score recalibration, variant calling followed by hard filtering to identify high-quality variants for downstream analyses.
- SnpEffv4.1 was exploited to determine in silico impacts upon the protein function of candidate genes. The Fast QC files analyses generated more than 300 SNPs per patient and these SNPs were subsequently later committed into reduction variants analysis that we termed disease association filtration. This filtering process from more than 1000 SNPs generated total variants of two single SNPs associated with two single proteins that we reported in another publication.

RESULTS

Patterns of Transition and Transversion Ratio (Ts/Tv) in CKD Patient Samples

Figure 1 depicts base counts and base substitutions of nucleic acids resulting from FASTQC data analysis of CKD patients. In principle, on a molecular basis, if the process is random, transversions, i.e., (purine–pyrimidine changes) should be observed twice as often as transitions (purine to purine or pyrimidine to pyrimidine changes) solely due to the accessible mutations. However, aside from insertional and deletional mutations, nucleotides substitutions are plagued with bias thereby justifying the profile of our results characterized by the dominance of transition substitutions over transversion and our data goes in the same line in these observations with the ratio being 1:4 (Ts/Tv) and previous works from human genome samples reported that transversion substitutions are approximately four times rarer transitions substitutions [37,38].
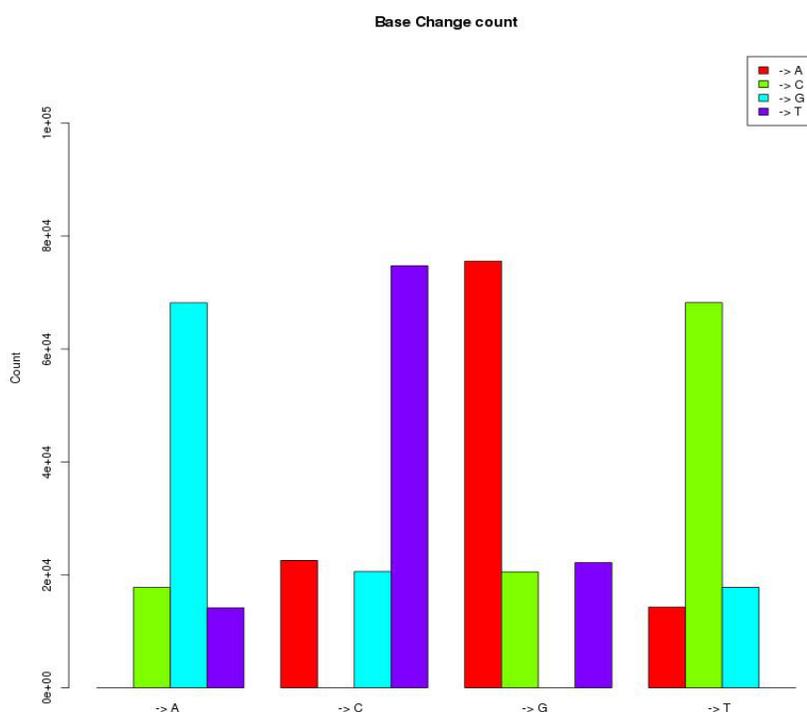

Figure 1 Substitution patterns between nucleotides pairs. The figure presents the breakdown of variants by nucleotide change highlighting the proportion of transitions and transversions in exome sequencing from CKD patients. A→G>T→C>C→T>G→A. This data shows a domination of distribution of transition far way over the transversion roughly by margin of 1:4

Transition and Transversion Rate by Sample in CKD Patients as Revealed in Exome Sequencing

In Figure 2, we reported some biological metrics representing the estimation of the (Ts/Tv) ratio per patient. These metrics are also quite useful for quality control and one would expect them to be roughly consistent across the genomes from individuals to individuals of the same ethnicity even between different methods of genomes sequencing and the application of these metrics to normal diploid genomes is relatively clear. We can also notice that the ratio of (Ts/Tv) for the known variants is a little bit higher than those of unknown variants and the biological significance of this discrepancy remains unclear. Our data on the other hand demonstrated that these biological

metrics estimation of (Ts/Tv) is accurate because, in principle, random sequencing errors do appear as novel gene variants during the call set. From our data, it is revealing that the ratio metrics of (Ts/Tv) that we presented in Figure 2 as well as in Table 1 are consistent between the patients which is roughly 2.75.
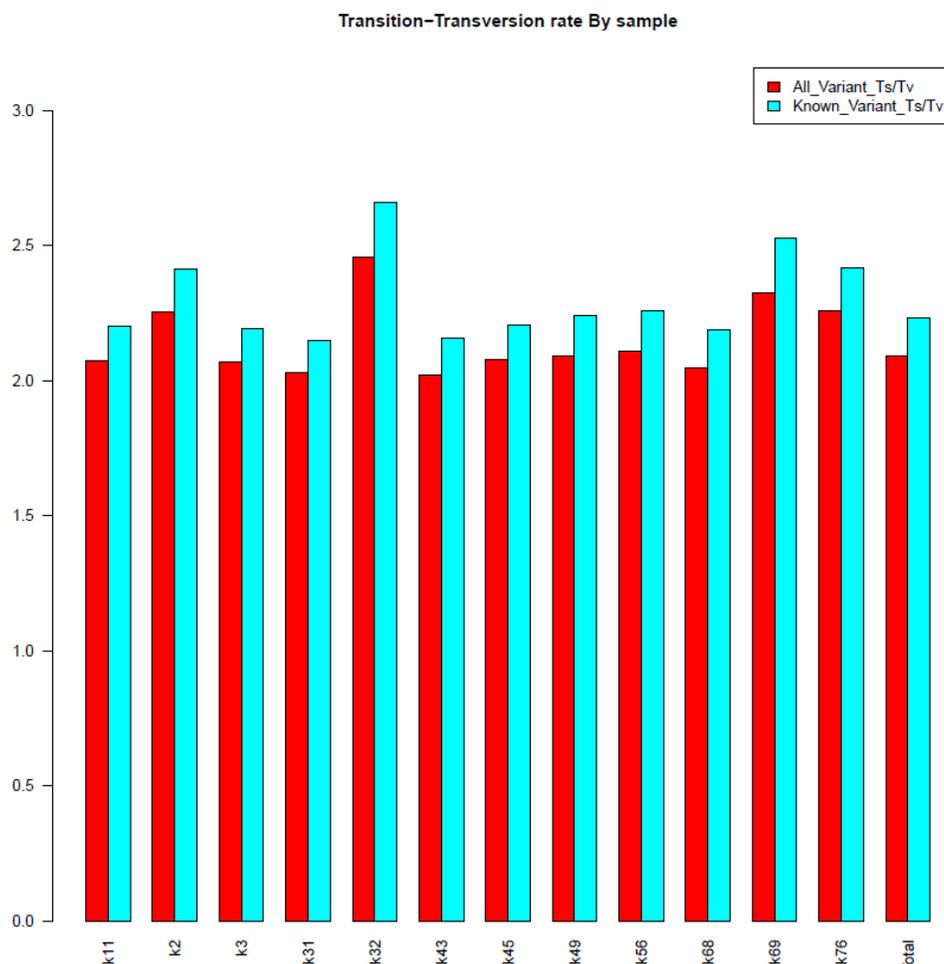
Transition–Transversion rate By sample

Figure 2 Ratio representation of Ts/Tv for each patient. The blue color represents the known variants and red color for total variants. The ratio is computed as the number of transition SNPs divided by the number of transversion SNPs for all known variants as well as unknown variants

Table 1 Summary of exonic coverage in the twelve CKD patients sequenced exomes. Coverage was defined as the percentage of bases in the exome that have at least 5 reads with a Phred-like consensus score of greater than zero at that position. The total size for the autosomal exons is 65,471,109 bp, which is the total length of the autosomal exons, defined as all protein coding gene entries in Ensembl core database version 50 [13]. The Ensembl database version 50 is based on the NCBI human genome assembly build 36 as well as its annotations (GenBank). doi: 10.1371/journal.pgen. 1001111.t001. 13 is reference in The Characterization of Twenty Sequenced Human Genomes publication

Patients ID	Mapped Reads	% Mapped	Exonic Coverage autosome	All Variants Transition	All Variants Transversion	Know Variants Transition	Know Variants Transversion	Ts/Tv Ratio (All variants)	Ts/Tv Ratio (Known variants)
k2	1,87,15,604	99.9298	14.80x	59519	26405	55796	23141	2.254	2.411
k3	4,43,34,278	99.9301	49.66 x	117092	56606	109356	49862	2.069	2.193
k11	4,39,58,602	99.9357	52.41 x	116983	56442	109378	49696	2.073	2.201
k31	5,44,27,722	99.9362	58.47 x	134553	66310	126324	58756	2.029	2.15
k32	10,80,540	99.7557	0	8948	3643	8447	3178	2.456	2.658
k43	5,61,54,770	99.9336	58.70 x	144952	71703	136212	63179	2.022	2.156
k45	4,31,16,046	99.9463	51.44 x	123771	59556	116166	52683	2.078	2.205
k49	4,42,99,004	99.9143	46.59 x	108038	51703	100736	44970	2.09	2.24
k56	4,19,74,572	99.938	43.03 x	101338	48057	94747	41983	2.109	2.257
k68	5,52,99,402	99.9331	57.39 x	140280	68447	131264	59949	2.049	2.19
k69	95,87,168	99.9174	2.37 x	42359	18209	39839	15754	2.326	2.529
k76	2,08,37,604	99.9286	19.30 x	67130	29731	62957	26055	2.258	2.416
Total				1164963	556812	1091222	489206	2.092	2.231

Length of INDELS Counts Distribution

The number of insertions and deletions, lengths, and frequency distribution are shown in Figure 3 for all the patients. Apart from substitutions, genomic DNA is also the site of random indels. Indels occur much less frequently than substitutions, therefore a larger amount of DNA sequences is needed to characterize them and our CKD dataset model fits very well apropos for these characterizations of Indels and their sizes. The maximum Indels length we observed was 107 bp with a total count close to zero. It is conceivable to assume that Indels frequency with larger gap lengths should occur more rarely than those with shorter lengths, and our data in Figure 3 replicates exactly the trend where we observed that those Indels with less than 3 bp have a total count of more than $1e^{+05}$ whereas those with larger gap are present at a level of trace and we will tackle the biological significance of this discrepancy in the discussion. In most cases, deletions are more deleterious than insertions but insertions and deletions are otherwise generally governed by the same genomic factors. Deletions are responsible for an array of genetic disorders, including some cases of male infertility, two-thirds of cases of Duchenne muscular dystrophy, and two-thirds of cases of cystic fibrosis [39].

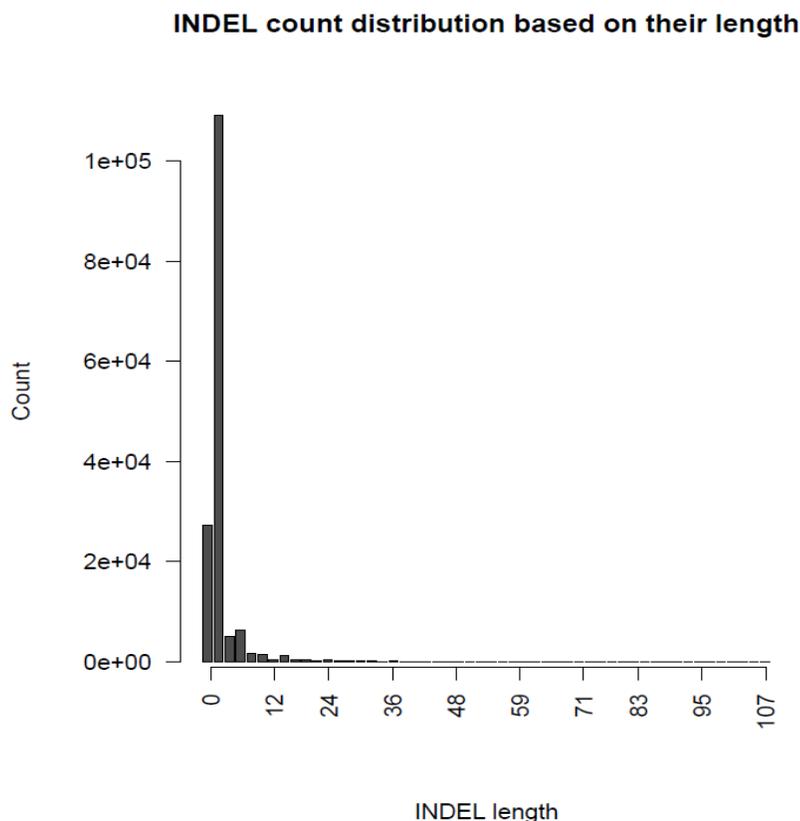## INDEL count distribution based on their length

Figure 3 Frequency distribution of insertion and deletion event lengths (in bp). The X-axis depicts the categories of the event length and on Y-axis shows, the number or frequency of count (raw counts) for observed indel events from each event length, type, and species category. The X-axis spans up to 107 codons and indels were observed up to 36 codons length

Nucleotides Bases Substitution Patterns

In our study of our patients with CKD disease, we were interested in discovering new genes involved in the pathophysiology of kidney diseases. We inventory multiple variants concomitant to some mutational events like the generation of Indels, nucleotides interconversion, and their frequencies. If we denote by T_α, the total number of types α nucleotides (α=A, G, T, and C) in the total genes pool from our CKD patients and $N_{\alpha \to \beta}$ as the number of times a nucleotide is mutated from α types to β types in the same genes pool, then we can set of rates of substitution ($K_{\alpha \to \beta}$) can be formulated as:

$$K_{\alpha \to \beta} = (N_{\alpha \to \beta})/T_\alpha \tag{1}$$

Where $R_{\alpha \to \beta}$ is the rate of nucleotides of α types mutating to β types in the CKD genes pool.

Alternatively, instead of assessing how often one type of nucleotide will be mutating to another one, we can just assume that, since a mutation has already occurred, it will be good to find the relative frequency that this mutation reached one of the three other types. In other words, instead of normalizing "$N_{\alpha \to \beta}$" (the number of times a nucleotide is mutated from α types to β types) by T_α, we will normalize it with S_α, symbolizing the total number of

mutations that have occurred in type α nucleotides. We can define $P_{\alpha\to\beta}$ as the proportion of substitutions in that gene pool.

$$P_{\alpha\to\beta} = \frac{N_{\alpha\to\beta}}{S_\alpha} = \frac{N_{\alpha\to\beta}}{\sum_{\alpha\neq\beta} N_{\alpha\to\beta}} \tag{2}$$

Equations (1) and (2) describe two different sets of statistics of the nucleotide's interconversion in the CKD gene pool, and dividing equation (1) by (2) will establish a relationship between the two sets of statistics. We reported the values for T_α and S_α for each type of nucleotide in Table 2. Moreover, it has also been reported that the rates of nucleotide substitution in the human genome vary according to different genomic environments, most especially with regions of different G+C contents [40].

Neighboring Effects on the Substitutions

It has been proposed that nucleotide substitutions have a neighboring bias, i.e., the chance that a specific nucleotide is mutated and the type of nucleotide that it is mutated to are affected by the adjacent flanking nucleotides [41]. For example, the single nucleotide substitution A→C is more than twice as frequent in the di-nucleotides TpA than in ApA. The four transitional substitutions: C→T, G→A, A→G, T→C, and the transversion of T→A are also significantly affected by the 5' neighboring base [42]. One needs to be cautious when interpreting the rates of that substitution that result in CpG islands since it is likely that the majority of the resulting CpG di-nucleotides have mutated to TpG soon after the original substitution event. However, such secondary substitutions do not affect the calculated rates for other substitutions that also result in TpG such as ApG→TpG or TpA→TpG.

DISCUSSION

Nucleotides Interconversion Patterns and their Relations with Transition and Transversion

Recent estimates indicate that Single Nucleotide Polymorphisms (SNPs) occur approximately every 200 bases in the human genome (Ensembl release 53.36°). SNPs have been associated with human evolution, drug sensitivity, and disease susceptibility [21-24]. An international effort has been undertaken through the HapMap project for the determination of common patterns of DNA sequence variation in the human genome and its relation to common diseases [43,44]. Then, the occurrence of SNPs through many mechanisms of mutation *viz*: nucleotides interconversion, substitution insertion, and deletions are important avenues to explore and decipher the causes of most endemic diseases like CKD. To achieve this goal, we run Whole Exome Sequencing on the Illumina platform from human CKD genomic DNA. We reported in Figure 1, that the base change counts representing the distribution pattern of nucleotides interconversion in CKD patients showed a domination of transition proportion compared to that of transversion as ascertained by the nucleotides profile: A→G>T→C>C→T>G→A which is roughly 1:4. Transition is the interconversion of one ring nucleotide with another one ring nucleotide or two rings nucleotides with two rings, whereas transversions involve interchanges of one-ring and two-ring structures (A↔C, A↔T, G↔T, G↔C). We should also take note that, when we are talking about nucleotides interconversion, we directly discussing the concepts of transition and transversion. Approximately two out of three Single Nucleotide

Polymorphisms (SNPs) are of transitions feature highlighting the ground of the nature of the conservation of transitional events [37, 38]. Even though the number of possible transversions is twice as many as the number of transitions, leading to a Ts/Tv ratio of 0.5, if mutations occurred at equal rates, the actual Ts/Tv ratio will differ by genomic regions due to the environment of the genetic mutation sites [44]. It was hypothesized that transitions are less severe than transversion concerning the chemical properties switch between the original and mutant amino acids" [45] or "tend to cause changes that conserve the chemical properties of amino acids" [46]. On the other hand, transversions are less common than transitions on the ground of chemical structure constrain of the substitution of one by two rings or vice versa, and our current data in Figure 2 support this assertion. For studies dealing with the human genome, the Ts/Tv ratio is around 3.0 for SNPs inside exons and about 2.0 elsewhere [47], and this ratio also differs between synonymous and non-synonymous SNPs [48]. The nature of primers included in exome capture kits often and time overstretch to the intervening sequences and thus carry out during the amplification of some intronic regions and thereby explaining the variation of the Ts/Tv ratio. In Figure 2, we observed on average a Ts/Tv ratio for each CKD patient of around 2.75 for known variants and between 2.0 to 2.2 for unknown variants. The Ts/Tv ratio of SNPs inside these target regions is expected to lie between 2.0 and 3.0 with those values depending on the fraction of exons inside target regions [47] and our observations are in parfait agreement with the literature. However, any value of Ts/Tv ratios in exome sequencing below 2 should be cause for concern, because if the Ts/Tv ratio is too low, this means that your call set likely has more false positives.

The reason for this widespread distribution of Ts/Tv bias throughout the genome remains unknown. Two main hypotheses militate in favor of the bias: the mutational hypothesis and the selective hypothesis. The mutational hypothesis stipulates that, in transition substitutions, the levels of polymerases are higher than in transversion substitution as was ascertained by some works [42]. These authors observed that transitional bias is observed together in non-coding and coding segments and mutation analysis shows higher transitional rates [48], whereas the selective hypothesis posits that natural selection disfavors transversions and this later hypothesis repose on codon usage, nonsynonymous transitions are more likely to conserve important biochemical properties of the original amino acid [42]. For example, a mutation that changes the charge of an amino acid is a radical change, whereas the one that does not is a conservative change. However, this provides only indirect evidence for the selective hypothesis and the extent to which radical/conservative distinctions are predictive of fitness remains unclear. Radical changes do occur less frequently than conservative ones during protein evolution. The nucleotides interconversion that we reported in Table 2 and Table 3 replicate other authors' findings based on the higher rate of transition over transversion as ascertained by the state and the magnitude of conversion of the nucleotides: (67% A→G, 64% G→A, 67% T→C, and 67% C→T). Those nucleotides conversion that mirrors the effects of a transversion is represented by a lower state of interconversion as exemplified by: 13% T→A, 12% A→T, 19% C→G, 17% C→G, 17% G→T and 20% T→G. We can see that the surprising feature in humans' variant lists is that, C->T changes (with C as a reference, and T as a variant) are more frequent than T->C changes. Likewise, G->A changes are more frequent than A->G changes and this is reflected in our Table 2. So why the reciprocal changes do not equate to the initial changes? The explanation lies in the fact that the major mechanism for new mutations (in warm-blooded animals) as well as human beings is the deamination of 5'-methyl C to uracil in RNA

(equivalently T in DNA) producing a change of (C->T) or, on the complementary strand, (G->A). This was first studied for CpG dinucleotide sites, but it also occurs at lower rates throughout the genome at any C sites whether followed by G or not. More often, we expect the reference genome to include the most common allele, which is also likely to be the ancestral allele. Thus, if C->T mutations are more common than T->C mutations, we expect to see an imbalance of C->T versus T->C changes. CKD patients are plagued with increased oxidative stress [49,50]. In the state of oxidative stress, there is an increased concentration of malondialdehyde in the urine generated by lipids peroxidation [51] and this product impaired the function of antioxidant systems because of low levels of superoxide dismutase and glutathione (GSH) peroxidase has been reported in hemodialysis patients [52]. These products also induce chemical changes in proteins, lipids, and nucleic acids. Oxidative stress can induce DNA or nucleic acid damage, such as base and sugar modifications, covalent crosslinks, and single and double-stranded breaks [53]. The DNA bases, especially Guanine (G), are particularly susceptible to oxidation, leading to oxidized guanine products. Nucleobase modifications most frequently involve 8-hydroxy-2'-deoxyguanosine (8-OH-dG), one of the most abundant oxidative products of nucleic acids [54]. With all these in mind, we replicated these later observations in our data where 67% of Guanine (G) is converted to Adenine (A). We should also note that 67% of Adenine (A) was converted to Guanine (G). The most common base substitution arising from oxidative damage of DNA is a GC→AT transition [55,56]. This substitution is also the most abundant genetic change induced as a consequence of oxidative DNA damage [57,58] and this transition is mediated via a rise in (8-oxo-dG leading to G→T leaving unidentified the genesis of GC→AT mutations [59-61]. Recently, several studies have suggested that oxidized cytosines-5-hydroxylysine (5-OH-C) and/or 5-hydroxyuracil (5-OH-U) might plausibly contribute significantly to the GC→AT mutations observed in *Escherichia coli*. Oxidation of cytosine can give rise to 5,6-dihydroxy-5,6-dihydrocytosine (Cg), an unstable DNA lesion that can break down further to form 5-OH-C, 5-OH-U, and 5,6-dihydroxy5,6-dihydrouracil (Ug) [62]. The three lesions, 5-OH-C, 5-OH-U, and Ug, have been identified in both untreated DNA and DNA that has been treated with an oxidizing agent [62,63].

Table 2 Entries are inferred as percentage of changes of nucleotides in CKD patients

To percentage numbers				
From	A	C	G	T
A	-	20	67	12
C	17	-	19	64
G	64	19	-	17
T	13	67	20	-

Table 3 Proportional distribution of nucleotide patterns in CKD patients. The nucleotide substitution presents unequal distribution

To number of times					Total
From	A	C	G	T	
A	-	22530	75550	14277	112357
C	17780	-	20520	68199	106499
G	68171	20576	-	17763	106510
T	14144	74738	22154	-	111036

Total	260095	117844	118224	100239	-

Relation between Oxidative Stress, Oxidative Deamination, Nucleotides Conversion, and Ti/Tv Ratio

From the start of this discussion, we demonstrated the intimate relationship between nucleotide interconversion and the mechanisms of transition and transversion, so to speak, the relative ratio of (Ts/Tv). We also reported the neighboring influence bias, on nucleotide substitutions. Oxidative stress is considered as an imbalance between pro- and antioxidant species, which results in molecular and cellular damage. Oxidative stress plays a crucial role in the development of age-related diseases. Oxidative damage to DNA nucleotide substitutions and these damages occur also in hemodialysis patients exacerbating any chance of proper cellular homeostasis at the podocyte level and developing so to speak a vicious circle [58].

CONCLUSION

We run whole exome sequencing in CKD patients with treated diabetes and hypertension and established a repertoire of SNPs, a profile nucleotides interconversion, Transition and Transversion ratio, and state of INDELS lengths. The data demonstrated a correlation profile between the data of Transition and Transversion vis a vis to nucleotides interconversion. The nucleotide conversion that replicates the effects of a transversion is expressed at lower levels of interconversion, but these levels are still enough to provoke radical amino acid changes to initiate the disease. The INDELS Length counts distribution favors the shorter length Indels with less than 3 bp than the larger length INDELS~107 bp. Lastly, the (Ts/Tv) ratio for all known variants for the 12 samples remained above those of unknown variants. Finally, this analysis of different aspects of mutations provides us with conceptual and mechanistic insights into different relations between nucleic acid interconversions and the evolution of transition and transversion ratio.

DECLARATIONS

Conflict of Interest

The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Ethics Approval

This study is part of a subgroup of an investigation started a while ago on hemodialysis patients from outpatient clinics of King Khaled Hospital and published with ethics approval issued to Dr. Alaraj .

Author Disclosure Statement

No competing financial interest exists

ACKNOWLEDGMENTS

REFERENCES

[1] Khorana, Har G., et al. "Polynucleotide synthesis and the genetic code." *Cold Spring Harbor symposia on quantitative biology*, Vol. 31, 1966.

[2] Nirenberg, Marshall, et al. "The RNA code and protein synthesis." *Cold Spring Harbor symposia on quantitative biology*, Vol. 31, 1966.

[3] Crick, Francis HC. "The origin of the genetic code." *Journal of molecular biology,* Vol. 38, No. 3, 1968, pp. 367-79.

[4] Duchêne, Sebastián, Simon YW Ho, and Edward C. Holmes. "Declining transition/transversion ratios through time reveal limitations to the accuracy of nucleotide substitution models." *BMC evolutionary biology,* Vol. 5, 2015, pp. 1-10.

[5] Gojobori, Takashi, Wen-Hsiung Li, and Dan Graur. "Patterns of nucleotide substitution in pseudogenes and functional genes." *Journal of molecular evolution,* Vol. 18, 1982, pp. 360-69.

[6] Kumar, Sudhir. "Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates." *Genetics,* Vol. 143, No. 1, 1996, pp. 537-48.

[7] Lynch, Michael. "Rate, molecular spectrum, and consequences of human mutation." *Proceedings of the National Academy of Sciences,* Vol. 107, No. 3, 2010, pp. 961-68.

[8] Lyons, Daniel M., and Adam S. Lauring. "Evidence for the selective basis of transition-to-transversion substitution bias in two RNA viruses." *Molecular biology and evolution,* Vol. 34, No. 12, 2017, pp. 3205-15.

[9] Petrov, Dmitri A., and Daniel L. Hartl. "Patterns of nucleotide substitution in Drosophila and mammalian genomes." *Proceedings of the National Academy of Sciences,* Vol. 96, No. 4, 1999, pp. 1475-79.

[10] Rosenberg, Michael S., Sankar Subramanian, and Sudhir Kumar. "Patterns of transitional mutation biases within and among mammalian genomes." *Molecular biology and evolution,* Vol. 20, No. 6, 2003, pp. 988-93.

[11] Wakeley, John. "The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance." *Trends in ecology & evolution,* Vol. 11, No. 4, 1996, pp. 158-62.

[12] Freeland, Stephen J., and Laurence D. Hurst. "The genetic code is one in a million." *Journal of molecular evolution,* Vol. 47, No. 3, 1998, p. 238.

[13] Woolfe, Adam, James C. Mullikin, and Laura Elnitski. "Genomic features defining exonic variants that modulate splicing." *Genome biology,* Vol. 11, 2010, pp. 1-23.

[14] Emond, Mary J., et al. "Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis." *Nature genetics,* Vol. 44, No. 8, 2012, pp. 886-89.

[15] Guo, Yan, et al. "The effect of strand bias in Illumina short-read sequencing data." *BMC genomics,* Vol. 13, 2012, pp. 1-11.

[16] Guo, Yan, et al. "Exome sequencing generates high quality data in non-target regions." *BMC genomics,* Vol. 13, No. 1, 2012, pp. 1-10.

[17] Guo, Yan, et al. "Three-stage quality control strategies for DNA re-sequencing data." *Briefings in bioinformatics,* Vol. 15, No. 6, 2014, pp. 879-89.

[18] Wang, Gao T., Bo Peng, and Suzanne M. Leal. "Variant association tools for quality control and analysis of large-scale sequence and genotyping array data." *The American Journal of Human Genetics,* Vol. 94, No. 5, 2014, pp. 770-83.

[19] Bainbridge, Matthew N., et al. "Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities." *Genome biology,* Vol. 12, No. 7, 2011, pp. 1-12.

[20] Yang, Ziheng, and Rasmus Nielsen. "Synonymous and nonsynonymous rate variation in nuclear genes of mammals." *Journal of molecular evolution,* Vol. 46, 1998, pp. 409-18.

[21] Welsgraber, K. H., et al. "Detection of normal Al in affected subjects and evidence for a cysteine for arginine substitution in the variant Al." *The Journal of Biological Chemistry,* Vol. 258, 1983, pp. 2508-13.

[22] Shah, Prediman K., et al. "High-Dose Recombinant Apolipoprotein A-IMilano Mobilizes Tissue Cholesterol and Rapidly Reduces Plaque Lipid and Macrophage Content in Apolipoprotein E–Deficient Mice: Potential Implications for Acute Plaque Stabilization." *Circulation*, Vol. 103, No. 25, 2001, pp. 3047-50.

[23] Galton, D. J., et al. "Identification of putative beneficial mutations for lipid transport." *Zeitschrift fur Gastroenterologie,* Vol. 34, 1996, pp. 56-58.

[24] Chen, Jun, et al. "Hunting for beneficial mutations: conditioning on SIFT scores when estimating the distribution of fitness effect of new mutations." *Genome Biology and Evolution,* Vol. 14, No. 1, 2022.

[25] Barbujani, Guido, and David B. Goldstein. "Africans and Asians abroad: genetic diversity in Europe." Annual Review of Genomics and Human Genetics, Vol. 5, 2004, pp. 119-50.

[26] Bell, John. "Predicting disease using genomics." *Nature,* Vol. 429, No. 6990, 2004, pp. 453-56.

[27] Goldstein, David B., and Gianpiero L. Cavalleri. "Understanding human diversity." *Nature,* Vol. 437, No. 7063, 2005, pp. 1241-42.

[28] Ng, Pauline C., and Steven Henikoff. "Accounting for human polymorphisms predicted to affect protein function." *Genome research,* Vol. 12, No. 3, 2002, pp. 436-46.

[29] Robert, Jacques, et al. "Predicting drug response and toxicity based on gene polymorphisms." *Critical reviews in oncology/hematology,* Vol. 54, No. 3, 2005, pp. 171-96.

[30] Stenson, Peter D., et al. "Human gene mutation database: 2003 update." *Human mutation,* Vol. 21, No. 6, 2003, pp. 577-81.

[31] Amberger, Joanna, et al. "McKusick's online Mendelian inheritance in man (OMIM®)." *Nucleic acids research,* Vol. 37, No. 1, 2009, pp. D793-D96.

[32] Schaefer, C., et al. "Disease-related mutations predicted to impact protein function. Disease-related mutations predicted to impact protein function." *BMC Genomics,* Vol. 13, 2012.

[33] Marchini, Jonathan, Peter Donnelly, and Lon R. Cardon. "Genome-wide strategies for detecting multiple loci that influence complex diseases." *Nature genetics,* Vol. 37, No. 4, 2005, pp. 413-17.

[34] Kraft, Peter, and David J. Hunter. "Genetic risk prediction—are we there yet?." *New England Journal of Medicine,* Vol. 360, No. 17, 2009, pp. 1701-03.

[35] Fazaludeen, Mohammad F., et al. "Chromosome HeatMap in CDK Patients as Defined by Multiregional Sequencing on Illumina MiSeq Platform." *European Journal of Medical and Health Sciences,* Vol. 2, No. 6, 2020.

[36] Alaraj, Moh'D., Nasim Alaraj, and Tarek D. Hussein. "Early Detection of Renal Impairment by Biomarkers Serum Cystatin C and Creatinine in Saudi Arabia." *Journal of Research in Medical and Dental Science,* Vol. 5, No. 1, 2017, pp. 37-45.

[37] Nachman, Michael W., and Susan L. Crowell. "Estimate of the mutation rate per nucleotide in humans." *Genetics,* Vol. 156, No. 1, 2000, pp. 297-304.

[38] Kondrashov, Alexey S. "Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases." *Human mutation,* Vol. 21, No. 1, 2003, pp. 12-27.

[39] Smith, Robert A., et al. "Cancer screening in the United States, 2017: a review of current American Cancer Society guidelines and current issues in cancer screening." *CA: a cancer journal for clinicians*, Vol. 67, No. 2, 2017, pp. 100-21.

[40] Zhang, Zhaolei, Paul Harrison, and Mark Gerstein. "Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome." *Genome research,* Vol. 12, No. 10, 2002, pp. 1466-82.

[41] Bulmer, Michael. "Neighboring base effects on substitution rates in pseudogenes." *Molecular biology and evolution,* Vol. 3, No. 4, 1986, pp. 322-29.

[42] Zhang, Zhaolei, and Mark Gerstein. "Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes." *Nucleic acids research,* Vol. 31, No. 18, 2003, pp. 5338-48.

[43] Wang, David G., et al. "Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome." *Science,* Vol. 280, No. 5366, 1998, pp. 1077-82.

[44] Cotton, R. G., and A. D. Auerbach. "Axton Metal (2008) The human variome project." *Science,* Vol. 322, 2008, pp. 861-62.

[45] Rosenberg, Michael S., Sankar Subramanian, and Sudhir Kumar. "Patterns of transitional mutation biases within and among mammalian genomes." *Molecular biology and evolution,* Vol. 20, No. 6, 2003, pp. 988-93.

[46] Wakeley, John. "The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance." *Trends in ecology & evolution,* Vol. 11, No. 4, 1996, pp. 158-62.

[47] Bainbridge, Matthew N., et al. "Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities." *Genome biology*, Vol. 12, No. 7, 2011, pp. 1-12.

[48] Denver, Dee R., et al. "Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of Caenorhabditis elegans." *Journal of molecular evolution,* Vol. 58, 2004, pp. 584-95.

[49] Modaresi, Atieh, Mohsen Nafar, and Zahra Sahraei. "Oxidative stress in chronic kidney disease." *Iranian journal of kidney diseases,* Vol. 9, No. 3, 2015, pp. 975-91.

[50] Dhanakoti, S. N., and H. H. Draper. "Response of urinary malondialdehyde to factors that stimulate lipid peroxidation in vivo." *Lipids,* Vol. 22, No. 9, 1987, pp. 643-46.

[51] Toborek, Michal, et al. "Effect of hemodialysis on lipid peroxidation and antioxidant system in patients with chronic renal failure." *Metabolism,* Vol. 41, No. 11, 1992, pp.1229-32.

[52] Zima, Tomás, et al. "Antioxidant enzymes-superoxide dismutase and glutathione peroxidase-in haemodialyzed patients." *Blood purification,* Vol. 14, No. 3, 1996, pp. 257-61.

[53] Maynard, Scott, et al. "Base excision repair of oxidative DNA damage and association with cancer and aging." *Carcinogenesis,* Vol. 30, No. 1, 2009, pp. 2-10.

[54] Ames, B. N., M. K. Shigenaga, and TM OXIDANTS HAGEN. "antioxidants." *and the degenerative diseases of aging. Proceedings of the National Academy of Sciences of the United States of America,* Vol. 90, 1993, ppp.7915-22.

[55] Schaaper, Roel M., Bryan N. Danforth, and Barry W. Glickman. "Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the Escherichia coli lacI gene." *Journal of molecular biology,* Vol. 189, No. 2, 1986, pp. 273-84.

[56] Schaaper, Roe1 M., and Ronnie L. Dunn. "Spontaneous mutation in the Escherichia coli lacI gene." *Genetics,* Vol. 129, No. 2, 1991, pp. 317-26.

[57] Tkeshelashvili, L. K., et al. "Mutation spectrum of copper-induced DNA damage." *Journal of Biological Chemistry,* Vol. 266, No. 10, 1991, pp. 6401-06.

[58] Moraes, E. C., S. M. Keyse, and R. M. Tyrrell. "Mutagenesis by hydrogen peroxide treatment of mammalian cells: a molecular analysis." *Carcinogenesis,* Vol. 11, No. 2, 1990, pp. 283-93.

[59] Wood, Michael L., et al. "Mechanistic studies of ionizing radiation and oxidative mutagenesis: genetic effects of a single 8-hydroxyguanine (7-hydro-8-oxoguanine) residue inserted at a unique site in a viral genome." *Biochemistry*, Vol. 29, No. 30, 1990, pp. 7024-32.

[60] Shibutani, Shinya, Masaru Takeshita, and Arthur P. Grollman. "Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG." *Nature,* Vol. 349, No. 6308, 1991, pp. 431-34.

[61] Hashimoto, Takashi, et al. "Hyoscyamine 6β-hydroxylase, an enzyme involved in tropane alkaloid biosynthesis, is localized at the pericycle of the root." *Journal of Biological Chemistry,* Vol. 266, No. 7, 1991, pp. 4648-53.

[62] Dizdaroglu, M., et al. "Formation of cytosine glycol and 5, 6-dihydroxycytosine in deoxyribonucleic acid on treatment with osmium tetroxide." *Biochemical Journal,* Vol. 235, No. 2, 1986, pp. 531-36.

[63] Wagner, J. Richard, Chia-Chieh Hu, and Bruce N. Ames. "Endogenous oxidative damage of deoxycytidine in DNA." *Proceedings of the National Academy of Sciences,* Vol. 89, No. 8, 1992, pp. 3380-84.