Journal of Cancer Science & Therapy
ISSN: 1948-5956

Chu and Wei, J Cancer Sci Ther 2019, 11:6

Research Article Open Access

Synonymous Mutations that Increase the tRNA Adaptation Index (tAI) are Generally Suppressed in Human Oncogenes

Duan Chu and Lai Wei*

College of Life Sciences, Beijing Normal University, No. 19 Xinjiekouwai Street, Haidian District, Beijing, P.R. China

Abstract

Background: Different synonymous codons are decoded at distinct rates during translation elongation due to the difference in tRNA availability, term tRNA adaptation index (tAI). Codons with higher tAI values are translated faster. Synonymous mutations do not change the amino acid (AA) but do sometimes change the tAI. That means synonymous mutations are able to alter the translation efficiency of host genes. In the cancer field, synonymous mutations are often automatically ignored. However, the increased translation of oncogenes or the decreased translation of tumor suppressor genes (TSG) might also lead to oncogenesis even when the protein sequences are unchanged. These changes in translation level could be induced by synonymous mutations.

Results and Discussion: We downloaded the single nucleotide polymorphisms (SNPs) in human populations. The ancestral state is parsed according to genomic alignments between human, macaque and mouse. We found that in the normal human populations, derived synonymous mutations that increase tAI are strongly suppressed in oncogenes but slightly favored in TSG. In oncogenes, mutation sites with higher conservation levels are more likely to decrease the tAI.

Conclusion: Our results indicate that the synonymous mutations in the human genome are not strictly neutral. The potentially increased translation of oncogenes and the decreased translation of TSG caused by synonymous mutations are suppressed in normal human populations. This is an indirect evidence that the synonymous-induced translational changes might be related to oncogenesis and should not be ignored in the cancer studies.

Keywords: Synonymous mutations; Human population; tRNA adaptation index (tAI); Oncogenes; Tumor suppressor genes (TSG)

Introduction

Synonymous mutations are those mutations in coding sequence (CDS) that do not change the amino acid (AA). However, this does not mean that synonymous mutations are free from natural selection [1]. Apart from the small fraction of synonymous mutations that affect mRNA splicing [2], another important feature for synonymous codons is that different synonymous codons are decoded at different rates during translation elongation [3-6] due to the difference in tRNA availability, term tRNA adaptation index (tAI) [7]. Optimized codons (with higher tAI values) tend to have higher translation efficiency (faster elongation speed). Thus, synonymous mutations that change the tAI would result in altered translation rate, which could be potentially subjected to natural selection. In the cancer field, synonymous mutations are often automatically ignored because much attention has been paid to the AA changes that are associated with cancer. However, the increased translation of oncogenes or the decreased translation of tumor suppressor genes (TSG) might also lead to oncogenesis even when the protein sequences are unchanged. These changes in translation level could be induced by synonymous mutations. So that we intuitively consider that those synonymous mutations increasing the tAI should be suppressed in oncogenes in normal human populations. In our previously published work, we reported that "nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selection than expectation" [1]. In this current follow-up study, we further divided the well-annotated cancer-related genes into oncogenes and tumor suppressor genes (TSG) and we mainly focus on synonymous mutations [8].

We downloaded the single nucleotide polymorphisms (SNPs) in human populations and extracted those mutations in highly expressed genes in HeLa cells. The ancestral state of mutations is parsed according to genomic alignments between human, macaque and mouse. We found that in the normal human populations, derived synonymous mutations that increase tAI are strongly suppressed in oncogenes but slightly favored in TSG. In oncogenes, mutation sites with higher conservation levels are more likely to decrease the tAI.

Our results indicate that the synonymous mutations in the human genome are not strictly neutral. The potentially increased translation of oncogenes and the decreased translation of TSG caused by synonymous mutations are suppressed in normal human populations. This is an indirect evidence that the synonymous-induced translational changes might be related to oncogenesis and should not be ignored in the cancer studies.

Methods

Data collection

We collected the recent version of all human SNP data from NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) (Version: Build 150. Last downloaded: January 2018). The common and all SNPs are labeled distinctly in the website. The list of 719 human cancer-related genes was downloaded from the latest version of cancer gene census website (CGC, https://cancer.sanger.ac.uk/census/). After removing a few ambiguously annotated genes, we finally obtained 312 oncogenes and

***Corresponding author:** Lai Wei, College of Life Sciences, Beijing Normal University, No. 19 Xinjiekouwai Street, Haidian District, Beijing, P.R. China, Tel: +86-10-58807647; E-mail: weilai_bnu@163.com

Received May 06, 2019; **Accepted** June 04, 2019; **Published** June 11, 2019

Citation: Chu D, Wei L (2019) Synonymous Mutations that Increase the tRNA Adaptation Index (tAI) are Generally Suppressed in Human Oncogenes. J Cancer Sci Ther 11: 208-212.

Copyright: © 2019 Chu D, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

314 tumor suppressor genes (TSG) (Additional file 1: Table S1). The reference genome sequences of human (*H. sapiens*, version hg19) and mouse (*M. musculus*, version mm10) were downloaded from UCSC Genome Browser (genome.ucsc.edu), and the reference genome of rhesus macaque (*M mulatta*, Ensembl version v89) was downloaded from Ensembl Genome Browser (www.ensembl.org).

Annotation of human SNPs

We annotated the SNP sites using the hg19 human genome downloaded from UCSC Genome Browser (genome.ucsc.edu). If a SNP hits multiple isoforms of the same gene, the transcript with the longest CDS (canonical transcript) was retained. The canonical transcript of each gene was defined by the software SnpEff (version 4.2) [9]. If a SNP does not hit any genes, it is annotated as intergenic. All the information of a given SNP in CDS including the position on CDS, the amino acid before and after mutation, were inferred from the output file of SnpEff. The mutations related to splicing effects (indicated by the software) were discarded because we should separate the selection on synonymous mutations from the constrain on splicing changes.

Conservation analysis

Conservation level of genomic positions is measured by phyloP score (file hg19.100way.phyloP100way.bw, downloaded from UCSC Genome Browser, genome.ucsc.edu). Briefly, sites with higher conservation level have higher phyloP scores. The orthologous sites (genomic coordinates) between human and mouse or between human and rhesus macaque were transferred with liftOver [8] based on the pairwise genomic alignments. The lift Over chain files were downloaded from UCSC Genome Browser website(http://hgdownload.soe.ucsc.edu/goldenPath/hg19/liftOver/). Bedtools (version 2.25) [10] was used to extract sequences of a give region according to the reference genome.

Calculation of gene expression level in human HeLa cells

We searched for the public database and chose an NGS dataset conducted in human HeLa cells (GES63591) [11]. The mRNA-Seq library (si-control) was used to define the gene expression level. We aligned the mRNA-Seq NGS reads to the hg19 reference genome using STAR (version 2.7) [12]. The uniquely mapped reads were kept for downstream analysis. The read counts of each gene were calculated by htseq-count (version 0.5.4) [13]. In this gene expression calculation, the canonical transcript of each gene was chosen, and all the reads overlapped with exon regions were counted. Highly expressed genes (7712 genes in total) are defined as gene with reads count > 200. Among the 7712 highly expressed genes in HeLa cells, 104 are oncogenes and 157 are TSG, the remaining 7451 genes are termed "other genes". We focused on highly expressed genes because they are better annotated (since poorly annotated genes are not possible to have many mapped NGS reads).

History and definition of tAI

To quantitatively measure the extent of codon bias, the codon adaptation index (CAI) was first invented [14]. CAI considers the codon frequencies appearing in the highly expressed genes. The factors that are ignored by CAI are 1) the tRNA pool that could actually translate each codon and 2) the wobble interactions between codon and anticodon. These factors could indeed affect the decoding efficiency. For example, the adenosines located at the first anticodon positions (A34NN) are deaminated into inosines (I34NN), which could form wobble (and also weaker) interactions with C, U or A [15]. To fill the gap left by CAI, tRNA adaptation index (tAI) was then created [7]. tAI could either describe a codon or a gene. The codon level tAI was name wi in the original literature [7]. tAI of a codon is calculated by the following steps, 1) seek for the corresponding tRNA copy numbers of a codon; 2) weight each tRNA copy number if wobble base pairing appears; 3) sum up the weighted tRNA copy numbers; 4) normalize by the maximum value within the same amino acid (within synonymous codons). The weight of each type of wobble interaction is decided by the selection constraint (*sij*) [7]. Distinct *sij* values suitable for different evolutionary clades were also defined [16]. tAI of each codon is generally in proportion to the tRNA that can translate this codon. This could be an estimate of translation efficiency at codon level. The tAI of each gene is the geometric mean of the tAI values of each codon. Similarly, tAI of a gene could be an estimate of gene level translation efficiency.

Statistical analysis and code availability

All statistical analyses (correlation tests, Fisher's exact tests) and the graphic work were conducted in R environment (http://www.R-project.org/). The codes used in this study are available under request.

Results

Mutations in coding regions of human genome

Our previous work has displayed the mutation patterns in human cancer-related genes [1] and our recent work is basically an extension of our previous study. The workflow for the preliminary data processing is summarized as follows (see Methods for details). We collected the recent version of human SNP data (normal human population) from NCBI dbSNP and annotated the SNP sites using the hg19 human genome downloaded from UCSC Genome Browser (Methods). To determine whether a mutation in human population is a derived mutation, the orthologous sites between human and mouse or between human and rhesus macaque were transferred with liftOver based on the pairwise alignments. Derived mutations must have the same nucleotide in the reference genome of human, macaque and mouse (Figure 1A). We retrieved the derived mutations in CDS of human genes. The mutation sites potentially related to splicing changes were further discarded (Methods). In the remaining set of mutations in CDS, synonymous mutations are those do not change amino acid (AA) and nonsynonymous mutations are those that change AA sequences.

Oncogenes and tumor suppressor genes (TSG)

From the cancer gene consensus, we obtained 626 well annotated cancer-related genes, including 312 oncogenes and 314 TSG (Methods and Additional file 1: Table S1). Initially, we define the human genes apart from these 626 genes as "other genes". However it is conceivable that the cancer-related genes are better annotated than the other genes, so that the annotation of mutation sites might be less precise in some poorly characterized genes. To conduct a fair comparison, we confined our analyses in the set of highly expressed genes in HeLa cells (Methods). Among the 7712 highly expressed genes in HeLa cells, 104 are oncogenes and 157 are TSG, the remaining 7451 genes are termed "other genes". We summarized the mutations in coding region of these genes: 408 nonsynonymous and 340 synonymous mutations in oncogenes (nsy/syn = 1.20), 861 nonsynonymous and 603 synonymous mutations in TSG (nsy/syn = 1.43), 34845 nonsynonymous and 20800 synonymous mutations in other genes (nsy/syn = 1.68). This result could be interpreted as the constraint on cancer-related genes especially oncogenes in normal human populations due to the deleterious effects of nonsynonymous mutations (Figure 1B).

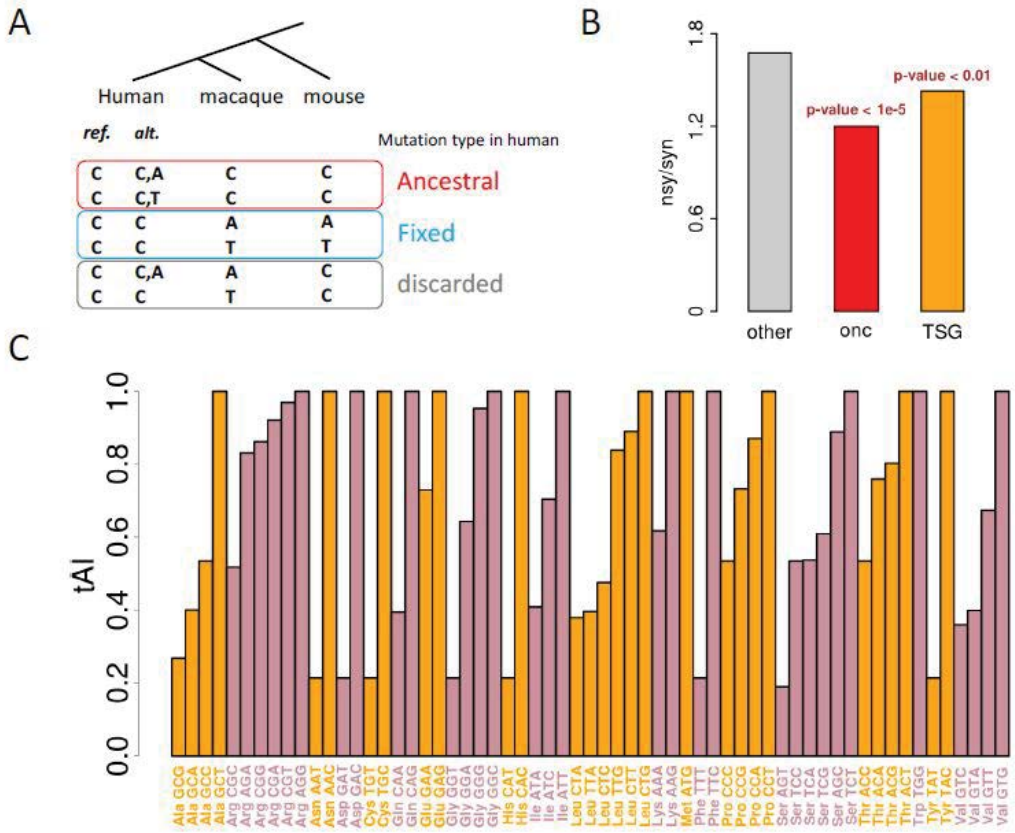

Figure 1: An overview of the methods and materials used in this study **(A)** Our scheme to define the ancestral state of a site and the derived mutations. The branch length in the phylogenetic tree is unscaled **(B)** nonsynonymous to synonymous ratios in oncogenes, TSG and other genes. P-values are calculated by Fisher's exact tests **(C)** tAI of the sense codons in human.

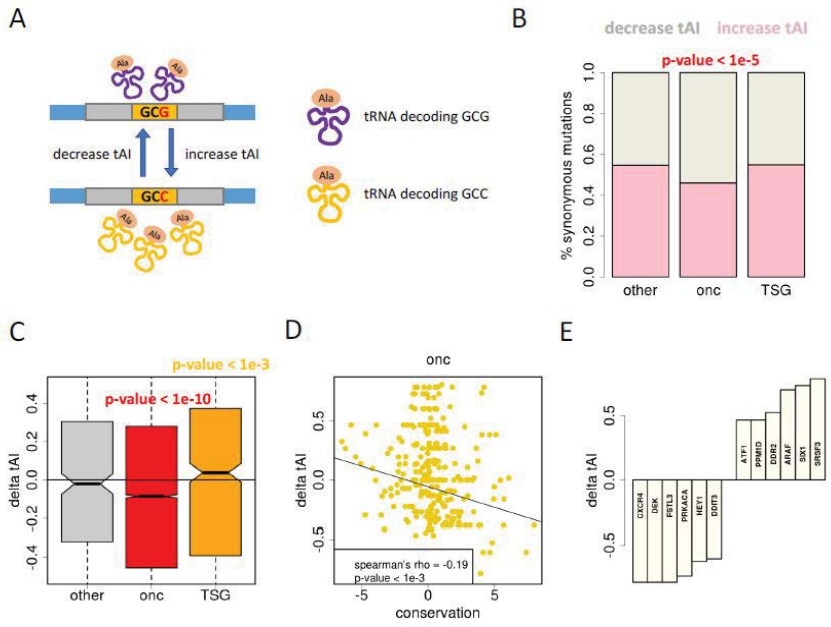

Figure 2: Constraint on synonymous mutations that increase or decrease the tAI values **(A)** A diagram telling the readers our definition of synonymous mutations that increase or decrease the tAI **(B)** Fractions of the synonymous mutations that increase or decrease the tAI values. P-values are calculated by Fisher's exact tests **(C)** Boxplots showing the delta tAI values of synonymous mutations in oncogenes, TSG and other genes. P-values are calculated by Wilcoxon rank sum tests **(D)** Spearman's correlation between the conservation level and delta tAI values among the synonymous mutations in oncogenes **(E)** Example of oncogenes with the lowest or highest delta tAI values.

Synonymous mutations that increase tAI are suppressed in oncogenes

tRNA adaptation index (tAI) describes the tRNA accessibility of a given codon (Figure 1C and Methods). Codons with higher tAI values are generally translated faster. Thus, synonymous mutations are able to alter the translation rate through changing the tAI values although the AA is unchanged (Figure 2A). Since the increased translation of oncogenes or the decreased translation of TSG might also lead to oncogenesis, we intuitively surmise that the synonymous mutations increasing the tAI should be suppressed in oncogenes in normal human populations. Among all the derived synonymous mutations in oncogenes, TSG and other genes, we classified them into two categories: tAI-up mutations (increased tAI) and tAI-down mutations (decreased tAI). We found that the fraction of tAI-up synonymous mutations is significantly lower in oncogenes than in other genes (Figure 2B) while no remarkable difference is observed between TSG and other genes (Fig. 2B). We next calculated the change of tAI value (delta tAI) for each synonymous mutation. Globally, we observed that the delta tAI values are significantly lower in oncogenes than other genes (Figure 2C) but significantly higher in TSG than other genes (Figure 2C). Our observations are plausible, the increased translation (contributed by tAI-up mutations) of oncogenes or the decreased translation (caused by tAI-down mutations) of TSG might be related to cancer growth or oncogenesis, so that these mutations are not likely to occur in the data of normal human populations.

We also found additional evidence to support our assumption. From the aspect of evolutionary biology, the more conserved genes or sites are usually functionally more important and selective constrained. In oncogenes, we investigated the correlation between the conservation level of the synonymous mutation sites (Methods) and the delta tAI values of these sites (Figure 2D). Interestingly, these two features show a strong negative correlation (Figure 2D). Remember that these patterns come from the normal human population data rather than cancer samples, so this correlation simply indicates that more conserved sites in oncogenes are less affordable for a (synonymous) mutation that increase the tAI, which could potentially enhance the translation of host oncogenes.

Discussion

Synonymous mutations were originally thought to be evolutionarily neutral since they do not change the amino acid sequences. With the appearance of next generation sequencing (NGS) technique, studies have identified a small set of synonymous mutations that might be subjected to natural selection due to their effect on mRNA splicing [2]. Our current study further broadened the knowledge of selection force acting on synonymous mutations.

We revealed that in the normal human population, the synonymous mutations that increase tAI are suppressed in oncogenes. Since increased translation (contributed by tAI-up mutations) of oncogenes or the decreased translation (caused by tAI-down mutations) of TSG might be related to cancer growth or oncogenesis, so that these mutations are not likely to occur in the data of normal human populations. To concretize our theory, we ranked the oncogenes (with synonymous mutations) by the mean delta tAI values and the top six and bottom six genes were displayed (Figure 2E). For example, oncogene *CXCR4* (C-X-C motif chemokine receptor 4) has the lowest delta tAI value, oncogene *SRSF3* (serine/argine-rich splicing factor 3) has the highest delta tAI value (Figure 2E). Note again that these observations come from the normal human population data rather than cancer samples, so the oncogenes with lower delta tAI values are less affordable for a synonymous mutation that increase the tAI, and vice versa. If one asks for a set of candidate oncogenes which are most likely to cause oncogenesis by synonymous mutations, we would putatively recommend the genes with lower delta tAI values in normal human populations like *CXCR4*.

Furthermore, it is almost the "tradition" that the synonymous mutations are automatically ignored in the cancer prevention or cancer diagnosis. On the contrary, our work demonstrates that in human populations, synonymous mutations that alter the tAI values are selectively constrained in oncogenes as well as TSG. This means that this kind of synonymous mutations might otherwise lead to cancer if they occur in oncogenes as commonly as they occur in other genes. Thus, the synonymous mutations should not be ignored in the cancer field. We speculate that some synonymous mutations in oncogenes (e.g. which severely increase tAI) should be noticed in cancer diagnosis and even be listed as potential causal mutations. Our idea combined with the patterns we found in oncogenes and TSG should be interesting to molecular biologists, cancer biologists as well as evolutionary biologists. Together with the fact that genomic studies on features like tAI are rarely conducted in the cancer researches, so that our work should be welcome by the broad community of this field.

Conclusion

Our results indicate that the synonymous mutations in the human genome are not strictly neutral. The potentially increased translation of oncogenes and the decreased translation of TSG caused by synonymous mutations are suppressed in normal human populations. This is an indirect evidence that the synonymous-induced translational changes might be related to oncogenesis and should not be ignored in the cancer studies.

Declarations

Ethics approval and consent to participate

All datasets used in this study were downloaded from publically available websites as described in the Methods section.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the NCBI website.

Competing interests

The authors declare they have no competing interests.

Funding

This research was financially supported by the National Natural Science Foundation of China (Grant no. 31770213). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

LW designed and supervised this research. Both DC and LW analyzed the data. DC and LW wrote this article.

Acknowledgements

We thank all members in Wei Lab for their constructive suggestions to this project.

References

1. Chu D, Wei L (2019) Nonsynonymous, synonymous and nonsense mutations in human cancer related genes undergo stronger purifying selections than expectation. BMC Cancer 19: 359.
2. Supek F, Minana B, Valcarcel J, Gabaldon T, Lehner B (2014) Synonymous mutations frequently act as driver mutations in human cancers. Cell 156: 1324-1335.
3. Comeron JM (2004) Selective and mutational patterns associated with gene expression in humans: Influences on synonymous composition and intron presence. Genetics 167: 1293-1304.
4. Dana A, Tuller T (2014) The effect of tRNA levels on decoding times of mRNA codons. Nucleic Acids Res 42: 9171-9181.
5. Sorensen MA, Kurland CG, Pedersen S (1989) Codon usage determines translation rate in *Escherichia coli*. J Mol Biol 207: 365-377.
6. Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, et al. (2015) Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. Mol Cell 59: 744-754.
7. Dos Reis M, Savva R, Wernisch L (2004) Solving the riddle of codon usage preferences: A test for translational selection. Nucleic Acids Res 32: 5036-5044.
8. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, et al. (2006) The UCSC genome browser database: Update 2006. Nucleic Acids Res 34: D590-D598.
9. Cingolani P, Platts A, Wang L, Coon M, Nguyen T, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; Iso-2; Iso-3. Fly (Austin) 6: 80-92.
10. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.
11. Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, et al. (2015) N(6)-methyladenosine modulates messenger RNA translation efficiency. Cell 161: 1388-1399.
12. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29: 15-21.
13. Anders S, Pyl PT, Huber W (2015) HTSeq--A Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166-169.
14. Sharp PM, Li WH (1987) The codon adaptation index - A measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15: 1281-1295.
15. Novoa EM, Pavon-Eternod M, Pan T, De Pouplana L (2012) A role for tRNA modifications in genome structure and codon usage. Cell 149: 202-213.
16. Sabi R, Tuller T (2014) Modelling the efficiency of codon-tRNA interactions based on codon usage bias. DNA Res 21: 511-525.