Journal of Genetics and Genomics 50 (2023) 447–449

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Letter to the editor

Synonymous somatic mutations that alter proximal out-of-frame downstream ATGs are associated with aberrant gene expression levels in cancer cells

Synonymous mutations are mutations that occur in the coding sequence of a gene but do not change the amino acid sequence of the encoded protein. It is now understood that synonymous mutations can have a variety of functional consequences by altering the expression of genetic information and are often non-neutral (Shen et al., 2022). For example, synonymous mutations can regulate the binding of transcription factors (Stergachis et al., 2013), the stability of messenger RNA (mRNA) molecules (Presnyak et al., 2015; Chen et al., 2017), the translational efficiency and accuracy (Qian et al., 2012; Sun and Zhang, 2022), and the co-translational protein folding (Zhang et al., 2009; Buhr et al., 2016). In this study, using genomic and transcriptomic data collected from thousands of cancer samples, we investigated if synonymous mutations could cause aberrant gene expression in cancer cells via translation initiation-mediated mechanisms.

In eukaryotes, recognition of the translation initiation site is based on a "scanning" mechanism using the 43S preinitiation complex (PIC), which is comprised of a 40S ribosomal subunit, several eukaryotic initiation factors (eIFs), a methionyl initiator transfer RNA (Met-tRNAi), and a guanosine triphosphate (Hinnebusch, 2014). The PIC is recruited by the mRNA 5′-cap and migrates along the mRNA one nucleotide (nt) at a time, searching for the AUG codon. Consistent with this model, insertion of an upstream AUG codon (uAUG) reduces translation initiation at the annotated AUG codon (aAUG) of a reporter gene because the uAUG can retain the PICs that otherwise would initiate translation at the aAUG (Kozak, 2002). Furthermore, translation initiation at uAUGs often causes premature translation termination (especially when they are in a different reading frame from the aAUG) and therefore, can activate the nonsense-mediated mRNA decay (NMD) pathway to degrade mRNA (Barbosa et al., 2013), reinforcing the inhibitory effect of uAUGs at the translational level.

Notably, in a recent study, we reported that not only uAUGs but also proximal downstream AUGs (dAUGs) could inhibit translation at the aAUG (Li et al., 2022). For example, an out-of-frame dAUG created with its adenine at the +6 position (relative to the A[+1] in the aAUG codon of the reporter gene) can reduce the protein production rate by ~50% in the budding yeast. The inhibitory effect of proximal out-of-frame dAUGs diminishes over increasing aAUG-dAUG distance, undetectable beyond ~17 nt. Furthermore, when a dAUG codon resides in a stronger sequence context (i.e., with adenine or guanine at the −3 position relative to the A[+1] in the dAUG), the proximal out-of-frame dAUGs could cause more severe inhibition on the aAUG translation initiation. The underlying mechanism of the distance- and context-dependent translational inhibition by proximal dAUGs is that PIC uses small-amplitude 5′−3′ and 3′−5′ oscillations (i.e., bidirectional scanning) to search for the AUG codon, a process that leads to competition for translation initiation between two closely spaced AUGs (Li et al., 2022).

Since human cancer cells often accumulate numerous somatic mutations, some point mutations in these cells could introduce or remove proximal out-of-frame dATGs, exerting influence on translation initiation and mRNA degradation. Therefore, in this study, we hypothesized that dATG-altering somatic mutations, even if synonymous, would affect gene expression in cancer cells and potentially contribute to cancer development. To test this hypothesis, we retrieved the point mutations identified in 9482 cancer samples spanning 33 cancer types from The Cancer Genome Atlas (TCGA, accessed on July 25th, 2022) (Cancer Genome Atlas Research Network et al., 2013). We focused on the 486,382 synonymous mutations to exclude potential interference from altered amino acid sequences and identified mutations that introduced out-of-frame dATGs (Fig. 1A).

Aberrant translation initiation at out-of-frame dAUGs would result in the degradation of mRNA via an NMD mechanism (Li et al., 2022), offering us an opportunity to infer the extent of the inhibitory effect caused by out-of-frame dAUGs from the reduction in the mRNA level. To determine the expression impacts of synonymous mutations that introduced dATGs, we retrieved transcriptome data from TCGA for the same set of cancer samples and estimated normalized mRNA levels by calculating the ratio of mRNA levels of genes bearing a synonymous dATG-introducing mutation to the median mRNA level of that gene in all other samples of the same cancer type (Fig. 1A). If a synonymous mutation has no effect on the mRNA level, we expect this expression ratio will be 1, while somatic mutations that increase mRNA levels will result in a ratio > 1, and somatic mutations that decrease mRNA levels will result in a ratio < 1. As a positive control, we identified 104,565 somatic mutations that introduced premature termination codons (PTC) in 9482 cancer samples from TCGA, and determined that this classical scenario for NMD activation reduced the mRNA level by 16% ($P < 2.2 \times 10^{-16}$, Mann-Whitney U test; Fig. 1B).

We then grouped the synonymous mutations according to their positions relative to the aATG: one group included synonymous mutations within 17 nt of the aATG (referred to as "proximal") and two additional groups included synonymous mutations in further downstream regions (18 nt−34 nt and 35 nt−51 nt, "distal"). Consistent with the reported inhibitory effect of proximal out-of-frame dAUGs (Li et al., 2022), we found that the introduction of proximal out-of-

https://doi.org/10.1016/j.jgg.2023.02.013

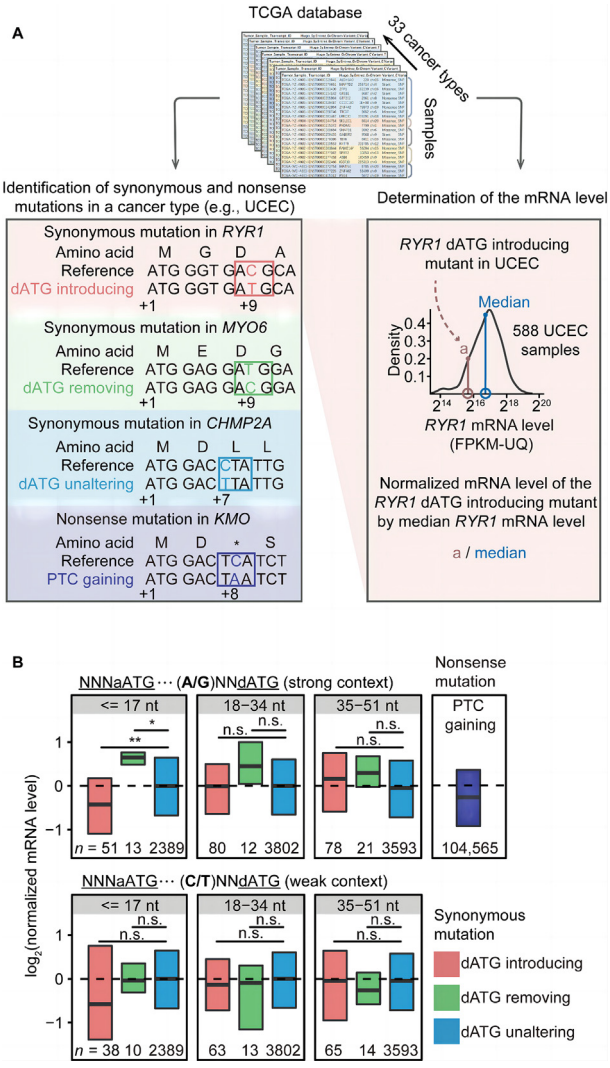

Fig. 1. Changes in the mRNA level in cancer cells upon synonymous somatic mutations that alter proximal out-of-frame dATGs. **A**: Schematic shows the identification of synonymous mutations that introduce or remove out-of-frame dATGs in the cancer genome and determination of the mRNA level of the mutated gene in the corresponding cancer sample. Somatic mutations were detected based on the MuTect2 pipeline, and dATG-altering mutations were identified based on the human reference genome GRCh38. Using the example of UCEC, some dATG-altering synonymous mutations are illustrated. The synonymous mutations that neither introduce nor remove dATGs ("dATG unaltering") and the nonsense mutations that introduced premature termination codons were also identified. The RNA-seq data were obtained from TCGA in the unit of FPKM-UQ. The mRNA level of genes bearing synonymous mutation was normalized by the median mRNA levels of that gene in all other samples of the same cancer type. **B**: Boxplots show the mRNA levels of genes bearing synonymous mutations in the cancer samples, grouped by the genomic region where the synonymous mutation occurred (relative to the A[+1] in the aATG, labeled in the top box) and by the sequence context (i.e., the nucleotide at the −3 position based on the human reference genome GRCh38) of the introduced or removed out-of-frame dATGs. Nonsense mutations are also shown; the effect size of reduction in the mRNA level caused by these nonsense mutations appeared small likely due to their low allele frequencies within a cancer sample. Outliers and whiskers are not shown. *P* values were given by the one-tail Mann-Whitney *U* tests. **, $P < 0.01$; *, $P < 0.05$; n.s., not significant. UCEC, uterine corpus endometrial carcinoma; FPKM_UQ, Fragments Per Kilobase of transcript per Million mapped reads Upper Quartile.

frame dATGs residing in a strong sequence context significantly reduced the mRNA level, compared to synonymous mutations that neither introduced nor removed out-of-frame dATGs ("dATG unaltering", $P = 0.004$, Mann-Whitney *U* test; Fig. 1B), while that of distal out-of-frame dATGs did not ($P = 0.41$ and 0.93, Mann-Whitney *U* tests; Fig. 1B). The same conclusion can be reached by estimating the Cliff's Delta effect sizes of these dATG-introducing synonymous mutations and examining their 95% confidence intervals (Table S1). These proximal dATG-introducing mutations reduced mRNA levels by 26%, qualitatively comparable to the reduction in the mRNA level incurred by somatic gains of PTCs (16%). In contrast, the reduction in the mRNA level was not statistically significant ($P = 0.06$, Mann-Whitney *U* test; Fig. 1B; Table S1) in cancer cells with synonymous somatic mutations that introduced proximal out-of-frame dATGs residing in a weak sequence context (i.e., with cytidine or thymine at the −3 position), consistent with the previous finding that the inhibitory effects on mRNA levels by proximal out-of-frame dATGs were dependent on translation activity (Li et al., 2022).

We further hypothesized that removing proximal out-of-frame dATGs from the coding sequence by synonymous somatic mutations could elevate the mRNA level due to the elimination of translation initiation from the original dAUG and the consequent diminished NMD activity. To test this hypothesis, we identified synonymous mutations that removed out-of-frame dATGs in TCGA samples (Fig. 1A). Indeed, synonymous mutations that removed proximal dATG significantly elevated mRNA levels when the original dATG resided in a strong sequence context (increasing by 57%, $P = 0.02$, Mann-Whitney *U* test; Fig. 1B; Table S1) but not in a weak sequence context ($P = 0.67$, Mann-Whitney *U* test; Fig. 1B; Table S1). By contrast, the synonymous somatic mutations that removed distal out-of-frame dATGs exhibited no statistically significant impact on the gene expression level of cancer cells, regardless of the sequence context (Fig. 1B; Table S1).

Some genes previously reported being associated with cancer development indeed bore dATG-altering mutations in TCGA samples. For example, a C>T mutation in *PTPN13* introduced an out-of-frame dATG at +5 position in a uterine corpus endometrial carcinoma sample, associated with a 3-fold reduction in its mRNA level. *PTPN13* encodes a protein tyrosine phosphatase and participates in cancer development as a tumor suppressor in the epithelial cells of multiple cancer types (Colbert et al., 2015). In a second example, a T>C mutation in *TACC2* removed an out-of-frame dATG at +8 position in a stomach adenocarcinoma sample, associated with a 2.6-fold increase in its mRNA level. *TACC2* acts as a potent oncogene in gastric cancer (Lasorella et al., 2017), and overexpression of *TACC2* has been reported to be associated with cancer recurrence and poor prognosis (Shakya et al., 2018). Collectively, these observations suggest that synonymous mutations that alter proximal out-of-frame dATGs are prevalent in cancer cells and potentially have profound impacts on cancer development.

Upon evaluating the possible effects of these two dATG-altering mutations on the development of cancer, we then aimed to conduct statistical analyses to test whether synonymous mutations which introduced or removed proximal out-of-frame dATGs as a group could potentially play a role in cancer development. To this end, we estimated the cancer effects for individual mutations using an established computational strategy (Cannataro et al., 2018; Mandell et al., 2023) for synonymous mutations located in various positions relative to the aATG (Fig. S1). We found that within 17 nt of the aATG, dATG-removing mutations exhibited greater cancer effects than dATG-unaltering mutations (increasing by a factor of 11, $P = 0.03$, Mann-Whitney *U* test; Fig. S1), but dATG-introducing mutations did not exhibit significantly different cancer effects from the dATG-unaltering mutations. As a negative control, in the further downstream regions (18 nt–34 nt downstream of the aATG), dATG-removing synonymous mutations no longer exhibited greater cancer effects than dATG-unaltering mutation (Fig. S1). Based on these observations, it appears that while both dATG-removing and dATG-introducing synonymous mutations have the potential to affect gene expression in cancer cells, the available empirical data suggest

that dATG-removing mutations in particular play a significant role in promoting cancer development, likely through increasing the expression levels of the genes they affect.

It is worth noting that in addition to the cap-dependent translation initiation mechanism which we have focused on in this study, a small fraction of eukaryotic mRNA molecules possesses internal ribosomal entry sites (IRES), which enable a cap-independent translation mechanism under some stress conditions (Yang and Wang, 2019). In principle, synonymous mutations that create or disrupt IRES could also exert an influence on gene expression in cancer cells. The role of such mutations in cancer development deserves further investigation.

To summarize, we reported that synonymous mutations, which for long were thought to have little functional consequence due to the amino acid sequence being unaltered, could lead to aberrant gene expression in cancer cells in a sequence context- and distance-dependent manner. The effects of synonymous mutations introducing or removing proximal out-of-frame dATGs likely come from competition for translation initiation between closely spaced AUG triplets. This mechanism thus provides a novel framework to predict functional important synonymous mutations and their effects on the expression of genetic information. Together with recent studies reporting that synonymous mutations are often strongly non-neutral (Shen et al., 2022), our study implicates the importance of carefully considering the potential effects of synonymous mutations on gene expression in human diseases.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by grants from the National Key Research and Development Program of China (2019YFA0508700) and the National Natural Science Foundation of China (32100443 and 31922014).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2023.02.013.

References

Barbosa, C., Peixeiro, I., Romao, L., 2013. Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 9, e1003529.
Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M.V., Komar, A.A., 2016. Synonymous codons direct cotranslational folding toward different protein conformations. Mol. Cell 61, 341–351.
Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., 2013. The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. 45, 1113–1120.
Cannataro, V.L., Gaffney, S.G., Townsend, J.P., 2018. Effect sizes of somatic mutations in cancer. J. Natl. Cancer Inst. 110, 1171–1177.
Chen, S., Li, K., Cao, W., Wang, J., Zhao, T., Huan, Q., Yang, Y.F., Wu, S., Qian, W., 2017. Codon-resolution analysis reveals a direct and context-dependent impact of individual synonymous mutations on mRNA level. Mol. Biol. Evol. 34, 2944–2958.
Colbert, P.L., Vermeer, D.W., Wieking, B.G., Lee, J.H., Vermeer, P.D., 2015. EphrinB1: novel microtubule associated protein whose expression affects taxane sensitivity. Oncotarget 6, 953–968.
Hinnebusch, A.G., 2014. The scanning mechanism of eukaryotic translation initiation. Annu. Rev. Biochem. 83, 779–812.
Kozak, M., 2002. Pushing the limits of the scanning mechanism for initiation of translation. Gene 299, 1–34.
Lasorella, A., Sanson, M., Iavarone, A., 2017. FGFR-TACC gene fusions in human glioma. Neuro Oncol. 19, 475–483.
Li, K., Kong, J., Zhang, S., Zhao, T., Qian, W., 2022. Distance-dependent inhibition of translation initiation by downstream out-of-frame AUGs is consistent with a Brownian ratchet process of ribosome scanning. Genome Biol. 23, 254.
Mandell, J.D., Cannataro, V.L., Townsend, J.P., 2023. Estimation of neutral mutation rates and quantification of somatic variant selection using cancereffectsizeR. Cancer Res. 83, 500–505.
Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al., 2015. Codon optimality is a major determinant of mRNA stability. Cell 160, 1111–1124.
Qian, W., Yang, J.R., Pearson, N.M., Maclean, C., Zhang, J., 2012. Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet. 8, e1002603.
Shakya, M., Zhou, A., Dai, D., Zhong, Q., Zhou, Z., Zhang, Y., Li, X., Bholee, A.K., Chen, M., 2018. High expression of TACC2 in hepatocellular carcinoma is associated with poor prognosis. Cancer Biomarkers 22, 611–619.
Shen, X., Song, S., Li, C., Zhang, J., 2022. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. Nature 606, 725–731.
Stergachis, A.B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., Raubitschek, A., Ziegler, S., LeProust, E.M., Akey, J.M., et al., 2013. Exonic transcription factor binding directs codon choice and affects protein evolution. Science 342, 1367–1372.
Sun, M., Zhang, J., 2022. Preferred synonymous codons are translated more accurately: proteomic evidence, among-species variation, and mechanistic basis. Sci. Adv. 8, eabl9812.
Yang, Y., Wang, Z., 2019. IRES-mediated cap-independent translation, a path leading to hidden proteome. J. Mol. Cell Biol. 11, 911–919.
Zhang, G., Hubalewska, M., Ignatova, Z., 2009. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol. 16, 274–280.

Jinhui Kong, Shuo Zhang, Wenfeng Qian
State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China

University of Chinese Academy of Sciences, Beijing 100049, China

Ke Li*
State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China

* Corresponding author.
E-mail address: like@genetics.ac.cn (K. Li).

28 December 2022
Available online 8 March 2023