The introduce of a cell report

- I will present my understanding of a cell report named “Local Correlations in Codon Preferences Do Not Support a Model of tRNA Recycling”.
- For my lack of basic knowledge，I may misunderstand some points.

  If i say something wrong，just correct me.
- I can’t understand how most of the formulas of this paper work , so i skipped them . Just forgive me.
- Now，just look at some of the important points together with me.

by
yangdechang

Brief

Hussmann and Press find that observed excesses in pairs of coding sequence occurrences of the same synonymous codon are a generic consequence of the existence of spatial variation in codon preferences and therefore cannot be interpreted as evidence for the recycling of individual tRNA molecules during translation.

Highlights

- Synonymous codon usage patterns are known to vary across the genome and within genes
- Mathematically, this variation implies a diagonal-positive local covariance signal
- That signal is thus not evidence for molecular tRNA reuse by the ribosome
- Rather, it reflects a complicated covariance structure across 61 codons

SUMMARY

- It has been proposed that patterns in the usage of synonymous codons provide evidence that individual tRNA molecules are recycled through the ribosome, translating several occurrences of the same amino acid before diffusing away. The claimed evidence is based on counting the frequency with which pairs of synonymous codons are used at nearby occurrences of the same amino acid, as compared to the frequency expected if each codon were chosen independently from a single genome-wide distribution.

We show that such statistics simply measure variation in codon preferences across a genome.

As a negative control on the potential contribution of pressure to exploit tRNA recycling on these signals, we examine correlations in the usage of codons that encode different amino acids.

We find that these controls are statistically as strong as the claimed evidence and conclude that there is no information evidence that tRNA recycling is a force shaping codon usage.

This is the translation of the summary, I put it here to help you understand the aims of this paper。

- 有人提出了同义密码子的使用提供了在翻译少数几个同样的氨基酸时，在tRNA散开之前核糖体中的独立tRNA分子可被循环利用的证据。这种声称的证据是基于计量同义密码子对在[illegible]每一个密码子都是从一个全基因组分布中独立地选择的情况下的期望频率作对比。我们发现了这样的统计只能简单地测量基因组中的密码子偏好的变化。至于利用这些tRNA循环使用的压力信号的可能性分布的负控制的相关内容，我们检查了编码不同的氨基酸的密码子的之间的使用的相关性。我们发现了这些控制在统计学上是和声称的证据一样强的，并且还总结出了：并没有信息学上的证据表明tRNA的循环使用是一种塑造密码子使用的推动力。

INTRODUCTION

It introduced the basic information firstly.

Due to degeneracies in the genetic code, sets of synonymous

codons are translated into the same amino acid.

Despite the fact that substitutions between synonymous codons in a coding sequence do not change the amino acid sequence of the translated protein, synonymous codons are not used with equal frequencies in the genomes of many organisms.

The extent and directions of codon usage biases vary between organisms, between genes within an organism’s genome, and within genes.

Many theories have been advanced that invoke the mechanics of the complex chain of processes that lead from packaged DNA to translated protein to explain the observed trends,
They including, but not limited to these points:

- mutational bias,
- bias in repair or heteroduplex mismatch resolution mechanisms,
- selection for enhanced translational elongation speed or translational accuracy via the coupling of codon usage frequencies to tRNA abundance differentials,
- selection to enhance mRNA stability or to minimize mRNA secondary structure in the neighborhood of binding sites for the translation initiation,
- selection to maintain control over splicing.

The relative importance of these mechanisms in shaping the structure of codon usage biases remains poorly understood.

Then the author present another paper firstly, and try to prove the views of that paper is not right.

- Here is the views of that paper.

- In a recent paper , Cannarozzi et al. (2010) make such an inference about the dynamics of translation.
- They examine all coding sequences of the genomes of several organisms and measure several related statistics, which are based on counting the frequency with which a given pair of codons is used to encode pairs of occurrences of the same amino acid that are located close to each other in a coding sequence.
- They observe that the same codon is used for two nearby occurrences more often than would be expected if every codon choice was drawn independently from a single genome-wide distribution.

- Furthermore, they observe that nearby pairs consisting of two distinct codons that occur more often than expected tend to be codons that are translated by the same isoaccepting tRNA species.
- They interpret these results as evidence for the intriguing hypothesis that consecutive codon choices are not made independently but instead experience selective pressure to use codons from the same isoaccepting class.
- They speculate that such reuse allows a single tRNA molecule to translate multiple codons before diffusing away from the ribosome, perhaps via a physical association between the ribosome and aminoacyl-tRNA synthetases

RESULTS

- Positive Diagonal Entries Are a Generic Indicator of Nonuniform Codon Preferences.
- Signal that Survives Gene-by-Gene Shuffling Is Also Nonspecific

- The set of statistics considered do not provide specific support for this interpretation.
- The statistics are unable to distinguish between a model of codon usage in which the choices of codon used at consecutive occurrences of an amino acid are not independent and a model in which consecutive choices are independent but drawn from distributions whose parameters vary across the genome with any spatial structure at scales longer than the distance between amino acid occurrences but shorter than the entire genome.

$$n_{genome}\left[E_{genome}\left[p_{local}^{2}\right]-E_{genome}\left[p_{local}\right]^{2}\right]$$

- Let plocal be the location-specific probability with which the codon is used.
- Let ngenome be the number of sequential pairs of occurrences of the amino acid in the genome.
- Let Egenome denote taking the expected value across all such pairs.

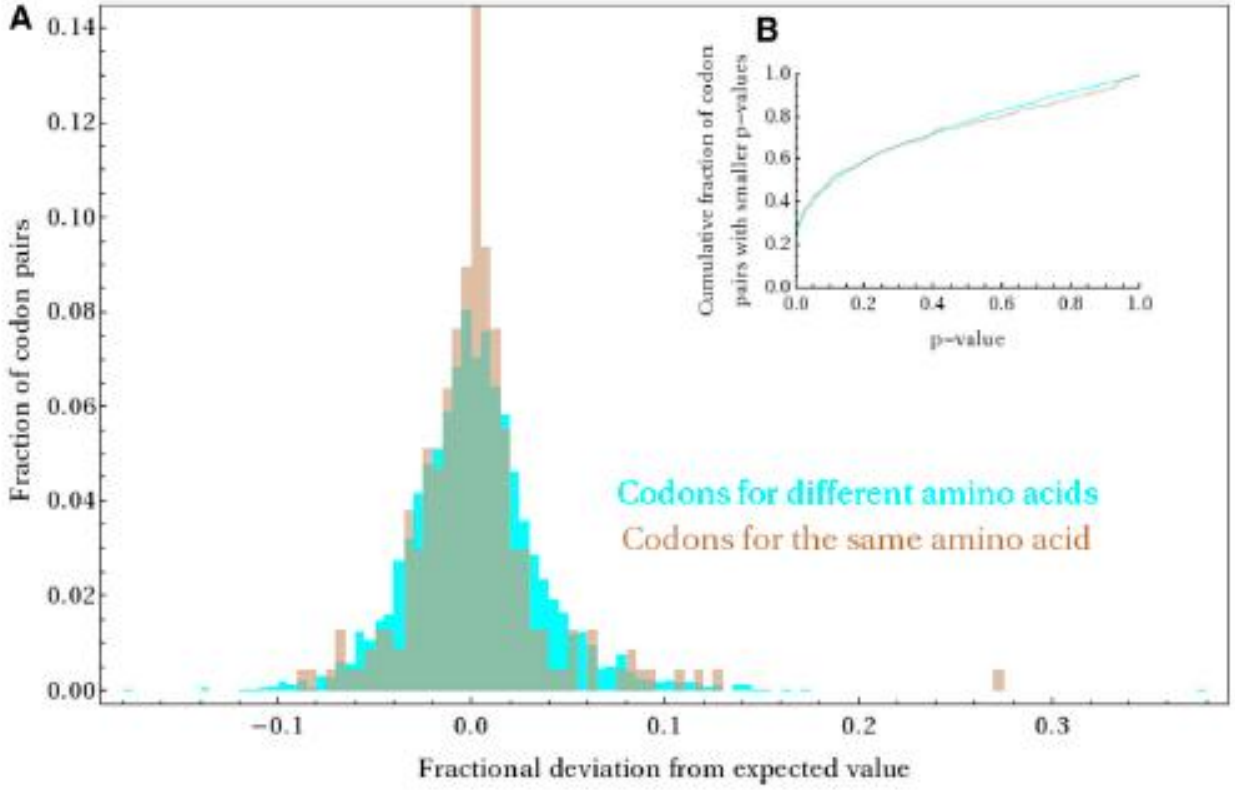


Figure . Most of the Apparent Signal in Codon Pair Usage Is due to Gene-Specific Codon Preferences

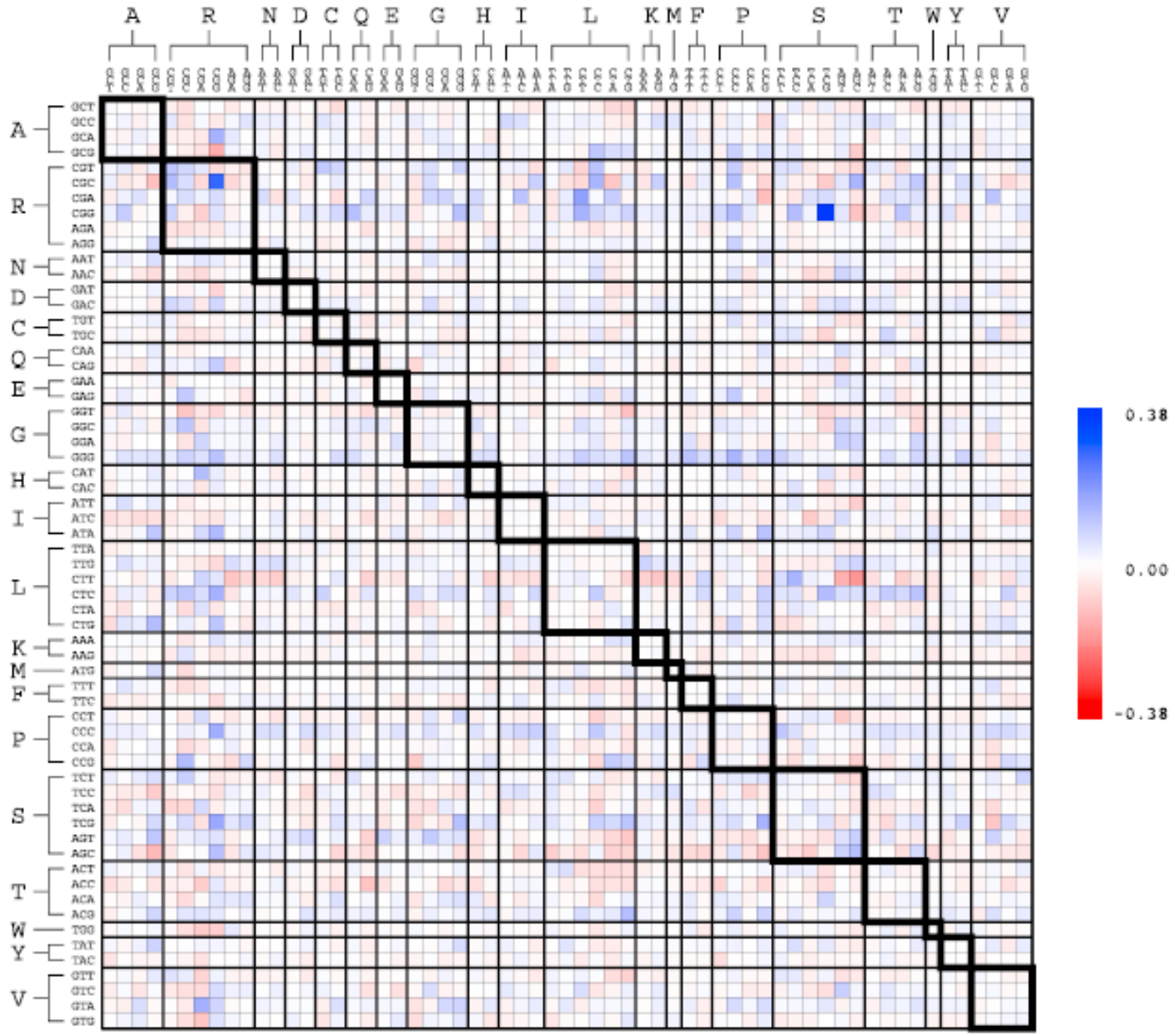
Fractional deviations of the data relative to this gene-specific null model are dramatically less extreme and less uniformly statistically significant than deviations over a genome-wide model

- The fact that universally positive values of the statistics are observed on the diagonals of matrices is now seen to be unremarkable.
- It is expected under any model of codon usage in which codon preferences are not uniform across a genome.

- Having presented this control, it should be noted that Cannarozzi et al.’s argument that “if the correlation effect was simply due to the accumulation of frequent codons in genes with biased codon composition, this effect should also be highest for frequent codons and not observed for rare codon” misstates the effect that local bias in codon composition has on correlation effects.

$$\sum_{genes} n_{gene} \left(\mathrm{E}_{gene} \left[p_{local}^2 \right] - \mathrm{E}_{gene} \left[p_{local} \right]^2 \right),$$

$$\sum_{genes} n_{gene}^{(a_i, a_j)} \left(\mathrm{E}_{gene} \left[p_{local}^{(i)} p_{local}^{(j)} \right] - \mathrm{E}_{gene} \left[p_{local}^{(i)} \right] \mathrm{E}_{gene} \left[p_{local}^{(j)} \right] \right),$$



Fractional deviations of counts of actual usage of codon pairs in all coding sequences of S.cerevisiae with respect to counts expected under a shuffling of assignments of codons to amino acids within each gene. The thickly bordered diagonal blocks contain those pairs of codons that encode the same amino acid. These diagonal blocks are not a visually distinct subset of the full matrix.

- Whereas this signature could be caused by tRNA recycling, it could also simply indicate that local codon preferences are coupled, by selection, to the identities of tRNA species.

- Taken together, these observations suggest that values in the diagonal blocks can be explained entirely by local preference structure induced by non-tRNA-recycling mechanisms and therefore cannot be taken as specific evidence that tRNA recycling is a major force shaping codon choices.

EXPERIMENTAL PROCEDURES

- Databases

Yeast coding sequences were retrieved from release 60 of the Ensembl databases using the Ensembl Perl API

- Derivation of Expected Number of Pairs under Gene-Specific Shuffling

$$\mathrm{E}_{shuffle}\left[number\ of\ pairs\ i,j\right]=\mathrm{E}_{shuffle}\left[\sum^{N}\sum^{n_g-1}1^{(i,j)}_{g,k}\right]$$

$$\mathrm{E}_{shuffle}\left[number\ of\ pairs\ i,j\right]=\mathrm{E}_{shuffle}\left[\sum_{g=1}^{N}\sum_{k=1}^{n_g^{(a_i,a_j)}-1}1^{(i,j)}_{g,k}\right]$$

P

on j]

$$=\sum_{g=1}^{N}\sum_{k=1}^{n_g^{(a_i,a_j)}-1}\mathrm{E}_{shuffle}\left[1^{(i,j)}_{g,k}\right]$$

$$=\sum_{g=1}^{N}n_g^{(a_i,a_j)}\frac{c_{g,i}}{n_g^{(a_i)}}\frac{c_{g,j}}{n_g^{(a_j)}}.$$

$$=\begin{cases}\sum_{g=1}^{N}\frac{c_{g,i}c_{g,j}}{n_g}, & i\neq j\end{cases}.$$

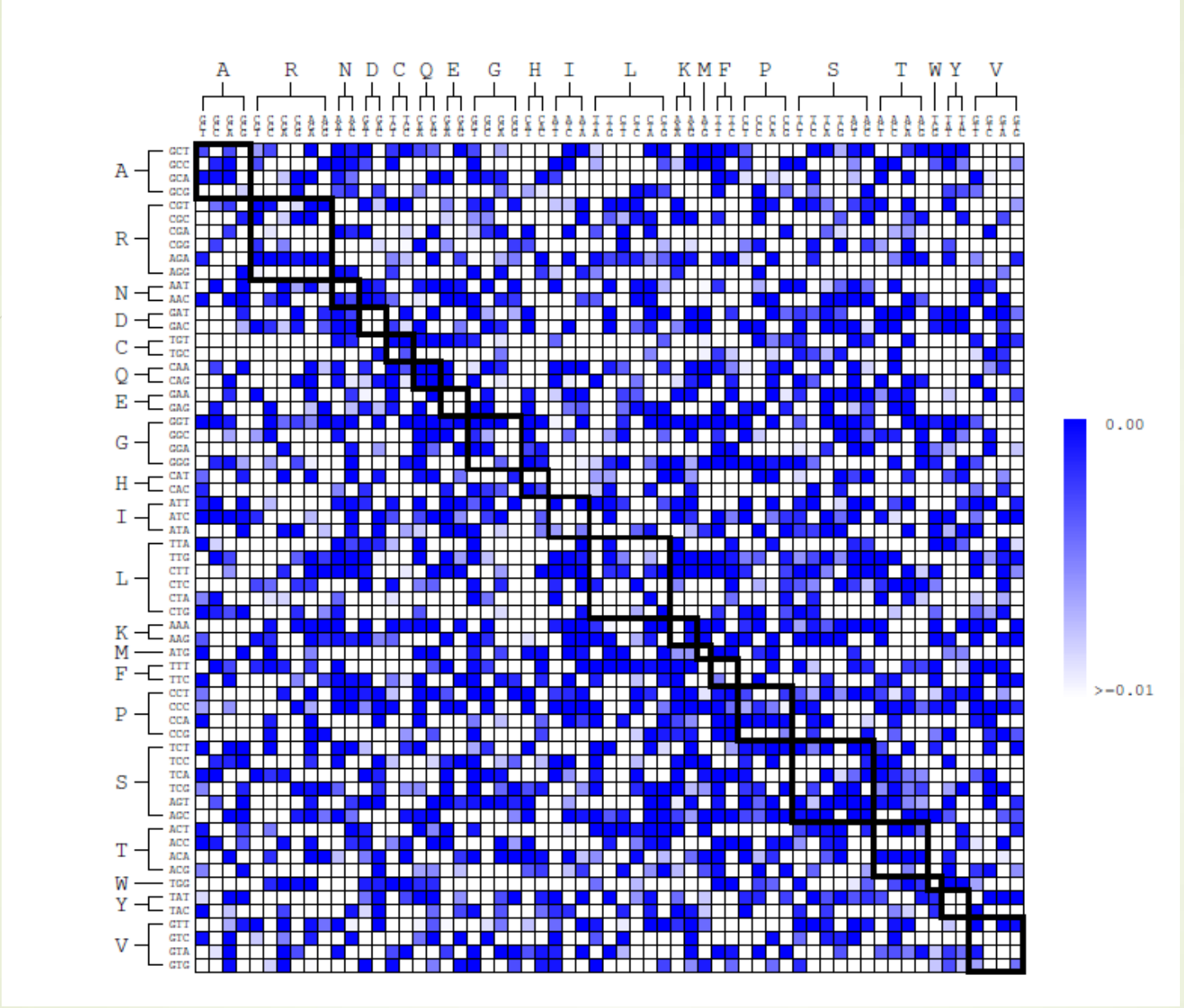
Supplemental Experimental Procedures

It includes two main part:

- Expected values of statistics assuming genome-wide preferences under model with gene-specic preferences
- tRNA pairing index and tRNA correlation as a function of distance

it is too complex to understand its formula,it may get the results that Expected values of statistics assuming genome-wide preferences under model with gene-specic preferences.

It may be represented by this picture.



Each entry in the matrix is shaded according to its empirical p-value

Statistical signicances of the fractional deviations

- tRNA correlation as a function of distance considers the deviation in usage of pairs of codons to encode a pair of occurrences of the same amino acid from the usage expected if each occurrence were drawn from the same genome-wide codon preference distribution as a function of the distance between the amino acid occurrences.

That’s all !

Thanks for your listening~