

Research article

Open Access

Strand bias in complementary single-nucleotide polymorphisms of transcribed human sequences: evidence for functional effects of synonymous polymorphisms

Hui-Qi Qu¹, Steve G Lawrence², Fan Guo¹, Jacek Majewski² and Constantin Polychronakos*^{1,3}

Address: ¹Endocrine Genetics Laboratory, The McGill University Health Center (Montreal Children's Hospital), Montréal, Québec, Canada, ²Department of Human Genetics, McGill University, Montréal, Québec, Canada and ³Department of Pediatrics, The McGill University Health Center (Montreal Children's Hospital), 2300 Tupper, Montréal, Québec H3H 1P3, Canada

Email: Hui-Qi Qu - hui.qi.qu@mail.mcgill.ca; Steve G Lawrence - steve.lawrence@mail.mcgill.ca; Fan Guo - fguo4@cs.mcgill.ca; Jacek Majewski - jacek.majewski@mcgill.ca; Constantin Polychronakos* - constantin.polychronakos@mcgill.ca

* Corresponding author

Published: 17 August 2006

Received: 09 March 2006

BMC Genomics 2006, 7:213 doi:10.1186/1471-2164-7-213

Accepted: 17 August 2006

This article is available from: <http://www.biomedcentral.com/1471-2164/7/213>

© 2006 Qu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Complementary single-nucleotide polymorphisms (SNPs) may not be distributed equally between two DNA strands if the strands are functionally distinct, such as in transcribed genes. In introns, an excess of A↔G over the complementary C↔T substitutions had previously been found and attributed to transcription-coupled repair (TCR), demonstrating the valuable functional clues that can be obtained by studying such asymmetry. Here we studied asymmetry of human synonymous SNPs (sSNPs) in the fourfold degenerate (FFD) sites as compared to intronic SNPs (iSNPs).

Results: The identities of the ancestral bases and the direction of mutations were inferred from human-chimpanzee genomic alignment. After correction for background nucleotide composition, excess of A→G over the complementary T→C polymorphisms, which was observed previously and can be explained by TCR, was confirmed in FFD SNPs and iSNPs. However, when SNPs were separately examined according to whether they mapped to a CpG dinucleotide or not, an excess of C→T over G→A polymorphisms was found in non-CpG site FFD SNPs but was absent from iSNPs and CpG site FFD SNPs.

Conclusion: The genome-wide discrepancy of human FFD SNPs provides novel evidence for widespread selective pressure due to functional effects of sSNPs. The similar asymmetry pattern of FFD SNPs and iSNPs that map to a CpG can be explained by transcription-coupled mechanisms, including TCR and transcription-coupled mutation. Because of the hypermutability of CpG sites, more CpG site FFD SNPs are relatively younger and have confronted less selection effect than non-CpG FFD SNPs, which can explain the asymmetric discrepancy of CpG site FFD SNPs vs. non-CpG site FFD SNPs.

Background

Single-nucleotide polymorphisms (SNPs) involve two complementary base substitutions, one on each DNA strand. Where the two DNA strands are functionally distinct (such as in transcribed sequences), the two complementary substitutions may not occur with equal frequency on each strand[1], due to transcription-related mutation/repair mechanisms or selective pressure from functional effects on mRNA. A↔G vs. C↔T asymmetry in the two DNA strands is well known to exist in prokaryotes[2]. In the human, there is an excess of C↔T over G↔A in mutations causing Mendelian disorders[3] while excess of A→G substitutions in the sense strand of transcribed intronic sequences was found when comparing a ~1.5 Mb region of human chromosome 7 to its chimpanzee orthologue[4]. Both reports attributed the bias to transcription-coupled repair (TCR), and further support for transcription-coupled effect has been provided by the correlation between strand bias in nucleotide composition of transcribed sequences with transcription levels[5]. However, the conflicting results observed within coding and intronic sequences have not been explored further. It is highly unlikely that TCR distinguishes between exons and introns. Furthermore, our current knowledge of TCR[6,7] suggest that its action would affect the proportion of A→G vs. T→C mutations, but should not affect other mutations. An alternative explanation for the observed discrepancy between exons and introns is that synonymous exonic substitutions in mammals may be under non-trivial selective pressures, as has been suggested by some recent studies[8,9]. An important effect of synonymous coding mutations is the association with gene splicing[10,11]. In humans, evidence of selection on synonymous variations may have a profound effect on how we view the role of synonymous variations in genetic disease and phenotypic variability. Further research is needed besides these studies: the analysis of disease-causing mutations[3] required assumptions about likelihood of coming to clinical attention based on chemical differences between substituted amino acids, while the work on

intronic sequences[4] was confined to a single ~1.5 Mb region and the genome-wide applicability of the results remains to be proven. Neither study explored differences between introns and exons to distinguish mutation/repair effects from alterations in RNA function. To our knowledge, strand asymmetry in human SNPs has not been fully examined for possible clues about the mutational mechanisms that created them and/or their potential functional significance. We therefore undertook a systematic examination of human coding SNPs in the fourfold degenerate (FFD) codon site and a random sample of intronic SNPs (iSNPs) for strand asymmetry between A↔G and C↔T polymorphisms.

Results

The identities of the ancestral bases and the direction of mutations were inferred from human-chimpanzee genomic alignment. To avoid bias from amino acid composition in the third codon position, only FFD SNPs were included in the analysis. On this basis, from the full list of Perlegen validated SNPs, 2,374 FFD SNPs involving A↔G or C↔T polymorphisms were identified for further investigation (Table 1). To increase the statistical power of this study, a larger number of iSNPs were included in the analysis. As edges of introns are known to be under selective constraint[8,12,13], all iSNPs investigated were chosen to be more than 200 bp from each intronic end. In addition, first introns have specific substitution patterns because they are enriched for CpG islands[8] which, being unmethylated[14], are not hypermutable. Also, iSNPs in first introns may be under purifying selection[8,12,13]. Therefore, iSNPs in first introns were not included in the subset.

To control the observed substitution rates for background nucleotide composition (see Methods), the nucleotide content was determined for all known human intronic sites and FFD sites of coding regions (Table 2). After the background correction, a large excess of A→G polymorphisms over the complementary T→C was found in both

Table 1: The asymmetry pattern of A↔G and C↔T iSNPs and FFD SNPs

Substitution type	A→G	T→C	G→A	C→T	Total
iSNP					
Non-CpG site	1411 (19.5%)	1030 (14.2%)	1200 (16.6%)	1052 (14.5%)	4693 (64.8%)
CpG site	678 (9.4%)	561 (7.8%)	664 (9.2%)	643 (8.9%)	2546 (35.3%)
Non-CpG vs. CpG*	$\chi^2 = 3.2, v = 1, p = 0.074$		$\chi^2 = 2.1, v = 1, p = 0.151$		
Total	2089 (28.9%)	1591 (22.0%)	1864 (25.8%)	1695 (23.4%)	7239 (100%)
FFD SNPs					
Non-CpG site	148(6.2%)	127(5.3%)	207(8.7%)	307(12.9%)	789(33.2%)
CpG site	316(13.3%)	163(6.9%)	654(27.5%)	452(19.0%)	1585(66.8%)
Non-CpG vs. CpG*	$\chi^2 = 10.9, v = 1, p = 0.001$		$\chi^2 = 50.1, v = 1, p < 0.001$		
Total	464(19.5%)	290(12.2%)	861(36.3%)	759(32.0%)	2374(100.0%)

* χ^2 test of the difference of complementary substitution between non-CpG site and CpG site.

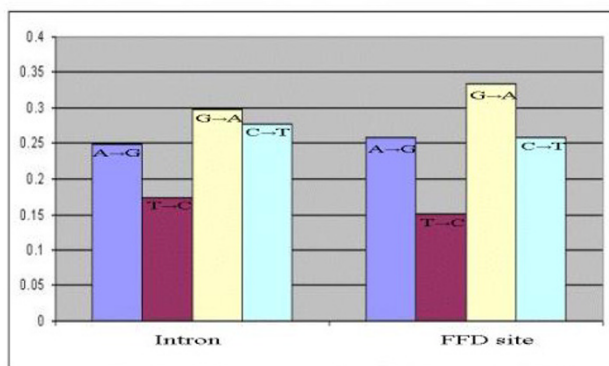
Table 2: The nucleotide composition of human genome introns and FFD codons

Region	A	T	G	C	Total
Intron*	126,530,426 (28.0%)	139,073,101 (30.7%)	94,268,606 (20.8%)	92,418,338 (20.4%)	452,290,471 (100.0%)
FFD codons	575,500 (19.5%)	615,324 (20.8%)	822,103 (27.8%)	939,137 (31.8%)	2,952,064 (100.0%)

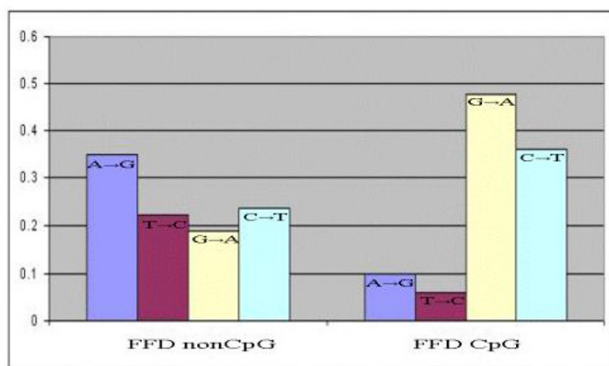
* The first introns as well as the first and last 200 bp of each intron were excluded;

iSNPs, $\chi^2 = 122.9$, $\nu = 1$, $p < 0.001$, ratio (95%CI) = 1.44 (1.35, 1.54), and FFD SNPs, $\chi^2 = 52.7$, $\nu = 1$, $p < 0.001$, ratio (95%CI) = 1.71 (1.48, 1.98) (Fig 1a). An excess of G→A changes over C→T was also observed in iSNPs, $\chi^2 = 5.0$, $\nu = 1$, $p = 0.025$, ratio (95%CI) = 1.08 (1.01, 1.15) and, more dramatically, at FFD SNPs: $\chi^2 = 27.2$, $\nu = 1$, $p <$

0.001, ratio (95%CI) = 1.30 (1.18, 1.43). We thus confirm, genome-wide and within *Homo Sapiens*, the strand bias in substitution rates, which has been found on a human chr. 7 region when compared to chimpanzee[4]. The excess of A→G polymorphisms resulting in iSNPs is concordant with the finding by Green et al [4]. This result can be explained by differential effect of TCR on transcribed and untranscribed DNA strands of genes.



a.



b.

Figure 1

The proportion of each type of A↔G and C↔T substitution (the ancestral allele vs. the recent allele of 7,239 iSNPs and 2,374 FFD SNPs. (a) The proportion of each type of substitution corrected by nucleotide compositions. (b) The FFD SNP distribution corrected by FFD codon compositions. The non-CpG site sSNP distribution is corrected by four types of FFD CpG codons, i.e. NDA, NNT|H, NDG, and NNC|H (D represents A or G or T, and H represents A or C or T). FFD CpG site SNP distribution is corrected by four types of FFD CpG codon compositions, i.e., NCA, NNT|G, NCG, and NNC|G.

In order to investigate the effect of hypermutable CpG dinucleotides, and to correct for their excess in exons over introns (Table 1), SNPs were next analyzed separately according to whether or not the polymorphism occurred within a CpG site. The hypermutability of CpG dinucleotides is well documented and results from methylation-induced deamination of 5-methyl cytosine[15]. If the deamination occurs on the sense strand, it results in [C→T]pG; if the cytosine deamination takes place on the antisense strand, it produces a Cp [G→A] on the sense strand. Thus, A SNP at a CpG site has the pattern of YpG or CpR (Y represents C or T, and R represents A or G). In introns, the mutational asymmetry does not differ between CpG and non-CpG sites (Table 1). Unlike iSNPs, a dramatic difference between CpG and non-CpG sites was noted in FFD SNPs. After correction for the codon compositions, different asymmetry pattern of G→A vs. C→T between non-CpG sites and CpG sites was noticed (Table 3, Fig 1b). Excess C→T over G→A can be seen in non-CpG FFD SNPs, but not CpG FFD SNPs. Because this finding is present in exons but absent in introns, it is very unlikely that it can be explained by any transcription-related mutational and/or repair mechanism, but suggests selective pressure due to effects on the function of the mature transcript.

An obvious example of such an effect is the creation of an AT dinucleotide by a G→A mutation in a FFD site when a T is the first nucleotide of the next codon. AU dinucleotides are known to be targets of RNaseL endonucleolytic cleavage[16]. A|U dinucleotides at synonymous dicodon boundaries could allow more efficient 3'-5' degradation by endonucleolytic cleavage[17] and, consequently, drive purifying selection. Thus, our interpretation makes the prediction of fewer than expected G→A polymorphisms at FFD sites preceding a codon with a T in the first position. Our analysis indeed shows a dramatic deficit of

Table 3: The proportions of A→G and C→T FFD substitutions corrected by codon compositions

FFD SNPs	Substitution type	Observed number	Codon type*	Codon count	Corrected proportion	χ ² test	Asymmetry ratio (95%CI)
Non-CpG site	A→G	148	NDA	265,987	0.350	χ ² = 13.6, v = 1, p < 0.001	A→G vs. T→C 1.56 (1.23, 1.97)
	T→C	127	NNT H	355,455	0.225		
	G→A	207	NDG	689,990	0.189		
CpG site	C→T	307	NNC H	818,126	0.236	χ ² = 26.0, v = 1, p < 0.001	A→G vs. T→C 1.63 (1.35, 1.97)
	A→G	316	NCA	309,513	0.099		
	T→C	163	NNT G	259,869	0.061		
	G→A	654	NCG	132,113	0.479		
	C→T	452	NNC G	121,011	0.361		

* N represents A or C or G or T, D represents A or G or T, and H represents A or C or T

G→A polymorphisms that occur before a codon that starts with a T (Table 4).

Discussions and conclusion

It is of great interest that the C→T excess over the complementary G→A in non-CpG FFD SNPs is not seen in iSNPs or FFD SNPs that are part of a CpG. As iSNPs and FFD SNPs should confront the same transcription-coupled mechanisms, including TCR and transcription-coupled mutation (TCM)[18], the C→T excess of FFD SNPs must be driven by mechanisms other than mutational/repair factors. Alternatively, biologically significant effects of synonymous SNPs (sSNPs) on aspects of RNA function other than protein coding may exist and be subject to selective pressures. Unlike lower organisms, it is still contentious whether selection for translational efficiency does[19,20] or does not [21-24] play a major role in shaping codon usage (and therefore sSNP frequencies) in mammals. There is little variation in iso-acceptor tRNA gene numbers and the population sizes are likely too low to reflect very weak selective pressures[23]. On the other hand, translation may be affected by RNA secondary structure which, like splicing, mRNA stability, or other less well understood RNA functions, may be significantly altered by single-nucleotide changes. Such mechanisms have recently been suggested in a few studies[8,9,25]. If

sSNPs do have such biological effects, there is evidence to suggest that changes in mRNA secondary structure are likely to play an important role in mediating them[25,26]. Given the evidence of compromised mRNA stability in the presence of A|T dinucleotides at dicodon boundaries [16,17], G→A polymorphisms at FFD sites may have deleterious effects that C→T does not, thus creating selection pressure that favors C→T if the next codon begins with a T. In this report we show that this is indeed the case.

The different asymmetry pattern between non-CpG and CpG sites can be attributed to the hypermutability of the latter[27]. The effects of selection on the observed mutation patterns are most pronounced in relatively slowly mutating, non-CpG sites. Because of the hypermutability of CpG sites, more CpG site FFD SNPs are younger and have confronted less selective pressure than non-CpG FFD SNPs. For the same selection effect on A|T dinucleotides, A→G polymorphism may also confront more selection pressure than T→C, which can also explain why the A→G excess is not significantly different in FFD non-CpG and intronic CpG sites.

In conclusion, we confirm the genome-wide excess of A→G over T→C mutations previously reported in a small

Table 4: Decreased FFD SNPs at A|T dinucleotides

FFD SNPs	Substitution type	nonT	T	χ ² test
Non-CpG site	G→A	161	46	χ ² = 7.9, v = 1, p = 0.005
	A→G	95	53	
CpG site	G→A	505	149	χ ² = 19.1, v = 1, p < 0.001
	A→G	202	114	

region of chr. 7 [4], a finding that points to TCR as an important factor in human mutagenesis. More importantly, our analysis of FFD SNPs clearly suggests a mechanism that operates differentially in intronic vs. exonic sequences. We propose that selective pressure related to changes in mRNA stability is the most likely explanation. In view of the balance between selective and mutational pressures, we provide satisfactory explanation for the previous contradictory findings of mutation rates in humans [3,4,28]. Our finding further highlights the importance of not overlooking potential function by the sSNPs, which may not be as selectively neutral as is generally thought [29], an important consideration given the expected wealth of complex-disease association data to come out of the new genotyping technologies.

Methods

SNP information collection

Considering the possibility that some SNPs recorded by NCBI dbSNP database may not be reliable and result from DNA sequencing errors, we performed the investigation using the Perlegen dataset of DNA variation genotyping [30,31]. The SNPs were all identically ascertained by microarray resequencing of the genome, and verified in multiple populations. Only single nucleotide polymorphisms with two alleles were included. SNPs in sex chromosomes were not included in this study. Reference sequences of the SNPs in 22 pairs of human autosomes were bulk-downloaded from the NCBI dbSNP database build 124 [32].

The orientation of SNP reference sequence

The dbSNP reference sequences of iSNPs can not be aligned with mRNA sequence directly. Some FFD SNP reference sequences have intronic sequence included, and some genes have different mRNA transcripts from alternative splicing. Therefore, instead of aligning SNP sequence with mRNA sequence, we wrote Java scripts to determine the orientation of a dbSNP reference sequence in the DNA coding strand. The corresponding NCBI genome DNA contig sequence was first downloaded from the NCBI reference sequences [33]. Then, a SNP reference sequence was aligned with the contig sequence around the SNP contig position and the orientation in the contig sequence was determined. The orientation of mRNA sequence in the same contig sequence was acquired from the annotation of dbSNP. Based on these two orientations, the orientational relation of SNP reference sequence and mRNA sequence was known. The corresponding nucleotide polymorphism in the DNA coding strand were determined consequently.

Correction for nucleotide or codon compositions

In order to determine the relative rates of each substitution, the observed counts were corrected for the back-

ground frequencies of nucleotides or codons. Both the intronic and FFD nucleotide compositions were acquired from the 14,029 genes annotated by the CCDS database [34,35]. For background intronic nucleotide compositions, the first introns as well as the first and last 200 bp of each intron were excluded. As an example of correction, for A→G polymorphism, the observed number ($N_{A\rightarrow G}$) corrected by the frequency of adenine (P_A) was calculated as:

$$P_{A\rightarrow G} = \frac{N_{A\rightarrow G}/P_A}{N_{A\rightarrow G}/P_A + N_{T\rightarrow C}/P_T + N_{G\rightarrow A}/P_G + N_{C\rightarrow T}/P_C}$$

The corrected proportions of each type of polymorphisms within the A↔G-C↔T pair were calculated in the same way. For the computation of the asymmetry ratio of complementarity polymorphism, such as A→G vs. T→C,

$$\text{Ratio}_{A\rightarrow G \text{ vs. } T\rightarrow C} = \frac{N_{A\rightarrow G}/\text{Genome-count (A)}}{N_{T\rightarrow C}/\text{Genome-count (T)}}$$

The 95% CI was computed by logistic regression analysis.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

HQQ carried out most of the implementation and analysis, and drafted the manuscript. SGL participated in the SNP database mining. FG wrote the JAVA scripts. JM participated in the study design, the SNP database mining, and the manuscript revision. CP conceived of the study, participated in its design and coordination, and participated in preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Luc Marchand for technical support and the three anonymous reviewers for their valuable comments. Funded by Genome Canada, Genome Quebec, the Juvenile Diabetes Research Foundation International, and Canadian Institutes of Health Research. HQQ was supported by a post-doctoral fellowship from the Montreal Children's Hospital Foundation. JM is a recipient of a Canada Research Chair.

References

1. Frank AC, Lobry JR: **Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms.** *Gene* 1999, **238**(1):65-77.
2. Lobry JR: **Asymmetric substitution patterns in the two DNA strands of bacteria.** *Mol Biol Evol* 1996, **13**(5):660-665.
3. Krawczak M, Ball EV, Cooper DN: **Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes.** *Am J Hum Genet* 1998, **63**(2):474-488.
4. Green P, Ewing B, Miller W, Thomas PJ, Green ED: **Transcription-associated mutational asymmetry in mammalian evolution.** *Nat Genet* 2003, **33**(4):514-517.
5. Majewski J: **Dependence of mutational asymmetry on gene-expression levels in the human genome.** *Am J Hum Genet* 2003, **73**(3):688-692.

6. Jiricny J, Su SS, Wood SG, Modrich P: **Mismatch-containing oligonucleotide duplexes bound by the E. coli mutS-encoded protein.** *Nucleic Acids Res* 1988, **16(16)**:7843-7853.
7. Lamers MH, Perrakis A, Enzlin JH, Winterwerp HH, de Wind N, Sixma TK: **The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch.** *Nature* 2000, **407(6805)**:711-717.
8. Chamary JV, Hurst LD: **Similar rates but different modes of sequence evolution in introns and at exonic silent sites in rodents: evidence for selectively driven codon usage.** *Mol Biol Evol* 2004, **21(6)**:1014-1023.
9. Willie E, Majewski J: **Evidence for codon bias selection at the pre-mRNA level in eukaryotes.** *Trends Genet* 2004, **20(11)**:534-538.
10. Cartegni L, Chew SL, Krainer AR: **Listening to silence and understanding nonsense: exonic mutations that affect splicing.** *Nat Rev Genet* 2002, **3(4)**:285-298.
11. Pagani F, Baralle FE: **Genomic variants in exons and introns: identifying the splicing spoilers.** *Nat Rev Genet* 2004, **5(5)**:389-396.
12. Majewski J, Ott J: **Distribution and characterization of regulatory elements in the human genome.** *Genome Res* 2002, **12(12)**:1827-1836.
13. Keightley PD, Gaffney DJ: **Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents.** *Proc Natl Acad Sci U S A* 2003, **100(23)**:13402-13406.
14. Bird AP: **CpG-rich islands and the function of DNA methylation.** *Nature* 1986, **321(6067)**:209-213.
15. Strachan T RAP: **Human molecular genetics.** 2nd Ed edition. Oxford, BIOS Scientific; 1999.
16. Carroll SS, Chen E, Viscount T, Geib J, Sardana MK, Gehman J, Kuo LC: **Cleavage of Oligoribonucleotides by the 2',5'-Oligoadenylate-dependent Ribonuclease L.** *J Biol Chem* 1996, **271(9)**:4988-4992.
17. Carlini DB: **Context-dependent codon bias and messenger RNA longevity in the yeast transcriptome.** *Mol Biol Evol* 2005, **22(6)**:1403-1411.
18. Beletskii A, Bhagwat AS: **Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli.** *Proc Natl Acad Sci U S A* 1996, **93(24)**:13919-13924.
19. Comeron JM: **Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence.** *Genetics* 2004, **167(3)**:1293-1304.
20. Lavner Y, Kotlar D: **Codon bias as a factor in regulating expression via translation rate in the human genome.** *Gene* 2005, **345(1)**:127-138.
21. Eyre-Walker AC: **An analysis of codon usage in mammals: selection or mutation bias?** *J Mol Evol* 1991, **33(5)**:442-449.
22. Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T: **Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis.** *J Mol Evol* 2001, **53(4-5)**:290-298.
23. Duret L: **Evolution of synonymous codon usage in metazoans.** *Curr Opin Genet Dev* 2002, **12(6)**:640-649.
24. dos Reis M, Savva R, Wernisch L: **Solving the riddle of codon usage preferences: a test for translational selection.** *Nucleic Acids Res* 2004, **32(17)**:5036-5044.
25. Duan J, Antezana MA: **Mammalian mutation pressure, synonymous codon choice, and mRNA degradation.** *J Mol Evol* 2003, **57(6)**:694-701.
26. Chamary JV, Hurst LD: **Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals.** *Genome Biol* 2005, **6(9)**:R75.
27. Subramanian S, Kumar S: **Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes.** *Genome Res* 2003, **13(5)**:838-844.
28. Hanawalt PC: **Transcription-coupled repair and human disease.** *Science* 1994, **266(5193)**:1957-1958.
29. Kimura M: **The neutral theory of molecular evolution.** Cambridge [Cambridgeshire]; New York, Cambridge University Press; 1983:Bibliography: p. [328]-353.
30. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR: **Whole-Genome Patterns of Common DNA Variation in Three Human Populations.** *Science* 2005, **307(5712)**:1072-1079.
31. **Perlegen Sciences, Inc.** [<http://www.perlegen.com/>].
32. **NCBI dbSNP database** [<http://www.ncbi.nlm.nih.gov/SNP/>].
33. **NCBI reference sequences** [ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/].
34. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CV, Thomas DJ, Weber RJ, Haussler D, Kent WJ: **The UCSC Genome Browser Database.** *Nucleic Acids Res* 2003, **31(1)**:51-54.
35. **UCSC Genome Bioinformatics** [<http://genome.ucsc.edu/>].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

