

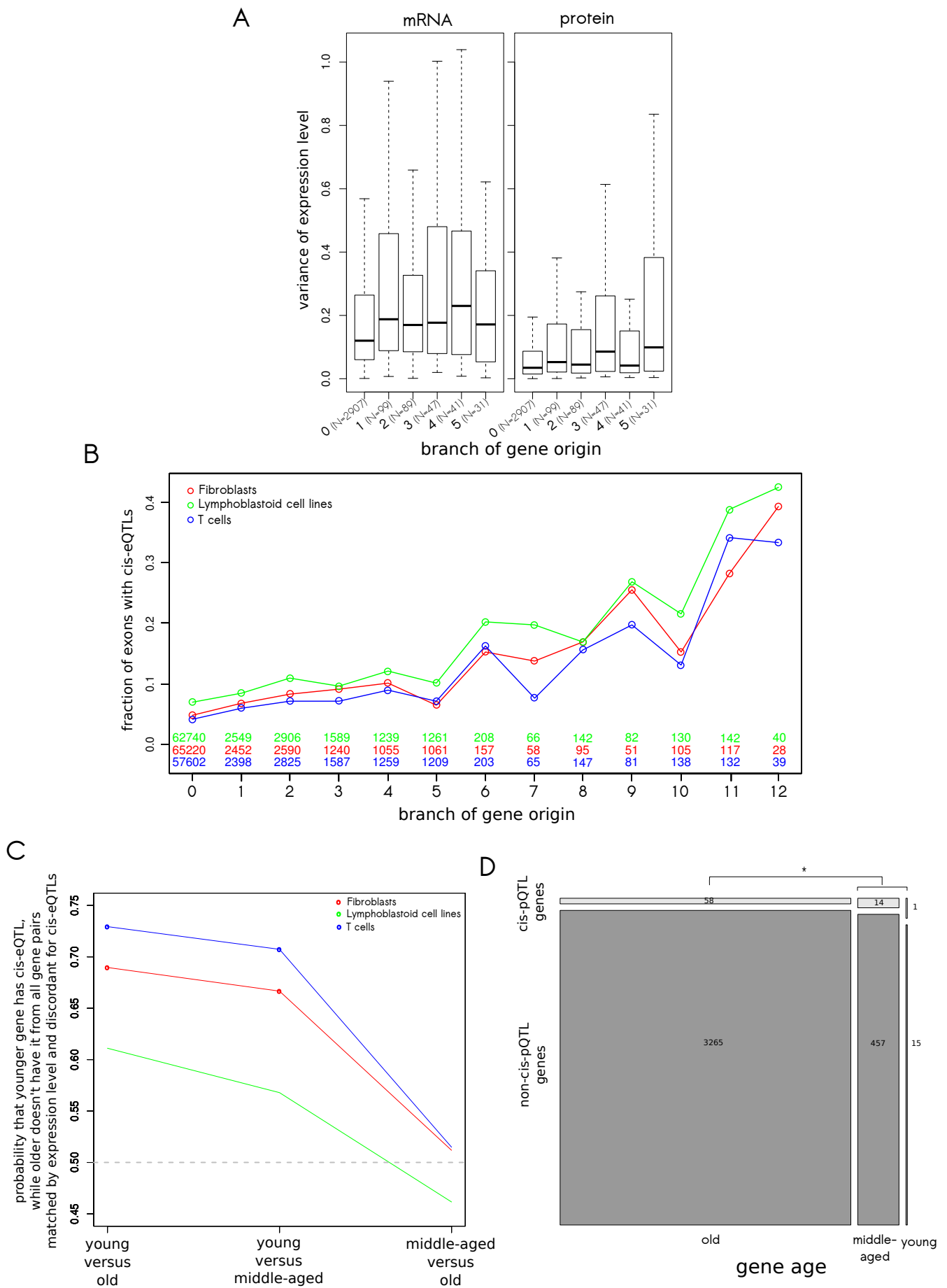
The American Journal of Human Genetics, Volume 95

Supplemental Data

## **Gene Age Predicts the Strength of Purifying Selection**

### **Acting on Gene Expression Variation in Humans**

**Konstantin Y. Popadin, Maria Gutierrez-Arcelus, Tuuli Lappalainen, Alfonso Buil, Julia Steinberg, Sergey I. Nikolaev, Samuel W. Lukowski, Georgii A. Bazykin, Vladimir B. Seplyarskiy, Panagiotis Ioannidis, Evgeny M. Zdobnov, Emmanouil T. Dermitzakis, and Stylianos E. Antonarakis**



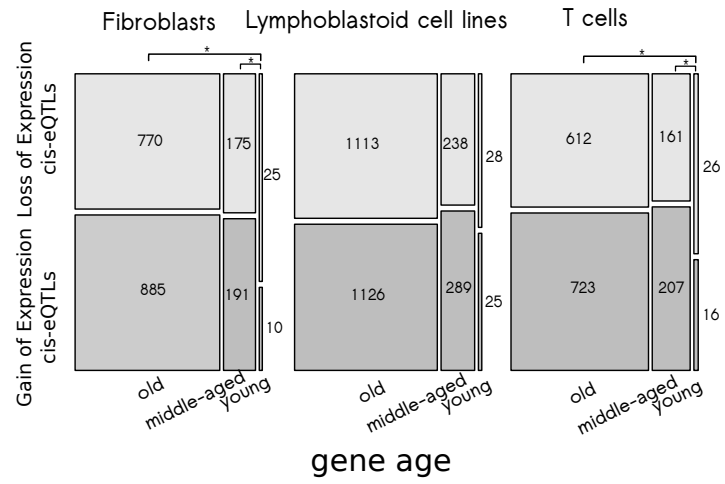
**Figure S1.**

(A) Variations in mRNA expression levels and protein levels are associated with gene age.

(B) Exons of young genes are enriched in cis-eQTLs. The number of expressed exons for each gene age category is presented above the x axis.

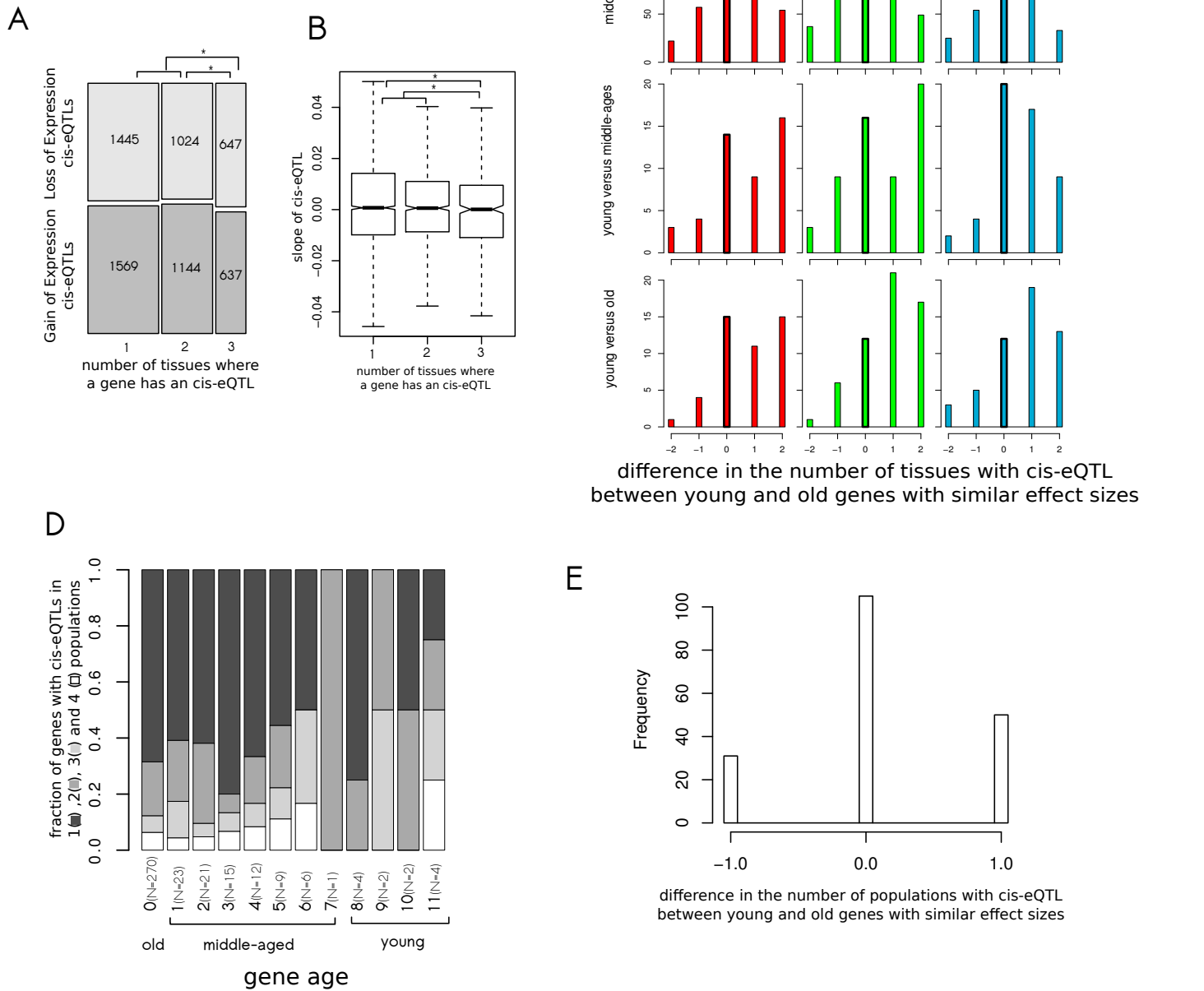
(C) Younger genes are enriched in cis-eQTLs irrespective of the expression level. P-values  $\leq 0.01$  (binomial test) are marked by circles.

(D) cis-pQTL genes are depleted in old versus young and middle-aged genes.



**Figure S2.**

Loss of Expression (LOE) cis-eQTLs are enriched among young genes in Fibroblasts and T cells of the GenCord collection.



**Figure S3.**

(A) Genes with cis-eQTLs in multiple GenCord cell types tend to have loss of expression (LOE) cis-eQTLs.

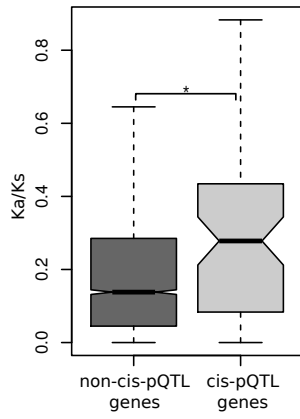
(B) Genes with cis-eQTLs in multiple GenCord cell types tend to have more negative slopes.

(C) cis-eQTL young genes are more tissue-shared than cis-eQTL old genes even if they have a similar effect size: histograms represent the difference between tissue-shared patterns of young and old genes with the similar effect size (2 on the X axis marks more tissue-shared cis-eQTLs of young genes, -2 marks more tissue-specific cis-eQTLs of young genes). All distributions are right-shifted compared to the expected zero bin, which is marked by a bold black margin.

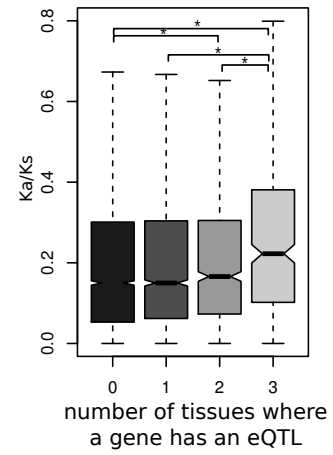
(D) Population-specificity of cis-eQTL genes is associated with gene age: the fraction of population-shared cis-eQTL genes increases among young genes.

(E) cis-eQTLs of young genes are more population-shared than cis-eQTLs of old genes even if they have a similar effect size: histogram represents the difference between population-shared patterns of young and old genes with the similar effect size (1 on the X axis marks more population-shared cis-eQTLs of young genes, -1 marks more population-specific cis-eQTLs of young genes).

A



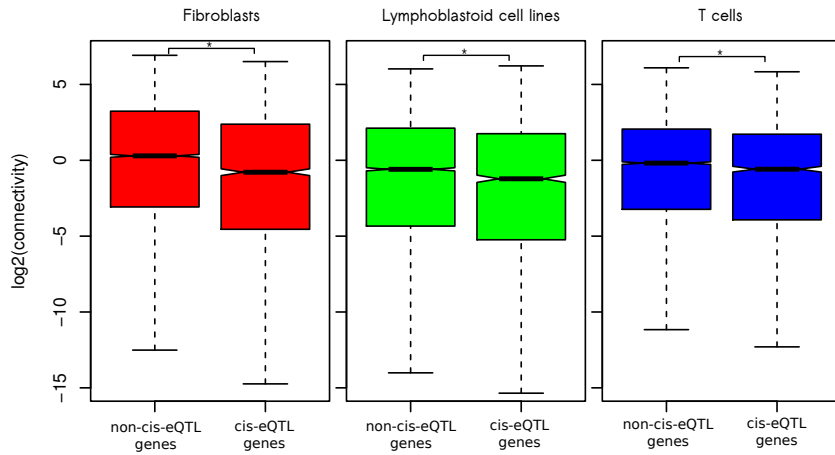
B

**Figure S4.**

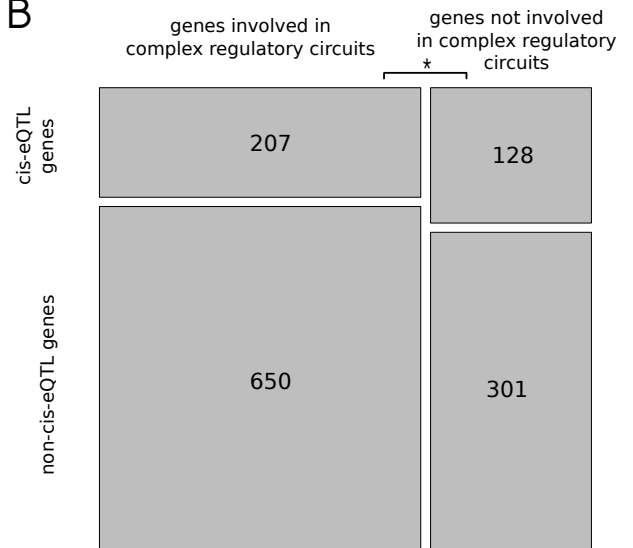
(A) Ka/Ks values of cis-pQTL genes are higher than Ka/Ks values of non-cis-pQTL genes.

(B) Ka/Ks of genes with cis-eQTLs in multiple tissues are higher.

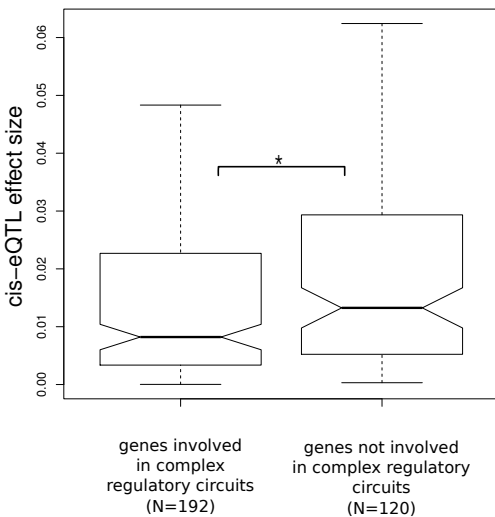
A



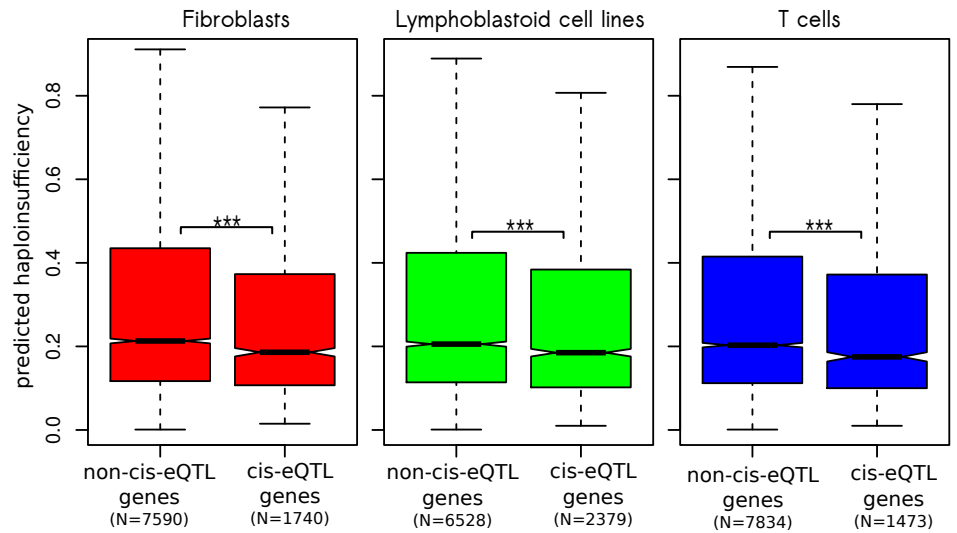
B



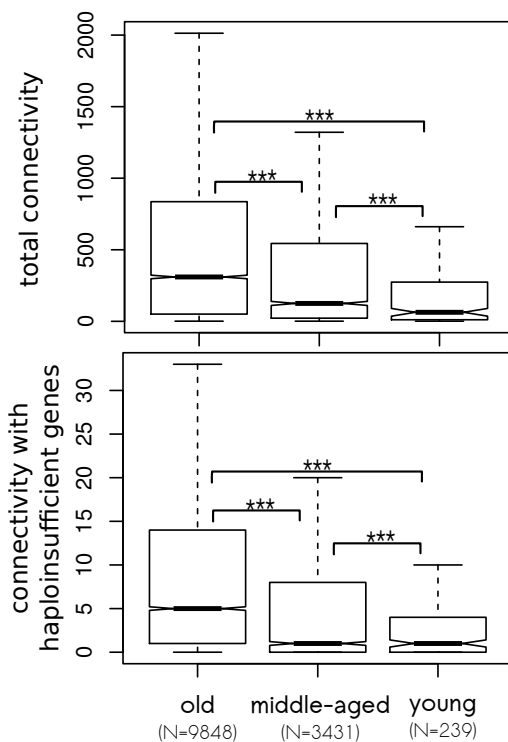
C



D



E



### Figure S5.

(A) Connectivity of genes in co-expression networks is lower in cis-eQTL genes.

(B) cis-eQTL genes have a deficit of genes involved in complex regulatory circuits.

(C) Genes involved in complex regulatory circuits have lower cis-eQTL effect size than genes, not involved in the circuits.

(D) Predicted haploinsufficiency is higher for non-cis-eQTL genes versus cis-eQTL genes in all three cell types of the GenCord collection.

(E) Total connectivity (the upper panel) as well as connectivity with haploinsufficient genes (the bottom panel) increases with gene age.

analyzed relationships controlled variables	branch ~ Ka/Ks	branch ~ connectivity	branch ~ expression level
none	0.306	-0.139	-0.099
Ka/Ks		-0.112	-0.061
connectivity	0.295		-0.076
expression level	0.231	-0.105	
both other variables except the analyzed one	0.227	-0.095	-0.041

**Figure S6.**

Partial Spearman correlations of gene age with Ka/Ks, connectivity and expression level. Bonferroni adjusted p-values, corresponding to the presented rho values, are color-coded: p-values < 1.0E-30, green; 1.0E-30 > p-values < 1.0E-10, cyan; 1.0E-10 > p-values < 1.0E-3, blue.