

Contrasts between adaptive coding and noncoding changes during human evolution

Ralph Haygood^{1,*}, Courtney C. Babbitt^{1,2}, Olivier Fedrigo^{1,2}, and Gregory A. Wray^{1,2}

¹Biology Department, Duke University, Durham, NC 27708

²Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708

*Corresponding author; Email: ralph.haygood@gmail.com

Proceedings of the National Academy of Sciences of the United States of America **107**:7853–7857 (2010)

[This version contains approximately 300 words, eight references, and one table that were cut from the version appearing in *PNAS*. The accompanying Supplementary Tables are numbered differently and collected in one file.]

Abstract

It has been proposed that changes in non-protein-coding, regulatory DNA sequences play distinctive roles in adaptive evolution. We analyzed correlations between gene functions and evidence for positive selection in a common statistical framework across several large surveys of coding or non-coding sequences throughout the human genome. Strong correlations with both classifications in gene ontologies and measurements of gene expression indicate that neural development and function have adapted mainly through noncoding changes. In contrast, adaptation via coding changes is dominated by immunity, olfaction, and male reproduction. Genes with highly tissue-specific expression have undergone more adaptive coding changes, suggesting that pleiotropic constraints inhibit such changes in broadly expressed genes. In contrast, adaptive noncoding changes do not exhibit this pattern. Our findings underline the probable importance of noncoding changes in the evolution of human traits, particularly cognitive traits.

Among the most fundamental unanswered questions about adaptive evolution are whether it proceeds primarily through changes in protein-coding DNA sequences or noncoding, regulatory sequences, whether the proportions of coding and noncoding changes vary appreciably among organismal traits, and, if so, why (Hoekstra and Coyne, 2007; Wray, 2007). For example, it has been argued that morphological adaptation occurs mainly via noncoding changes, on the grounds that many genes underlying development are active in many contexts, and a noncoding mutation is likely to alter a gene's activity in only one or a few contexts, avoiding pleiotropic constraints (Stern, 2000; Carroll, 2005; Carroll, 2008). These questions have been addressed largely using case studies of individual genes and traits. However, they can now be addressed on a genomic scale, and doing so is indispensable for assessing whether consistent, intelligible patterns exist. The human genome is especially suitable for such inquiry, for two reasons. First, there are now many surveys aiming to detect signatures of positive selection on sequences throughout the human genome (cf. Sabeti et al., 2006 and Kelley and Swanson, 2008 for reviews), most focusing on coding but a few on noncoding sequences. Second, the extensive functional annotations of the human genome often make it possible to infer something about the trait through which positive selection on a sequence arose (Sabeti et al., 2006; Kelley and Swanson, 2008). Here, we evaluate what several large surveys of the human genome imply about the roles of coding and noncoding changes in adaptive evolution, how these roles vary among gene functions, and what this variation suggests about its causes.

We analyzed three surveys of coding (Bustamante et al., 2005; Nielsen et al., 2005; Kosiol et al., 2008) and three of noncoding sequences (Pollard et al., 2006; Prabhakar et al., 2006; Haygood et al., 2007) (cf. "Methods" and Sup. Tab. 1–2). Each survey aims to detect a signature of positive selection, distinguishes selection on coding versus noncoding sequences, and treats thousands of sequences without a priori bias regarding function. (Several other large surveys are unsuitable for our purposes, because they flag regions typically containing both coding and noncoding sequences, or they treat only sequences with specified functions.) The surveys' data are diverse, and even

when they overlap, different surveys sometimes give different results, because their methods have different sensitivities, not only to signatures of positive selection but also to confounding factors (Li et al., 2008; Prabhakar et al., 2008; Prabhakar et al., 2009). This diversity of data and methods is what makes it valuable to consider these surveys collectively. Although each survey offers unique insights, we are interested in trends prevailing across all three coding or all three noncoding surveys, which are likely to represent consistent features of adaptation via coding or noncoding changes.

The trends we focus on are correlations between positive selection and gene functions according to the PANTHER (Mi et al., 2005) and GO (Gene Ontology Consortium, 2000) classifications and the Novartis Gene Expression Atlas (Su et al., 2004). Within each survey, certain functional categories and expression domains are enriched with or depauperate of genes scoring high for positive selection, suggesting that changes detected in the survey have played large or small roles, respectively, in adaptation of these functions. These functional annotations apply directly to proteins and hence coding sequences. In the absence of direct annotations of most noncoding sequences, we associated each noncoding sequence with the nearest coding sequence, as coding sequences are often regulated by nearby noncoding sequences (Wray et al., 2003). (Our analyses avoid potential bias arising from the fact that genes with certain functions tend to be larger or more isolated (Taher and Ovcharenko, 2009); cf. “Methods”.) Each survey was published with some analysis of functional enrichment, and some impression of patterns across surveys may be gained from these analyses (cf. Sabeti et al., 2006 and Kelley and Swanson, 2008). However, a more detailed and precise understanding is attainable by analyzing the surveys in a common statistical framework.

For each PANTHER or GO category, we computed the rank correlation between score for positive selection and membership in the category (rank-biserial correlation r_{rb}) and the standard error of the correlation within each survey (cf. “Methods”). We then computed the weighted mean of this correlation and the standard error of the mean across coding and across noncoding surveys, weighting so that surveys estimating the correlation more precisely contribute more heavily to the

mean. We are particularly interested in categories for which the mean correlation is significantly positive ($p_{\text{enr}} < 0.05$) or negative ($p_{\text{dep}} = 1 - p_{\text{enr}} < 0.05$) across coding or noncoding surveys, indicating that the category is, respectively, appreciably enriched with or depauperate of positive selection on coding or noncoding changes. We also computed a heterogeneity statistic across coding and across noncoding surveys, and when discussing categories enriched with or depauperate of positive selection across coding or noncoding surveys, we restrict attention to categories for which this statistic is nonsignificant across the same surveys ($p_{\text{het}} > 0.05$), indicating a lack of appreciable discord among the surveys. Figure 1 plots results for large PANTHER biological processes enriched across coding or noncoding surveys. Tables 1a and b list results for large and middle-sized PANTHER biological processes enriched across coding and noncoding surveys, respectively, and Supplementary Tables 3a and b are analogous for GO biological processes. The Supplementary Tables contain full results for large and middle-sized PANTHER and GO biological processes, PANTHER and GO molecular functions, and GO cellular components (Sup. Tab. 4–8).

The strongest pattern in these results is that neural development appears to have adapted primarily through noncoding changes. Across noncoding surveys, the PANTHER biological process “neurogenesis” is highly enriched with positive selection ($p_{\text{enr}} < 10^{-6}$). “Neurogenesis” has no subcategories in PANTHER, but finer resolution is available in GO biological processes, where “regulation of neuron differentiation”, “axon guidance”, “regulation of axonogenesis”, “brain development”, “neuron migration”, “positive regulation of neurogenesis”, and “negative regulation of neurogenesis” are enriched across noncoding surveys ($p_{\text{enr}} = 4.5 \times 10^{-5}$, 5.2×10^{-4} , 0.0024, 0.0026, 0.0072, 0.0079, and 0.014), whereas “axon guidance” and “negative regulation of neurogenesis” are depauperate across coding surveys ($p_{\text{dep}} = 0.021$ and $< 10^{-6}$). This pattern arises largely from different genes in different surveys. Of the 86 “neurogenesis” genes scoring high ($p < 0.05$) for positive selection in at least one noncoding survey, only eight score high in two surveys, and none do so in all three. (Note that these surveys generally treat different sequences associated with a given gene.)

Likewise, neural function appears to have adapted primarily through noncoding changes, although this pattern is weaker. The PANTHER biological process “other neuronal activity” is enriched with positive selection across noncoding surveys ($p_{\text{enr}} = 8.9 \times 10^{-4}$) and remains so when genes also in “neurogenesis” are excluded ($p_{\text{enr}} = 0.018$). (However, the distinction between neural development and function is fuzzy, such that genes in “other neuronal activity” but not “neurogenesis” may nonetheless influence neural morphology; examples scoring high in at least one noncoding survey include *CAMK1G* (Nopoulos et al., 2002) and *HTT* (huntingtin) (Paulsen et al., 2006).) The GO biological process “regulation of synaptic transmission” is enriched across noncoding surveys ($p_{\text{enr}} = 0.0094$), the GO molecular functions “GABA receptor activity” and “ionotropic glutamate receptor activity” are enriched across noncoding surveys ($p_{\text{enr}} = 0.027$ and 0.045) and highly depauperate across coding surveys ($p_{\text{dep}} < 10^{-6}$ for both), and the GO cellular component “synapse” is marginally enriched across noncoding surveys ($p_{\text{enr}} = 0.075$) and depauperate across coding surveys ($p_{\text{dep}} = 1.6 \times 10^{-4}$). No such category is enriched across coding surveys, apart from the olfaction-related categories mentioned below.

Other trends prevailing across noncoding surveys include enrichment with positive selection of several aspects of development in addition to neural development. Most conspicuously, the PANTHER biological process “muscle development” is enriched ($p_{\text{enr}} = 0.020$). Moreover, enriched categories such as the PANTHER biological process “oncogene” ($p_{\text{enr}} = 0.0017$) and the GO biological process “homophilic cell adhesion” ($p_{\text{enr}} = 1.1 \times 10^{-5}$) include high-scoring genes with developmental functions beyond neural or muscle development. However, there are also developmental categories enriched across coding surveys, such as the PANTHER biological processes “anterior/posterior patterning” and “segment specification” ($p_{\text{enr}} = 0.029$ and 0.034).

Consistent with earlier reports (Bustamante et al., 2005; Nielsen et al., 2005; Sabeti et al., 2006; Kosiol et al., 2008; Kelley and Swanson, 2008), the leading themes of adaptation through coding changes appear to be immunity and olfaction, represented by the PANTHER biological processes “immunity and defense” and “chemosensory perception” ($p_{\text{enr}} < 10^{-6}$ for both), the

GO biological processes “defense response” and “sensory perception of smell” ($p_{\text{enr}} = 0.0011$ and $< 10^{-6}$), and many other categories. Also conspicuous is sperm function, represented by the PANTHER biological process “spermatogenesis and motility” ($p_{\text{enr}} = 0.039$) and the GO cellular component “acrosome” ($p_{\text{enr}} = 0.0087$), among other categories. However, few such categories are depauperate of positive selection across noncoding surveys. Indeed, the PANTHER biological process “T-cell mediated immunity” is enriched with positive selection across both coding and noncoding surveys ($p_{\text{enr}} = 4.9 \times 10^{-5}$ and 0.016). In conjunction with the distinctive role of noncoding changes in neural adaptation, these results raise the possibility that a wider range of traits have been amenable to adaptation via noncoding than coding changes.

To see how adaptive coding and noncoding changes relate to gene expression, we turned to the Novartis Gene Expression Atlas. For each gene and each of the 73 non-cancerous tissues in the atlas, we computed a specificity score between 0 and 1 representing the specificity of the gene’s expression to the tissue (Haygood et al., 2007; Kosiol et al., 2008). The score is high if the expression is very specific to the tissue, and it is low even for the gene’s tissue of maximal expression if the gene is nearly as highly expressed in other tissues. For each gene, we also computed an evenness score between 0 and 1 representing the evenness of the gene’s expression across all 73 tissues. The score is high if the expression is not very specific to any tissue. A gene’s specificity and evenness scores are independent of the overall magnitude of its expression. Figure 2 illustrates these scores in the simplest context of two tissues.

For each tissue, we computed the rank correlation between score for positive selection and specificity to the tissue ($r_{r,s}$) and its standard error within each survey. We then computed the weighted mean of this correlation, its standard error, and the heterogeneity statistic across coding and across noncoding surveys. Figure 3 plots the results for all tissues and highlights two groups of tissues exhibiting strong contrasts (cf. Sup. Tab. 9). Consonant with the results presented above, specificity to neural tissues (mainly components of brain plus ganglia) is more correlated with adaptive noncoding than coding changes (Wilcoxon signed-rank test $p < 10^{-6}$), whereas

specificity to male reproductive tissues (components of testis plus prostate gland) is marginally the opposite ($p = 0.063$). These contrasts prevail almost without exception, as mean $r_{r,s}$ is greater across noncoding than coding surveys for all but one (dorsal root ganglion) of the 23 neural tissues, whereas the opposite holds for all but one (prostate gland) of the six male tissues.

Similarly, we computed the rank correlation between score for positive selection and evenness across all tissues ($r_{r,e}$) within each survey, the weighted mean of this correlation across coding and across noncoding surveys, and the associated standard errors and heterogeneity statistic. We repeated this computation restricting attention to the 25%, 10%, or 5% most and least evenly expressed genes per survey. Table 2 lists the results (cf. Sup. Tab. 10). As the evenness tail fraction decreases, mean $r_{r,e}$ across coding surveys becomes more negative, becoming significantly so when the fraction is 10% or 5% ($p_{\text{dep}} = 0.0057$ and 0.0076). Thus, very evenly expressed genes have experienced less positive selection on their coding sequences than very unevenly expressed genes, a phenomenon noted in Kosiol et al.'s (2008) survey. This pattern persists when only genes in significantly enriched PANTHER and GO categories are included in the computations (Sup. Tab. 10). No such pattern is discernible across noncoding surveys.

All the patterns we have mentioned persist when any one of the six surveys is excluded (Sup. Tab. 11–17). There are, of course, potential sources of noise or error, both in the surveys and in our analyses of them. For example, some sequences exhibiting accelerated evolution in the human lineage may do so due to nonadaptive processes (e.g., biased gene conversion (Pollard et al., 2006; Berglund et al., 2009)), and many genes have multiple functions or functions missing from current annotations. Moreover, human-specific gene duplications and deletions are beyond the scope of the surveys analyzed here but may have contributed much to human adaptation (Olson and Varki, 2003; Han et al., 2009). Assuming, however, that the strong contrasts we have identified are genuinely characteristic of adaptive evolution, at least in humans, what might account for them?

A simple possibility is that genes with functions enriched with positive selection across coding or noncoding surveys are associated with longer coding or noncoding sequences, respectively. All

else being equal, longer sequences should undergo larger numbers of adaptive changes. Length variation is potentially relevant to all the surveys analyzed here except perhaps Haygood et al.'s (2007), which treats approximately 5 kb of noncoding sequence per gene. Consistent with this idea, genes in noncoding-enriched PANTHER and GO categories are associated with 185 ± 5 kb of noncoding sequence (introns, UTRs, and half of flanking regions) on average versus 157 ± 5 kb for genes in coding-enriched categories. Much of this sequence may be nonfunctional, but Pollard et al.'s (2006) and Prabhakar et al.'s (2006) surveys treat noncoding sequences conserved across nonhuman species and hence presumptively functional, and genes in noncoding-enriched categories are associated with 1058 ± 50 bp and 4367 ± 145 bp of Pollard et al.'s and Prabhakar et al.'s sequences, respectively, on average versus 937 ± 67 bp and 3732 ± 180 bp for genes in coding-enriched categories. These differences suggest that some functions may have adapted more through noncoding than coding changes partly because they are affected by more noncoding sequence per gene, which constitutes a larger "target" for mutations. However, length variation is clearly not the only relevant factor. Some categories associated with longer noncoding sequences are not enriched with positive selection across noncoding surveys, even excluding Haygood et al.'s (e.g., the GO biological processes "cyclic-nucleotide-mediated signaling" and "positive regulation of developmental process" (Taher and Ovcharenko, 2009)). Moreover, genes in noncoding-enriched categories also have longer coding sequences, 1974 ± 22 bp on average, than genes in coding-enriched categories, 1909 ± 31 bp.

As mentioned above, it has been argued that for genes active in many contexts, noncoding changes are more likely than coding changes to be adaptive, because a noncoding mutation is more likely to enhance a gene's function in one context without degrading it in others, given that gene expression in different contexts is often governed by distinct noncoding sequences (Stern, 2000; Wray et al., 2003; Carroll, 2005; Carroll, 2008). It follows that if most genes with certain functions are active in many contexts, adaptation of these functions should occur mainly via noncoding changes. Although expression in many tissues need not entail conflicting demands in different con-

texts, our results regarding evenness offer modest support for these ideas. Very even genes, many of which presumably play important roles in many contexts, have tended to undergo fewer adaptive coding changes than very uneven genes, most of which presumably play important roles in fewer contexts. This pattern does not appear to hold for adaptive noncoding changes. (Conceivably, we lack statistical power to detect the pattern across noncoding surveys. However, it is detectable across coding surveys even when only the smaller set of genes in significantly enriched PANTHER and GO categories is analyzed; cf. Sup. Tab. 10.) Moreover, genes in noncoding-enriched PANTHER and GO categories have higher evenness scores, 0.62 ± 0.0026 on average, than genes in coding-enriched categories, 0.59 ± 0.0037 , although the difference is small.

It has also been argued that selection is typically more efficient on noncoding than coding mutations, because coding mutations are typically recessive, whereas noncoding mutations are typically co-dominant with respect to gene expression, although it is unclear whether they are also typically co-dominant with respect to organismal traits (Wray, 2007). This differential should be diminished for mutations on the X chromosome, which is hemizygous in males, suggesting that the X might be enriched with positive selection on coding but not noncoding changes. The evidence regarding these ideas from the surveys analyzed here is mixed (Sup. Tab. 18). Both across coding surveys and across noncoding surveys, the X is significantly heterogeneous ($p_{\text{het}} = 4.9 \times 10^{-4}$ and 5.0×10^{-4}), indicating appreciable discord among the surveys. The X is enriched in two of the three coding surveys (Bustamante et al., 2005; Nielsen et al., 2005) but also in one of the three noncoding surveys (Prabhakar et al., 2006). Further complicating interpretation is the existence of factors in addition to male hemizyosity potentially causing unusual patterns of evolution on the X (Vicoso and Charlesworth, 2006).

A full understanding of which traits adapt mainly via coding versus noncoding changes and why must await the maturity of functional genomics. In particular, far more functional annotations of noncoding sequences are needed. Nonetheless, our current findings are important. At a minimum, they strongly suggest that both coding and noncoding changes have played important and

distinctive roles in human adaptation. One implication is that studying adaptive coding changes alone, which have long been a major focus of research in evolutionary genetics, can yield an incomplete and unbalanced picture of adaptive evolution, which can be significantly extended and enriched by studies of adaptive noncoding changes. More such studies in nonhuman species are needed to reveal which of the patterns we have found in humans also exist in other species.

The finding that neural adaptation has occurred mainly via noncoding changes is particularly important, in view of the remarkable cognitive innovations in the human lineage. It is consistent with the hypothesis advanced thirty-five years ago that the major phenotypic differences between humans and chimpanzees reflect changes in gene regulation rather than protein structure (King and Wilson, 1975). Noncoding sequences flagged by the surveys analyzed here and associated with neural development and function are excellent candidates for research on the genetics and evolution of human cognition.

Methods

Surveys and scores: To analyze the noncoding surveys (Pollard et al., 2006; Prabhakar et al., 2006; Haygood et al., 2007), we associated each noncoding sequence treated in these surveys with the nearest coding sequence in the UCSC Known Genes collection (Karolchik et al., 2003). (Pollard et al.'s (2006) survey includes sequences overlapping known coding sequences, which we excluded from our analyses.) For Pollard et al.'s (2006) and Prabhakar et al.'s (2006) surveys, some genes were associated with multiple sequences, in which case we combined Pollard et al.'s or Prabhakar et al.'s p -values for the sequences into a p -value for the gene using Simes' method (Simes, 1986): the p -value of the gene is the minimum Benjamini–Hochberg-adjusted p -value of the sequences. This is important, in that although these surveys have no a priori bias regarding function, genes with certain functions tend to be larger or more isolated and hence associated with more sequences treated in these surveys (Taher and Ovcharenko, 2009). If this phenomenon were not accounted for when converting sequence scores into gene scores, such functions would tend to be enriched with high-scoring genes purely by chance, in the absence of positive selection or

any other evolution-accelerating process. Simes' method avoids this potential bias, because the more sequences are associated with a gene, the more their p -values are discounted by Benjamini–Hochberg adjustment.

No sequence-based test for positive selection is perfectly reliable. In addition to false positives occurring purely by chance, nonadaptive processes can mimic signatures of positive selection (Pollard et al., 2006; Li et al., 2008; Prabhakar et al., 2008; Berglund et al., 2009; Prabhakar et al., 2009). An important motivation for our analyses is that to a substantial extent, different tests are sensitive to different confounding factors (Li et al., 2008; Prabhakar et al., 2008; Prabhakar et al., 2009). For example, Pollard et al.'s (2006) survey detects accelerated evolution in the human lineage relative to several other lineages, assessing relative acceleration in a region of interest with respect to a set of reference regions. Such acceleration might result from relaxation of negative selection or biased gene conversion (BGC), as Pollard et al. discuss (Pollard et al., 2006; Berglund et al., 2009). In contrast, Prabhakar et al.'s (2006) and Haygood et al.'s (2007) surveys assess acceleration with respect to estimates of the local rate of neutral evolution, which is unlikely to reflect relaxation of negative selection or BGC (Prabhakar et al., 2006; Haygood et al., 2007; Prabhakar et al., 2008; Prabhakar et al., 2009). Indeed, Haygood et al.'s null model explicitly accommodates relaxation of negative selection, and Haygood et al. found no evidence that BGC influences their results (Haygood et al., 2007, text supplement, p. 8).

Four of the six surveys are sensitive to positive selection in the human lineage alone, but Bustamante et al.'s (2005) and Nielsen et al.'s (2005) surveys are sensitive to positive selection in the chimpanzee lineage too. However, of the patterns discussed here, only enrichment of the X chromosome is supported by these two but not Kosiol et al.'s (2008) coding survey.

See Supplementary Tables 1 and 2 for more information about the surveys.

Functional categories and expression domains: In each survey, we mapped as many genes as we could first to UniProt (UniProt Consortium, 2009; ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/taxonomic_divisions/uniprot_sprot_human.dat and uniprot_trembl_

human.dat; downloaded 11/25/08) identifiers and then to PANTHER (Mi et al., 2005; ftp://ftp.pantherdb.org/sequence_classifications/current_release; updated 02/22/08) and GO (Gene Ontology Consortium, 2000; http://cvswb.geneontology.org/cgi-bin/cvswb.cgi/go/gene-associations/gene_association.goa.human.gz?rev=HEAD; submitted 09/18/08) categories (excluding placeholder categories, e.g., “biological process unclassified”). (We do not present results for PANTHER pathways, because relatively few genes have been classified into them.) For each category, we computed the rank-biserial correlation r_{rb} between score for positive selection and membership in the category. These computations involved only genes that were treated in the given survey and that we were able to map to at least one category of the relevant kind. r_{rb} measures association between an ordinal variable and a dichotomous variable (Kraemer, 1982); it is proportional to the standard (Pearson) correlation coefficient between the ranks of the ordinal variable and any two values, say, 0 and 1, for the dichotomous variable. We treated score for positive selection as an ordinal rather than continuous variable to avoid issues that might arise from the diversity of scoring functions among the surveys. r_{rb} is a linear function of the Mann–Whitney U statistic, and a test of $r_{rb} = 0$ is equivalent to a Mann–Whitney test (Bustamante et al., 2005; Nielsen et al., 2005; Haygood et al., 2007; Kosiol et al., 2008). For each category, we estimated the standard error of r_{rb} as the standard deviation of r_{rb} over 1000 bootstrap replicates per survey. (Bootstrapping is appropriate because the sample distributions of correlations involving dichotomous variables need not be nearly normal (Harris and Kolen, 1988).)

Similarly, in each survey, we mapped as many genes as we could to Novartis (Su et al., 2004; http://symatlas.gnf.org/suppl.html#reqdata_geneatlas) probes. We took means over multiple arrays per tissue and maxima over multiple probes per gene to obtain the expression of a gene in a tissue. For each gene, we computed specificity and evenness scores as illustrated in Figure 2 but in the context of all 73 non-cancerous tissues in the atlas. For each tissue, we computed the rank (Spearman) correlation $r_{r,s}$ between score for positive selection and specificity to the tissue. We also computed the rank correlation $r_{r,e}$ between score for positive selection and evenness across all

tissues. We estimated the standard errors of $r_{r,s}$ and $r_{r,e}$ from 1000 bootstrap replicates per survey.

Fixed-effects meta-analyses: Our analyses of patterns across surveys are essentially standard fixed-effects meta-analyses (Hedges, 1994). Given a category and a set of surveys, the weighted mean of r_{rb} is $\sum_i w_i (r_{rb})_i$, where $w_i = (1/\text{stderr}(r_{rb})_i)^2 / \sum_j (1/\text{stderr}(r_{rb})_j)^2$. It follows that the standard error of the mean is $1/(\sum_i (1/\text{stderr}(r_{rb})_i)^2)^{1/2}$. We tested mean $r_{rb} = 0$ by comparing $(\text{mean } r_{rb})/(\text{stderr mean } r_{rb})$ to the standard normal distribution; we label the upper and lower one-tailed p -values p_{enr} and p_{dep} , respectively. (To check the validity of these z -test p -values, we also tested mean $r_{rb} = 0$ by resampling each $(r_{rb})_i$ from the bootstrap replicates used to estimate its standard error and recording the sign of the resulting mean r_{rb} , obtaining generally good agreement with the z -test p -values, particularly for large categories.) Under the null hypothesis $(r_{rb})_i = (r_{rb})_j$ for every i and j , the heterogeneity statistic $\sum_i w_i ((r_{rb})_i - \text{mean } r_{rb})^2$ has approximately the chi-squared distribution with two degrees of freedom, assuming three surveys; we label the chi-squared p -value p_{het} . Analogous formulas and procedures apply to mean $r_{r,s}$ and mean $r_{r,e}$.

Random-effects meta-analyses: Fixed-effects meta-analyses are sometimes criticized on the grounds that homogeneity across studies is unrealistic, and p_{het} is an unsatisfactory indicator (Field, 2001; Hunter and Schmidt, 2004). Supplementary Tables 19–25 parallel Supplementary Tables 4–10 and contain results from a random-effects approach (Field, 2001; Hunter and Schmidt, 2004). Here, mean r_{rb} is as above but with $w_i = n_i / \sum_j n_j$, where n_i is the number of categorized genes in the i th survey. More importantly, r_{rb} is regarded as a random variable over surveys, the standard error of the mean is $(\sum_i n_i ((r_{rb})_i - \text{mean } r_{rb})^2 / 3 \sum_i n_i)^{1/2}$, and the test of mean $r_{rb} = 0$ is a t -test with two degrees of freedom, assuming three surveys. Again, analogous formulas and procedures apply to mean $r_{r,s}$ and mean $r_{r,e}$.

Although the fixed- and random-effects results differ in detail, they display the same major patterns. For example, according to both sets of results, the PANTHER biological processes “neurogenesis”, “other neuronal activity”, and “muscle development” are enriched with positive selection across noncoding surveys, “immunity and defense”, “chemosensory perception”, and “sper-

matogenesis and motility” are enriched across coding surveys, and “T-cell mediated immunity” is enriched across both.

Software: Our software, consisting of approximately 2700 lines of Ruby and 300 lines of C, is available upon request.

References

- Berglund, J., Pollard, K. S., and Webster, M. T., 2009. Hotspots of biased nucleotide substitutions in human genes. *PLoS Biology* **7**:45–62.
- Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S., Tanenbaum, D. M., White, T. J., Sninsky, J. J., Hernandez, R. D., Civello, D., Adams, M. D., Cargill, M., and Clark, A. G., 2005. Natural selection on protein-coding genes in the human genome. *Nature* **437**:1153–1157.
- Carroll, S. B., 2005. Evolution at two levels: On genes and form. *PLoS Biology* **3**:1159–1166.
- Carroll, S. B., 2008. Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**:25–36.
- Field, A. P., 2001. Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods* **6**:161–180.
- Gene Ontology Consortium, 2000. Gene Ontology: Tool for the unification of biology. *Nature Genetics* **25**:25–29.
- Han, M. V., Demuth, J. P., McGrath, C. L., Casola, C., and Hahn, M. W., 2009. Adaptive evolution of young gene duplicates in mammals. *Genome Research* **19**:859–867.
- Harris, D. J., and Kolen, M. J., 1988. Bootstrap and traditional standard errors of the point-biserial. *Educational and Psychological Measurement* **48**:43–51.
- Haygood, R., Fedrigo, O., Hanson, B., Yokoyama, K.-D., and Wray, G. A., 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nature Genetics* **39**:1140–1144.

- Hedges, L. V., 1994. Fixed effects models. Cooper, H., and Hedges, L. V. (Eds.), *The Handbook of Research Synthesis*, pp. 285–299. Russell Sage Foundation, New York, NY.
- Hoekstra, H. E., and Coyne, J. A., 2007. The locus of evolution: Evo devo and the genetics of adaptation. *Evolution* **61**:995–1016.
- Hunter, J. E., and Schmidt, F. L., 2004. *Methods of meta-analysis*, 2nd ed. Sage Publications, Thousand Oaks, CA.
- Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., Weber, R. J., Haussler, D., and Kent, W. J., 2003. The UCSC genome browser database. *Nucleic Acids Research* **31**:51–54.
- Kelley, J. L., and Swanson, W. J., 2008. Positive selection in the human genome: From genome scans to biological significance. *Annual Review of Genomics and Human Genetics* **9**:143–160.
- King, M.-C., and Wilson, A. C., 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**:107–116.
- Kosiol, C., Vinař, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R., and Siepel, A., 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genetics* **4**:1–17.
- Kraemer, H. C., 1982. Biserial correlation. Kotz, S., and Johnson, N. L. (Eds.), *Encyclopedia of Statistical Sciences*, pp. 276–280. John Wiley and Sons, New York, NY.
- Li, Y. F., Costello, J. C., Holloway, A. K., and Hahn, M. W., 2008. “Reverse ecology” and the power of population genomics. *Evolution* **62**:2984–2994.
- Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Murganujan, A., Doremieux, O., Campbell, M. J., Kitano, H., and Thomas, P. D., 2005. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Research* **33**:D284–D288.

- Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., White, T. J., Sninsky, J. J., Adams, M. D., and Cargill, M., 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* **3**:0976–0985.
- Nopoulos, P., Berg, S., Canady, J., Richman, L., Van Demark, D., and Andreasen, N. C., 2002. Structural brain abnormalities in adult males with clefts of the lip and/or palate. *Genetics in Medicine* **4**:1–9.
- Olson, M. V., and Varki, A., 2003. Sequencing the chimpanzee genome: Insights into human evolution and disease. *Nature Reviews Genetics* **4**:20–28.
- Paulsen, J. S., Magnotta, V. A., Mikos, A. E., Paulson, H. L., Penziner, E., Andreasen, N. C., and Nopoulos, P. C., 2006. Brain structure in preclinical Huntington’s disease. *Biological Psychiatry* **59**:57–63.
- Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J. S., Bejerano, G., Baertsch, R., Rosenbloom, K. R., Kent, J., and Haussler, D., 2006. Forces shaping the fastest evolving regions in the human genome. *PLoS Genetics* **2**:1599–1611.
- Prabhakar, S., Noonan, J. P., Pääbo, S., and Rubin, E. M., 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**:786–786.
- Prabhakar, S., Visel, A., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., Plajzer-Frick, I., Morrison, H., FitzPatrick, D. R., Afzal, V., Pennacchio, L. A., Rubin, E. M., and Noonan, J. P., 2008. Human-specific gain of function in a developmental enhancer. *Science* **321**:1346–1350.
- Prabhakar, S., Visel, A., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., Plajzer-Frick, I., Morrison, H., FitzPatrick, D. R., Afzal, V., Pennacchio, L. A., Rubin, E. M., and Noonan, J. P., 2009. Response to comment on “Human-specific gain of function in a developmental enhancer”. *Science* **323**:714.

- Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D., and Lander, E. S., 2006. Positive natural selection in the human lineage. *Science* **312**:1614–1620.
- Simes, R. J., 1986. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**:751–754.
- Stern, D. L., 2000. Evolutionary developmental biology and the problem of variation. *Evolution* **54**:1079–1091.
- Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R., and Hogenesch, J. B., 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**:6062–6067.
- Taher, L., and Ovcharenko, I., 2009. Variable locus length in the human genome leads to ascertainment bias in functional inference for noncoding elements. *Bioinformatics* **25**:578–584.
- UniProt Consortium, 2009. The universal protein resource (UniProt). *Nucleic Acids Research* **37**:D169–D174.
- Vicoso, B., and Charlesworth, B., 2006. Evolution on the X chromosome: Unusual patterns and processes. *Nature Reviews Genetics* **7**:645–653.
- Wray, G. A., 2007. The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics* **8**:206–216.
- Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A., 2003. The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution* **20**:1377–1419.

Table 1: PANTHER biological processes enriched with positive selection

a: Across coding surveys

category ¹	mean n^2	mean r_{rb}	stderr mean r_{rb}	p_{enr} ³
chemosensory perception	68	0.37	0.054	$< 10^{-6}$
immunity and defense ⁴	550	0.059	0.011	$< 10^{-6}$
T-cell mediated immunity ⁴	83	0.12	0.030	4.9×10^{-5}
interferon-mediated immunity ⁴	30	0.16	0.061	0.0037
other oncogenesis	23	0.15	0.058	0.0051
anterior/posterior patterning	29	0.12	0.064	0.029
segment specification	48	0.079	0.043	0.034
steroid metabolism	78	0.049	0.027	0.038
spermatogenesis and motility	50	0.068	0.039	0.039
induction of apoptosis	73	0.049	0.028	0.042

b: Across noncoding surveys

category ¹	mean n^2	mean r_{rb}	stderr mean r_{rb}	p_{enr} ³
neurogenesis	266	0.13	0.020	$< 10^{-6}$
other neuronal activity	54	0.14	0.044	8.9×10^{-4}
T-cell mediated immunity	42	0.11	0.051	0.016
oncogene	39	0.12	0.055	0.0017
muscle development	59	0.085	0.041	0.020
blood clotting	23	0.12	0.068	0.042
sulfur metabolism	31	0.11	0.064	0.047

¹Listed categories satisfy (1) each of at least five surveys treats at least 10 genes, (2) mean rank-biserial correlation (r_{rb}) between score for positive selection and membership in the category is significantly positive ($p_{enr} < 0.05$) across (a) coding or (b) noncoding surveys, and (3) heterogeneity is nonsignificant ($p_{het} > 0.05$) across the same surveys. When one such category contains others, the former is listed only if it still satisfies (1)–(3) when the latter are subtracted. 163 categories satisfy (1).

²Mean number of genes in the category per survey, rounded to nearest integer.

³Upper one-tailed p -value for one-sample z test, rounded upward; null hypothesis is mean $r_{\text{rb}} = 0$. p_{enr} is not adjusted for multiple comparisons, which is difficult to do correctly because categories overlap extensively, but $p_{\text{enr}} < 3.0 \times 10^{-4}$ would remain significant even under Bonferroni adjustment, which is extremely conservative.

⁴“T-cell mediated immunity” and “interferon-mediated immunity” are contained in “immunity and defense”, but the latter minus the former remains significantly enriched ($p_{\text{enr}} = 1.8 \times 10^{-4}$) across coding surveys.

Table 2: Expression evenness and positive selection

a: Across coding surveys

evenness tail fraction ¹	mean n^2	mean $r_{r,e}$	stderr mean $r_{r,e}$	p_{dep}^3
50%	7375	0.0095	0.0068	0.92
25%	3687	-0.0045	0.0098	0.33
10%	1475	-0.040	0.016	0.0057
5%	738	-0.053	0.022	0.0076

b: Across noncoding surveys

evenness tail fraction ¹	mean n^2	mean $r_{r,e}$	stderr mean $r_{r,e}$	p_{dep}^3
50%	4771	0.0078	0.0084	0.83
25%	2385	0.022	0.012	0.97
10%	954	0.014	0.019	0.77
5%	477	-0.013	0.027	0.32

¹Percentage of most and least even genes per survey included in computation of mean rank correlation ($r_{r,e}$) between score for positive selection and evenness across tissues in the Novartis Gene Expression Atlas. For example, 5% means 5% most even and 5% least even genes included. 50% amounts to all genes. For every percentage, heterogeneity is nonsignificant ($p_{\text{het}} > 0.05$) across coding and across noncoding surveys.

²Mean number of genes per survey included in computation, rounded to nearest integer.

³Lower one-tailed p -value for one-sample z test, rounded upward; null hypothesis is mean $r_{r,e} = 0$.

Figure legends

Figure 1: Large PANTHER biological processes enriched with positive selection across coding or noncoding surveys. Plotted categories are all but three of those for which (1) mean number of genes per survey is at least 50 across all surveys, (2) mean rank-biserial correlation (r_{rb}) between score for positive selection and membership in the category is significantly positive ($p_{\text{enr}} < 0.05$) across coding or noncoding surveys, and (3) heterogeneity is nonsignificant ($p_{\text{het}} > 0.05$) across the same surveys. The three such categories not plotted are “ectoderm development”, “neuronal activities”, and “mesoderm development”, which are not significantly enriched when their respective subcategories “neurogenesis”, “other neuronal activity”, and “muscle development” are subtracted. Error bars represent standard error of mean r_{rb} . Green blocks on error bars indicate one-tailed p -value for one-sample z test of mean r_{rb} : one block— $5 \times 10^{-4} < p < 0.05$; two blocks— $5 \times 10^{-6} < p < 5 \times 10^{-4}$; three blocks— $p < 5 \times 10^{-6}$; null hypothesis is mean $r_{rb} = 0$. Similarly, green dots on category names indicate two-tailed p -value for two-sample z test of mean r_{rb} across coding vs. noncoding surveys; null hypothesis is equality.

Figure 2: Illustration of specificity and evenness scores. The vector represents expression of a given gene in two tissues. The dashed line corresponds to equal expression in these tissues. The specificity scores of the gene are $s_1 = \cos^2\sigma_1$ and $s_2 = \cos^2\sigma_2$. The gene’s expression is higher in tissue 1, so $s_1 > s_2$. The evenness score of the gene is $e = \cos^2\varepsilon$. If the gene’s expression were lower in tissue 1 or higher in tissue 2, e would be greater.

Figure 3: Expression specificity and positive selection across coding and noncoding surveys. Neural tissues and male (reproductive) tissues are highlighted. Circles represent mean rank correlation ($r_{r,s}$) between score for positive selection and specificity to tissues in the Novartis Gene Expression Atlas. Horizontal lines indicate median of means. Green dots on group names indicate two-tailed p -value for Wilcoxon signed-rank test of means across coding vs. noncoding surveys: one dot— $5 \times 10^{-4} < p < 0.05$; three dots— $p < 5 \times 10^{-6}$; null hypothesis is equality of medians.

Figure 1

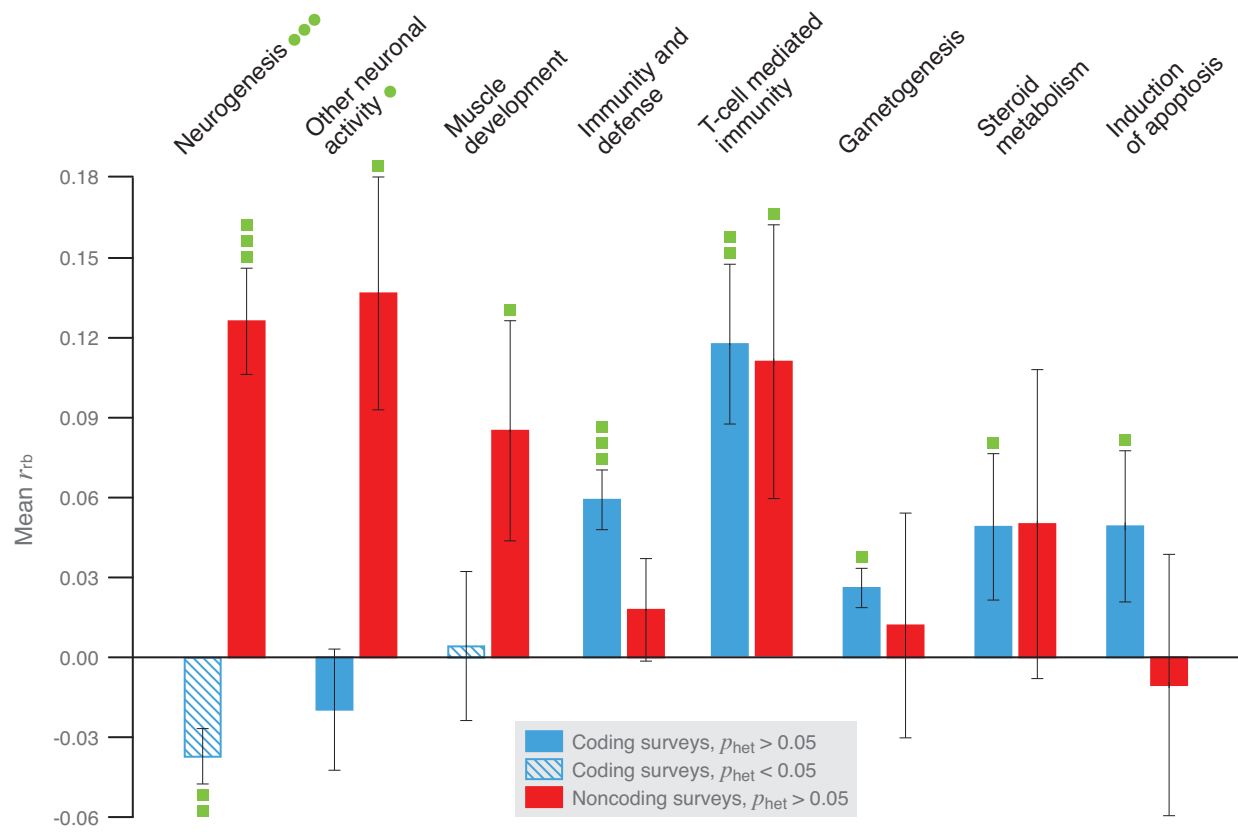


Figure 2

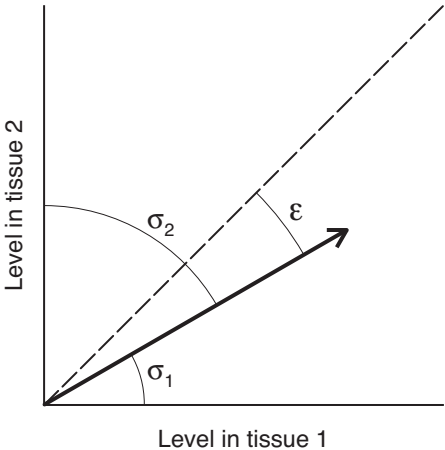


Figure 3

