Position-dependent Codon Usage Bias in the Human Transcriptome

by
Kaavya Subramanian

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Computer Science
(Honors Scholar)

Presented April 29, 2022
Commencement June 2022

AN ABSTRACT OF THE THESIS OF

Kaavya Subramanian for the degree of Honors Baccalaureate of Science in Computer Science presented on April 29, 2022. Title: Position-dependent Codon Usage Bias in the Human Transcriptome.

Abstract approved:____

David Hendrix

All life depends on the reliable translation of RNA to protein according to complex interactions between translation machinery and RNA sequence features. While ribosomal occupancy and codon frequencies vary across coding regions, well-established metrics for computing coding potential of RNA do not capture such positional dependence. Here, we investigate *position-dependent codon usage bias (PDCUB)*, which dynamically accounts for the position of protein-coding signals embedded within coding regions. We demonstrate the existence of PDCUB in the human transcriptome, and show that it can be used to predict translation-initiating codons with greater accuracy than other models. We further show that observed PDCUB is not accounted for by other common metrics, including position-dependent GC content, consensus sequences, and the presence of signal peptides in the translation product. More importantly, PDCUB defines a spectrum of translational efficiency supported by ribosomal occupancy and tRNA adaptation index (tAI). High PDCUB scores correspond to a tAI-defined translational ramp and low ribosomal occupancy, while low PDCUB scores exhibit a translational valley and the highest ribosomal occupancy. Finally, we examine the relationship between PDCUB intensity and functional enrichment. We find that transcripts with start codons showing the highest PDCUB are enriched for functions relating to development of musculoskeletal, cardiovascular, neurological, gastrointestinal, sensory, and other body systems. Furthermore, transcripts with high PDCUB are depleted for functions related to immune response and detection of chemical stimulus. These findings lay important groundwork to improve our understanding of the regulation of translation, the calculation of coding potential, and the classification of RNA transcripts.

Keywords: Codon usage bias, translation, start codon prediction, tRNA adaptation index, Ribo-seq

Corresponding e-mail address: subramka@oregonstate.edu

Position-dependent Codon Usage Bias in the Human Transcriptome

by
Kaavya Subramanian

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Computer Science
(Honors Scholar)

Presented April 29, 2022
Commencement June 2022

Honors Baccalaureate of Science in Computer Science project of Kaavya Subramanian presented on April 29, 2022

APPROVED:

David Hendrix, Mentor, representing Biochemistry and Biophysics

Stephen Ramsey, Committee Member, representing Electrical Engineering and Computer Science

Maude David, Committee Member, representing Microbiology

Toni Doolen, Dean, Oregon State University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, Honors College. My signature below authorizes release of my project to any reader upon request.

Kaavya Subramanian, Author

CONTRIBUTION OF AUTHORS

This project was done in collaboration with Dr. David Hendrix, Masters student Nathan Waugh, and undergraduate student Cole Shanks.

INTRODUCTION

Translation from RNA to protein is a fundamental and ubiquitous life process. Some studies estimate that $\approx 68\%$ of human genes do not encode proteins, and are transcribed as long noncoding RNAs (lncRNAs) or transcripts of unknown coding potential [1]. This raises the question of how the ribosome distinguishes mRNAs from lncRNAs that have open reading frames (ORFs) [2]. At the same time, other studies suggest that some lncRNAs are weakly translated [3-5]. These observations underscore the importance of understanding how the ribosome uses sequence features to distinguish mRNAs from lncRNAs, as well as to properly identify the start codon within an mRNA. Such sequence features could help explain the translation of short transcripts that encode small peptides [6, 7], and improve the design of mRNA vaccines [8].

The Kozak consensus sequence is one such motif, discovered by Marilyn Kozak in the 1980s [9]. This sequence characterizes and helps identify start codons in eukaryotic mRNAs and is regarded as a key feature that enhances translational potential. Another approach to quantifying mRNA coding potential is codon usage bias (CUB), which refers to the frequency of usage for each codon in the coding portion of a transcriptome relative to the frequency of synonymous codons. Computational models of CUB have been shown to perform well in predicting translational efficiency and also show an association with ribosome profiling data [10]. The codon adaptation index (CAI) is a widely recognized metric of CUB that assigns a score to each transcript based on its length and codon composition relative to overall CUB across the coding transcriptome [11]. The CAI model has been used as a baseline for other quantitative models of CUB, and has also been correlated with gene expression in select genomes [11].

A potential limitation of these approaches is that they do not consider the role of codon position or transcript length explicitly. Recently, studies of position-dependent codon usage bias (PDCUB) observed that codon usage is non-uniform with regard to transcript position in *E. coli [12]* and yeast [13]. Moreover, our own investigations demonstrated that a recurrent neural network, which we trained to distinguish mRNAs from lncRNAs based on sequence alone, was able to learn sequence-specific rules and make classification decisions approximately 100-200 nucleotides (nt) downstream of the start codon [14].

Here, we investigate the extent of PDCUB within the human coding transcriptome and whether PDCUB can function as a predictor of start codon location. We present a codon bias score that incorporates these position-dependent codon patterns, using only information observed in the first 300 nt downstream of a given AUG. We show that our score is capable of distinguishing start codons from non-start AUGs with greater accuracy than other methods. While our PDCUB score was constructed from general positional trends in codon usage, we show that it is correlated with translational efficiency data. We show that the highest PDCUB transcripts are associated with a translational ramp and low ribosomal occupancy, while the lowest are associated with a translational valley and high ribosomal occupancy. Moreover, we show that high- and low-PDCUB transcripts display a clear division in biological function, respectively corresponding to a need for high and low translational efficiency.

RESULTS

Visualization of PDCUB

We first performed a coarse-grained investigation of PDCUB in the human transcriptome using GENCODE data [15]. The frequencies of all 61 sense codons in protein-coding transcripts were visualized in successive 60-nt bins, up to 3000 nt downstream of the annotated start codon. We transformed the observed bin-specific codon frequencies to z-scores, defined relative to the global average frequency of codon occurrences within a bin, across the 50 bins downstream of the AUG. We observed that most codons displayed a distinct position-dependent z-score profile as we move downstream from the start codon, with some codons increasing in usage and others decreasing (Figure 1). This led us to evaluate a more refined model of PDCUB in the human transcriptome.

We investigated whether the heatmap of z-score profiles reveals other patterns in position-dependent codon usage when the codons are clustered according to distinct criteria. First, we organized codons by their amino acid chemical properties, but did not observe a strong correspondence between z-score profile and encoded amino acid chemical type (Figure 2). We next performed hierarchical clustering on the codons and found that they were predominantly organized by G and C (GC) nucleotide content. To test this observation, we compared our computed z-score dendrogram distance against a modified Hamming distance that compares the extent of variation of GC nucleotides between two codons (Figure 1). We found a strong correspondence between the number of common GC nucleotides and the dendrogram distance in Figure 1A, suggesting that the GC content of each codon is indeed a significant factor in the pattern of position-dependent z-scores (Figure 3).

Next, we grouped codons by their GC content and plotted z-score curves for each individual codon (Figure 4). We found that codons consisting exclusively of GC nucleotides show a strong enrichment near the start codon, while codons consisting exclusively of A or U (AU) nucleotides are depleted immediately after the start codon. Codons with either 1 or 2 GC nucleotides continue this trend, with GC-rich codons (2 GC nucleotides) showing greater similarity to GC-only codons, while GC-poor codons (1 GC nucleotide) show greater similarity to AU-only codons.

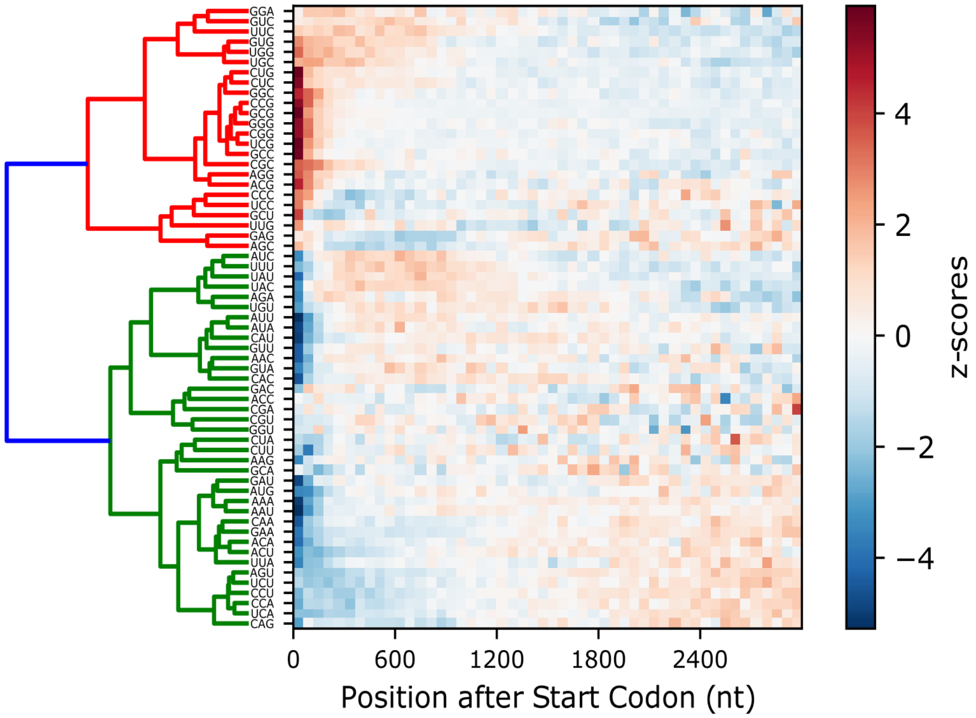


Figure 1: *A "birds eye view of position-dependent codon usage-bias in the human transcriptome (GENCDOE protein coding transcript models). A heatmap and dendrogram visualizing the z-scores computed from codon occurrence frequency in bins of size 60 nt.*

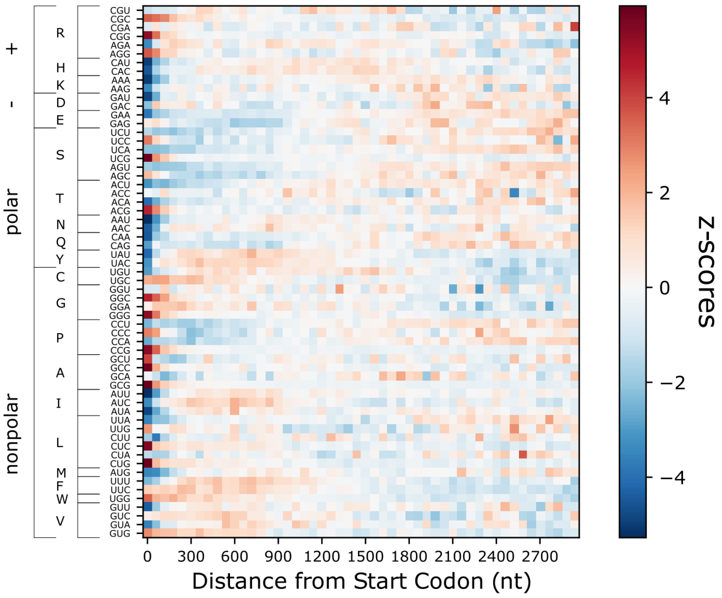


Figure 2: *Z-score heatmap computed from codon counts in bins of size 60 nt, for the first 3000 nt of all coding regions. Codons are grouped by amino acid and by the chemical category of the amino acid.*

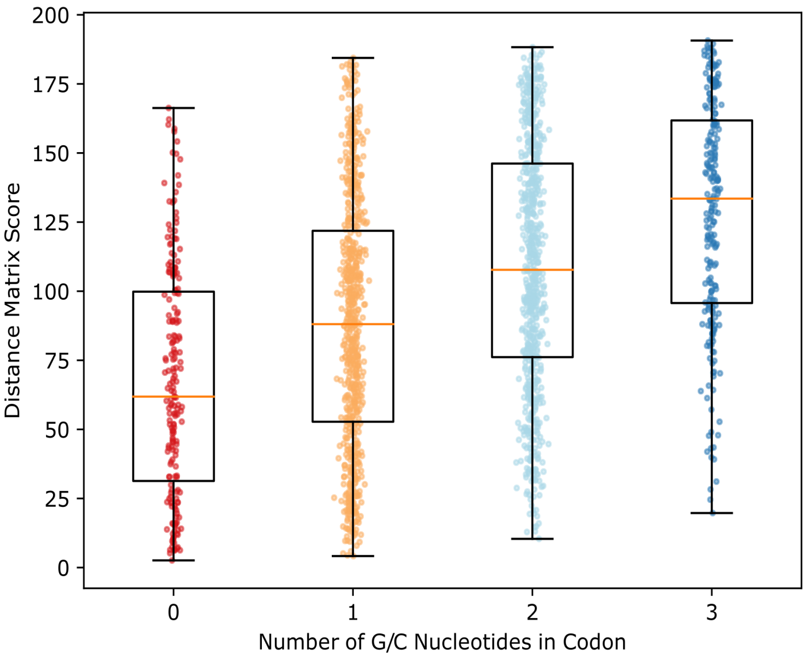


Figure 3: *Scatter-boxplot comparing the distance between codons based on the difference in GC nucleotides versus distance between codons based on hierarchical clustering of the position-dependent z-score profiles.*

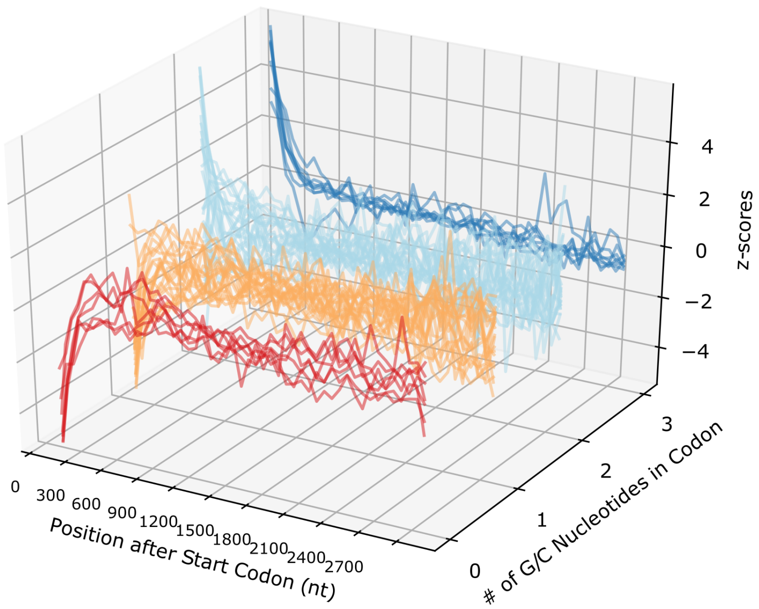
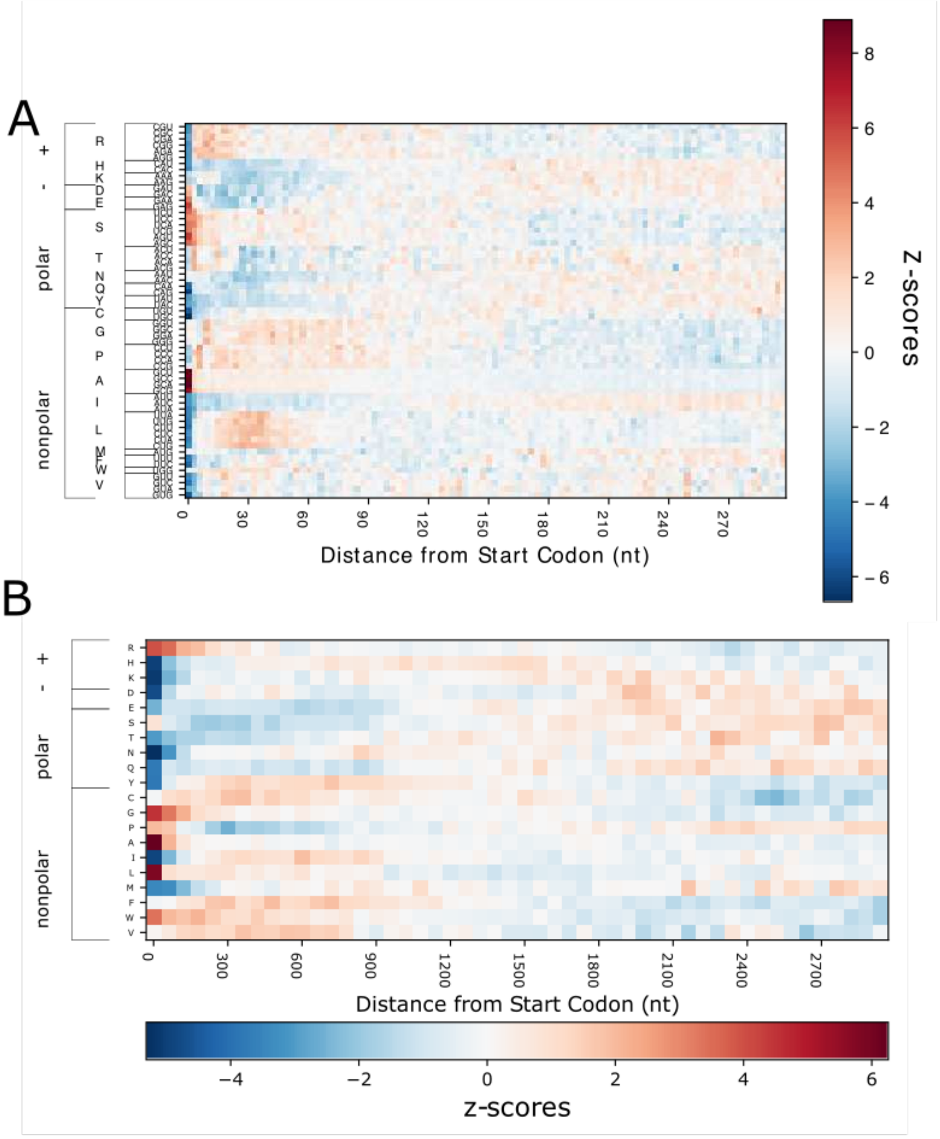


__Figure 4__: Curves show the z-scores computed from codon occurrence frequencies, shown in clusters corresponding to the number of G/C nucleotides within the codon.

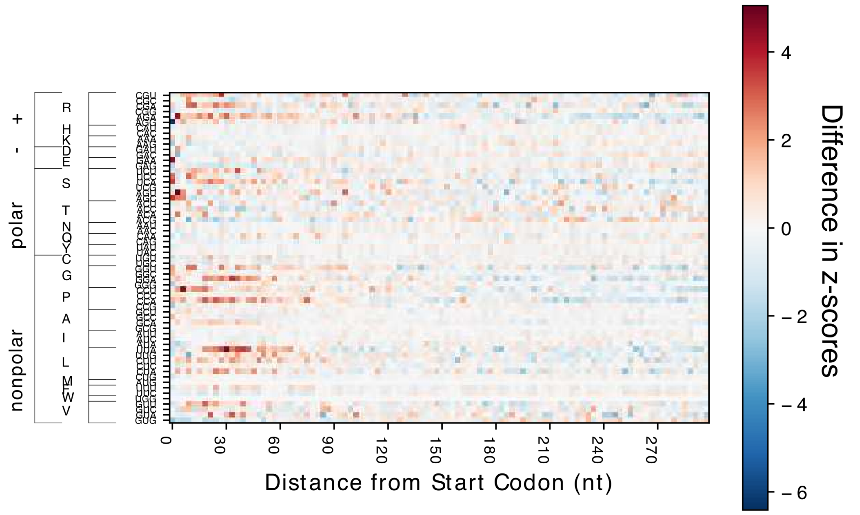
Observing PDCUB for Shuffled Sequences

We wanted to investigate whether these position dependent trends were preserved in amino acid sequences. We shuffled synonymous codons while preserving encoded amino acid ordering in our original dataset and repeated our initial analysis, grouping codons by amino acid and chemical category (Figure 5A). We observed that the z-scores for the shuffle sequences displays strong levels of enrichment and depletion immediately downstream of the start codon in a manner similar to the unshuffled sequences. We also noticed that amino acids in general became enriched, or depleted, which wasn't as prominent in Figure 1. The enrichment or depletion of one codon seemed to result in the overall enrichment or depletion of its parent amino acid. This position dependence at the amino acid level was also observed at a larger bin size (Figure 5B).

We continued to investigate by visualizing the difference between the shuffled codon z-scores and the unshuffled codon z-scores (Figure 6). We found that codons that became more enriched immediately downstream of the start codon after shuffling tended to have an ‘A’ or ‘U’ in its 3rd position. Similarly, codons that became depleted after shuffling tended to have a ‘G’ or ‘C’ in its 3rd position.



<u>Figure 5</u>*: *A. *Z-score heatmap computed from codon counts in bins of size 3 nt, for the first 300 nt of all coding regions. Codons have been shuffled.* ***B.*** *Z-score heatmap based on amino acid counts in bins of size 60 nt for the first 3000 nt of all coding regions.*



<u>Figure 6:</u> *Heatmap visualizing the difference in z-scores between shuffled and unshuffled codons, in bins of size 3 nts for the first 300 nts of all coding regions. Codons have grouped by amino acid and chemical group.*

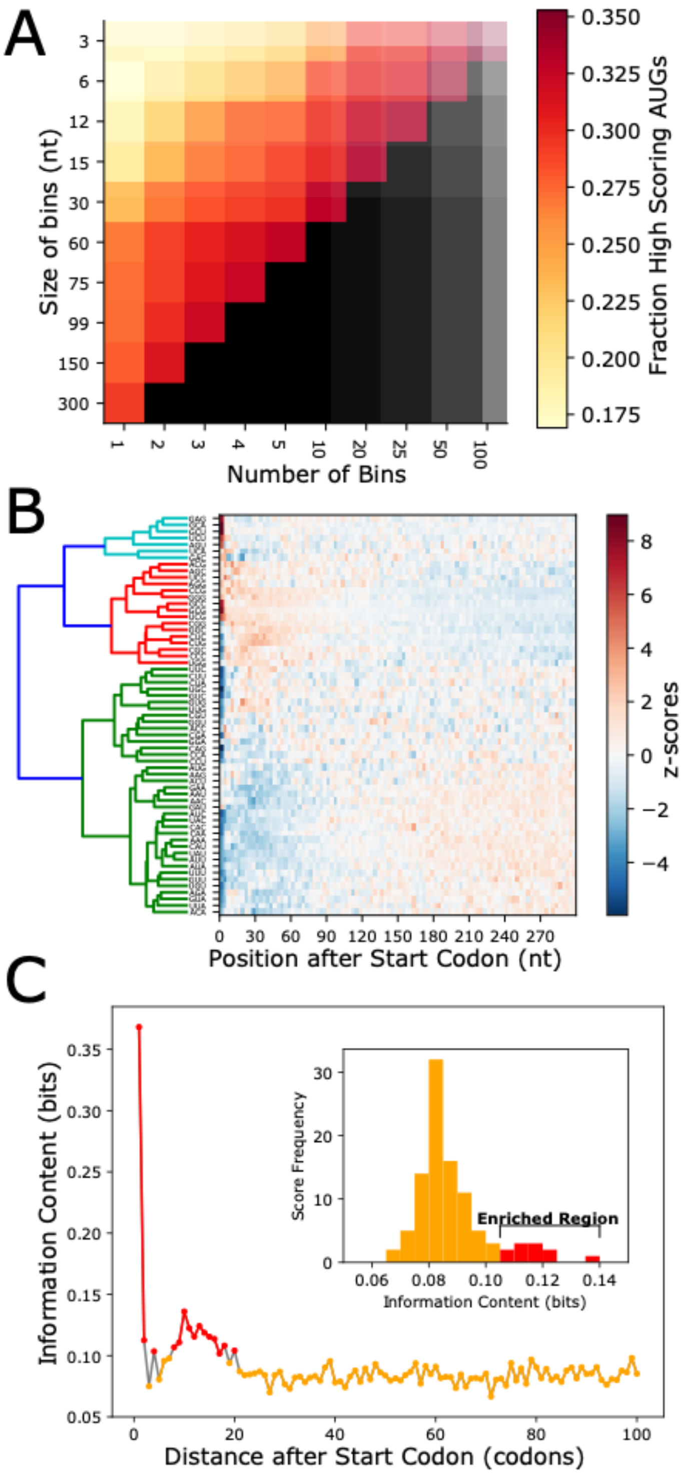
Optimizing PDCUB weight matrix

We used the prediction of start codons to assess the importance of the observed PDCUB patterns in the human transcriptome. We created training, validation, and test sets by subdividing the protein-coding transcripts from GENCODE using an 80:10:10 split. We created PDCUB position-specific scoring matrices (PSSMs, or weight matrices) to score the occurrence of a codon *c* at bin *b* of a potential open reading frame, relative to the global frequency of the nucleotide triplet corresponding to the codon *c*. We assigned the PDCUB weight matrix score (PDCUB score) to the series of triplets that follow each AUG in the transcripts across all frames, to discern whether the PDCUB score was sufficient to distinguish start codons from all other AUGs in a transcript.

We focused on the first 300 nt downstream of the start codon so that our score could be applied to most mRNA and lncRNA transcripts, including those with shorter ORFs. As a metric of predictive performance, we computed a recall, defined as the percentage of transcripts in which our PDCUB score identified the start codon as the highest-scoring AUG in that transcript. We observed that start codon recall depends on bin number and bin size, so we systematically computed the recall for combinations of both parameters (Figure 7A). We computed our PDCUB weight matrix using the training set using different combinations of bin number and size, selected an optimal set of parameters using performance on the validation set, and computed final percentages on the test set. We found that start codons were detected with maximum accuracy using one codon per bin – in other words, the combination of 100 single-codon bins provided the greatest predictive accuracy (Figure 7A). The z-score heatmap for this optimal bin size and number of bins is presented in Figure 7B. Hierarchical clustering of this heat map also corresponds to codon GC-content.

Information content of the PDCUB model

We used the PDCUB probability matrix to compute the information content as a function of nucleotide position or bin (see Methods). We observed enriched information content in the range from codons 0 to 20, relative to and downstream of the start codon (Figure 7C).



<u>Figure 7</u>*: A.** *A heatmap demonstrating the optimal bin size and number of bins within the first 300 nt of CDS regions of the human transcriptome favors 100 3-nt bins.* ***B. *A codon z-score heat map with hierarchical clustering for the 300 nt after the start codon.* ***C.*** *The information content of all transcripts, computed from position-dependent codon frequencies, shows a region of enriched information content relative to the background trinucleotide frequencies. Red line segments indicated some regions of enriched information content. The inset shows a histogram of the information content for the first 100 codons with a tail of enriched information content, shown in red.*

Start codon identification as a function of transcript length

We next investigated the role of transcript length in the percentage of correctly identified start codons – specifically, whether transcript length is correlated with the percentage of start codons identified by our PDCUB score as the highest-scoring AUG triplet in a given transcript. We compared the central tendencies (mean and median) for PDCUB scores computed for start codons and for other AUGs. We found that there was a significant difference between start codons and non-start AUGs for all transcript lengths (Figure 10).

We compared the length-dependent recall of the PDCUB score with four other predictive models. First, we compared the PDCUB model to a Kozak consensus sequence model, which relies on finding within a transcript the nucleotide sequence of fewest possible mismatches to the Kozak consensus sequence [16]. We also compared the PDCUB model to a weight matrix computed from base composition in the region of the Kozak sequence, –10 to +1 relative to the start codon. A null model that estimates the rate for the random selection of start codons amongst all AUG triplets, and a position-dependent GC-content model, were also used as controls. We found that the PDCUB model consistently performed best among the five models (Figure 8A). Even though PDCUB only considers the first 300 nt of the coding sequence, we observed improved performance with shorter transcripts. Figure 8B shows an example transcript, plotting the PDCUB score over all positions, and plotted separately for each reading frame. This example transcript conforms to the trend observed in the majority of transcripts of this length, with the start codon having the highest PDCUB score.

We followed up our investigation of position-dependent trends in amino acids by also comparing PDCUB to position dependent amino acid usage bias as a predictor of start codons. We also tested GC3_ (Figure 9). We found that PDCUB outperformed both the amino acid usage bias and the -GC3 model. However, both models performed better than the GC model. Both models also have a higher performance with shorter transcripts.

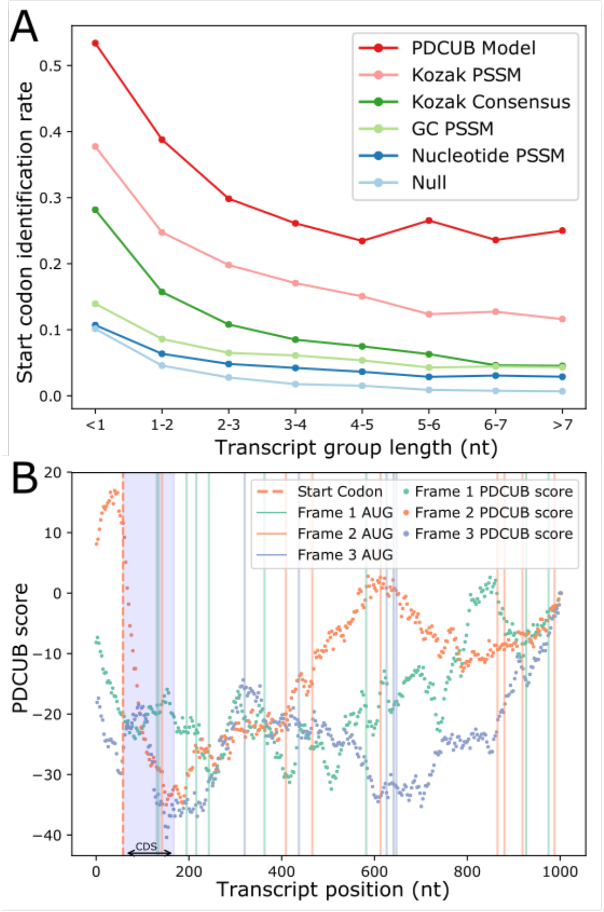
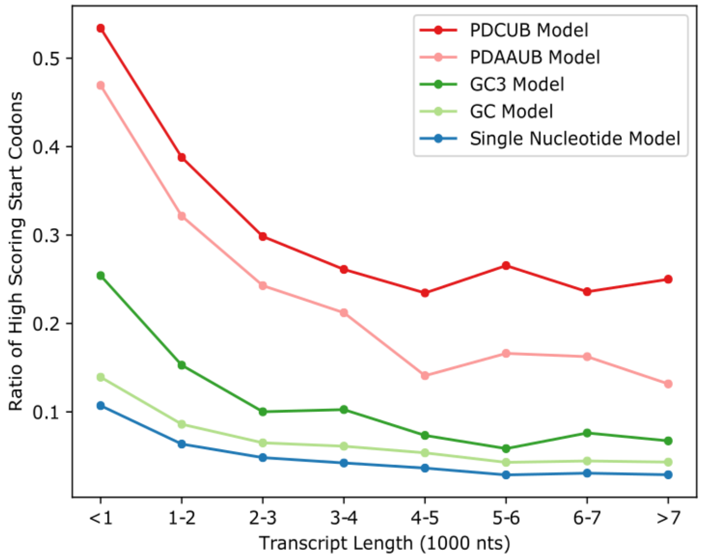


***Figure 8**: **A.** The rate of start codon identification compared for different models. The curves show what fraction are identified per transcript, comparing the PDCUB score computed from a PSSM compared to the Kozak consensus sequence AUGGCC, a PSSM computed from the positions -10 to +2 relative to the start codon for all transcripts, a PSSM computed from the position-dependent GC content within CDS regions relative to global GC content, a PSSM computed from single-nucleotide frequencies in CDS regions in CDS regions relative to global nucleotide frequencies, a random selection model. **B.** The PDCUB score plotted as a function of position, starting at each codon, and plotted separately for each reading frame, for an example short (~1000nt) transcript. Blue shaded area beginning at the start codon indicates the CDS region.*



***<u>Figure 9</u>**: The rate of start codon identification compared for additional different models. The curves show what fraction of start codons are identified per transcript, comparing the PDCUB score computed from a PSSM to the position-dependent amino acid usage model PSSM, a PSSM computed from the GC3 content within CDS regions relative to global GC3 frequencies.*

Comparison of mRNAs and lncRNAs with PDCUB scores

We next investigated the role of transcript length in the percentage of correctly identified start codons – specifically, whether transcript length is correlated with the percentage of start codons identified by our PDCUB score as the highest-scoring AUG triplet in a given transcript. We compared the central tendencies (mean and median) for PDCUB scores computed for start codons and for other AUGs. We found that there was a significant difference between start codons and non-start AUGs for all transcript lengths (Figure 10A).

We compared scoring mRNAs using the annotated CDS to lncRNAs by scoring the longest ORF in each GENCODE lncRNA with an ORF using the PDCUB scores. We found that in general, PDCUB was substantially higher in mRNAs than in lncRNAs (Figure 10B). We performed a Kolmogorov-

Smirnov test to evaluate the difference between the mRNA PDCUB distribution to the distribution for lncRNAs, and it resulted in a p-value of 0.731e-07. We further subdivided both the mRNAs and lncRNAs into different ranges of 1000 nt. Figure 10A shows the mean and median values for each of these bins, along with 95% confidence intervals.

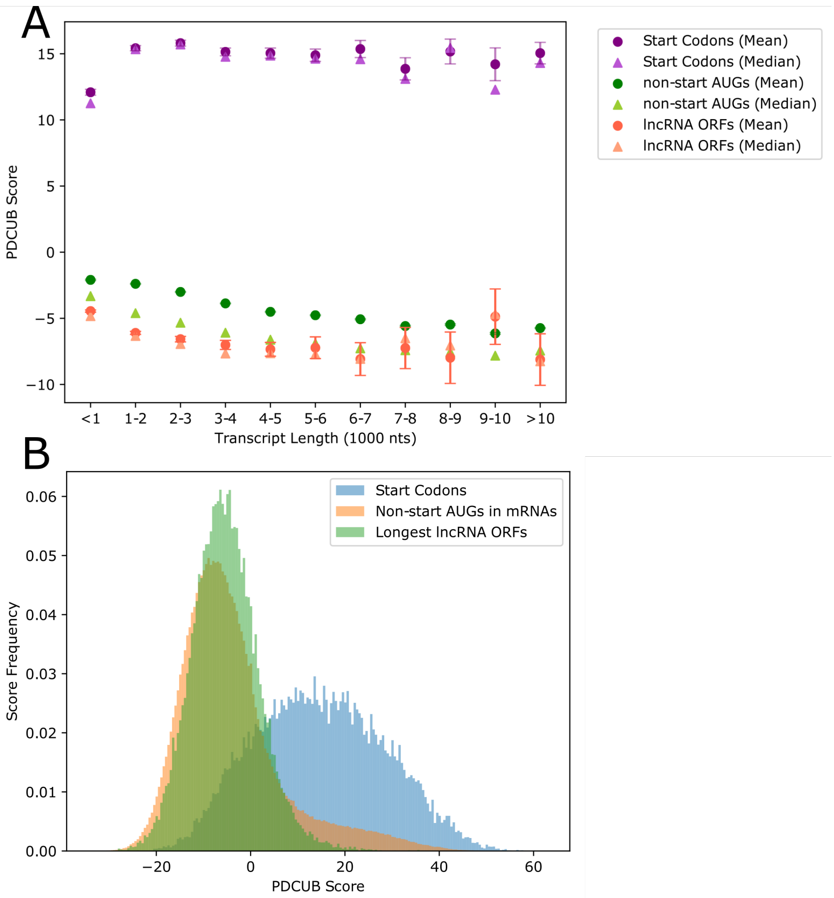


Figure 10: ***A.*** *Comparison of the mean and median PDCUB score for start codons in transcripts of different lengths in the test set, relative to non-start AUG triplets, in increments of 1000 nt. In addition, this also compares to the scores for the AUG in the longest ORF for lncRNAs using the same size bins.* ***B****. Histograms comparing the PDCUB score for start codons compared to non-start AUG triplets and the AUG in the longest ORF for lncRNAs. Mean and standard deviation from the non-start distribution is used to compute statistical significance of PDCUB scores.*

Functional enrichment in high-PDCUB transcripts

We identified a subset of transcripts with particularly high-significance PDCUB scores relative to the distribution of scores observed over all AUG triplets (Figure 10B) and named this our significant PDCUB set. We also divided our total transcript set into quintiles based on PDCUB score (see methods). We examined whether transcripts in these sets were enriched for biological functions or protein sequence patterns that might explain their scores.

Gene Ontology (GO) enrichment

We investigated whether the transcripts in our high-significance and low-significance PDCUB transcript sets, as well as in the various PDCUB quintiles (see methods), were enriched or depleted for particular biological functions. A table of significant Gene Ontology (GO) terms is provided in Supplementary Table 1. A similar analysis was performed for CAI quintiles as a control. Enrichment of significant GO terms associated with transcripts in the bottom PDCUB quintile is visualized in a scatterplot (Figure 11A). Similarly, GO terms that are enriched in the top quintile are shown in Figure 11B.

Growth and development. Among the highest-PDCUB transcripts, we observed a statistical enrichment of most developmental processes, with a corresponding depletion of these same processes among the lowest-PDCUB transcripts (Supplementary Table 1, blue terms). Examples include generation, development, differentiation, formation, and morphogenesis of the sensory, nervous, endocrine, gastrointestinal, circulatory, and musculoskeletal systems, as well as of various organs, glands, tissues, cells, and cellular components. Terms with the greatest enrichment ($> 5x$ expected) were nephron tubule formation and regulation of norepinephrine secretion, the latter of which is increasingly thought to be a uniquely important regulator of brain development [17]. The only exceptions to this trend were development-related terms involving skin and keratin, which were enriched only among low-PDCUB transcripts.

Signaling and transport. Nearly all significant GO terms pertaining to signaling and transport are enriched among high-PDCUB transcripts but depleted among mid- and low-PDCUB transcripts (Supplementary Table 1, purple terms). Exceptions to this trend primarily consist of terms tied to immunological function, which are depleted among high-PDCUB transcripts but enriched among mid- and low-PDCUB transcripts, as well as G-protein coupled receptor signaling pathway terms that are enriched across all PDCUB categories.

Immune response and defense. Significant GO terms pertaining to immune response are depleted among high-PDCUB transcripts and enriched among mid- and low-PDCUB transcripts (Supplementary Table 1, green terms). We found that “adaptive immune response” was the most statistically significant GO term in the analysis of both the significant PDCUB set and the top PDCUB quintile, and was strongly depleted in both cases. We found a single reversal of this trend in the term “phagocytosis”, which is the most depleted term in the second PDCUB quintile. Interestingly, terms relating to specific immune cells such as lymphocytes, leukocytes, granulocytes, monocytes, eosinophils, neutrophils, and natural killer cells are highly enriched among the middle quintile (40%-60% PDCUB scores) but neither significantly enriched nor depleted in the other quintiles.

Chemical stimulus and sensory perception. Significant GO terms related to stimulus *response* and *regulation* are all *enriched* among high-PDCUB transcripts (Supplementary Table 1, pink terms). By contrast, significant GO terms pertaining to sensory or stimulus *detection* and *perception* are all *depleted* among the high-PDCUB transcripts (Supplementary Table 1, yellow terms).

Among mid- and low-PDCUB transcripts the opposite situation holds true, with enrichment of GO terms involving sensory or stimulus detection and perception, and depletion of terms involving stimulus response and regulation.

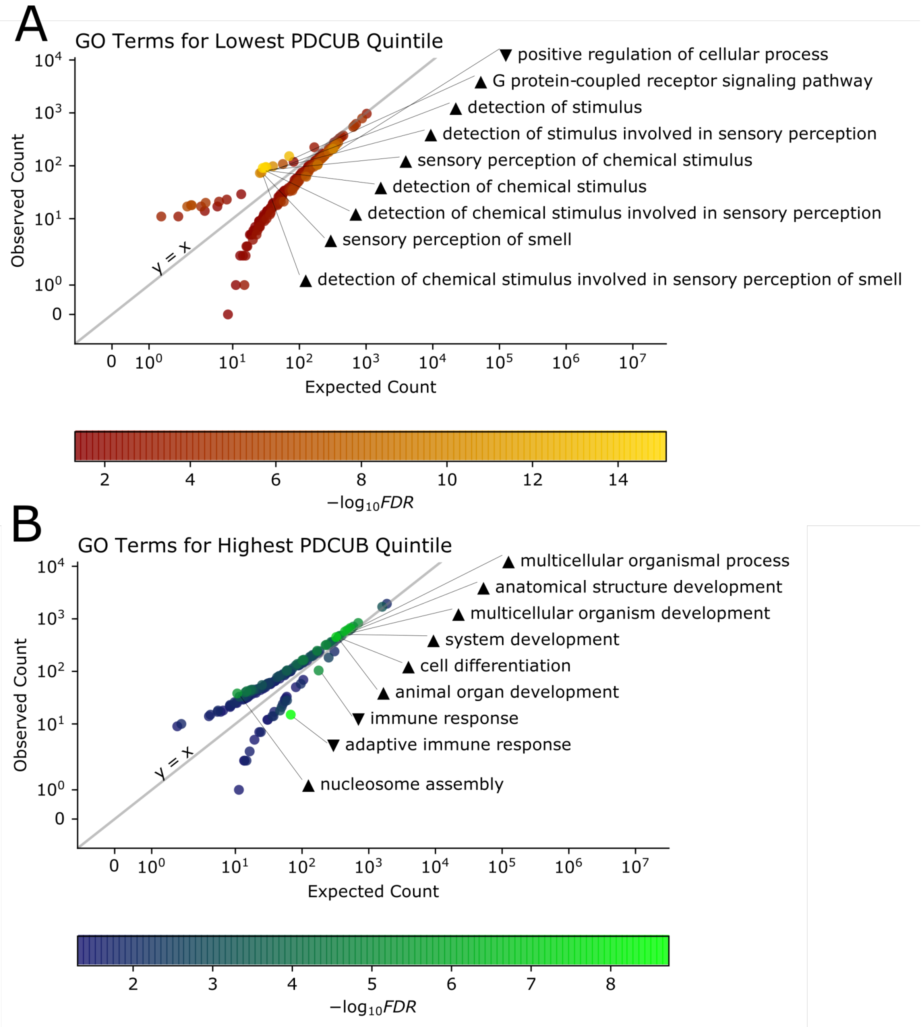
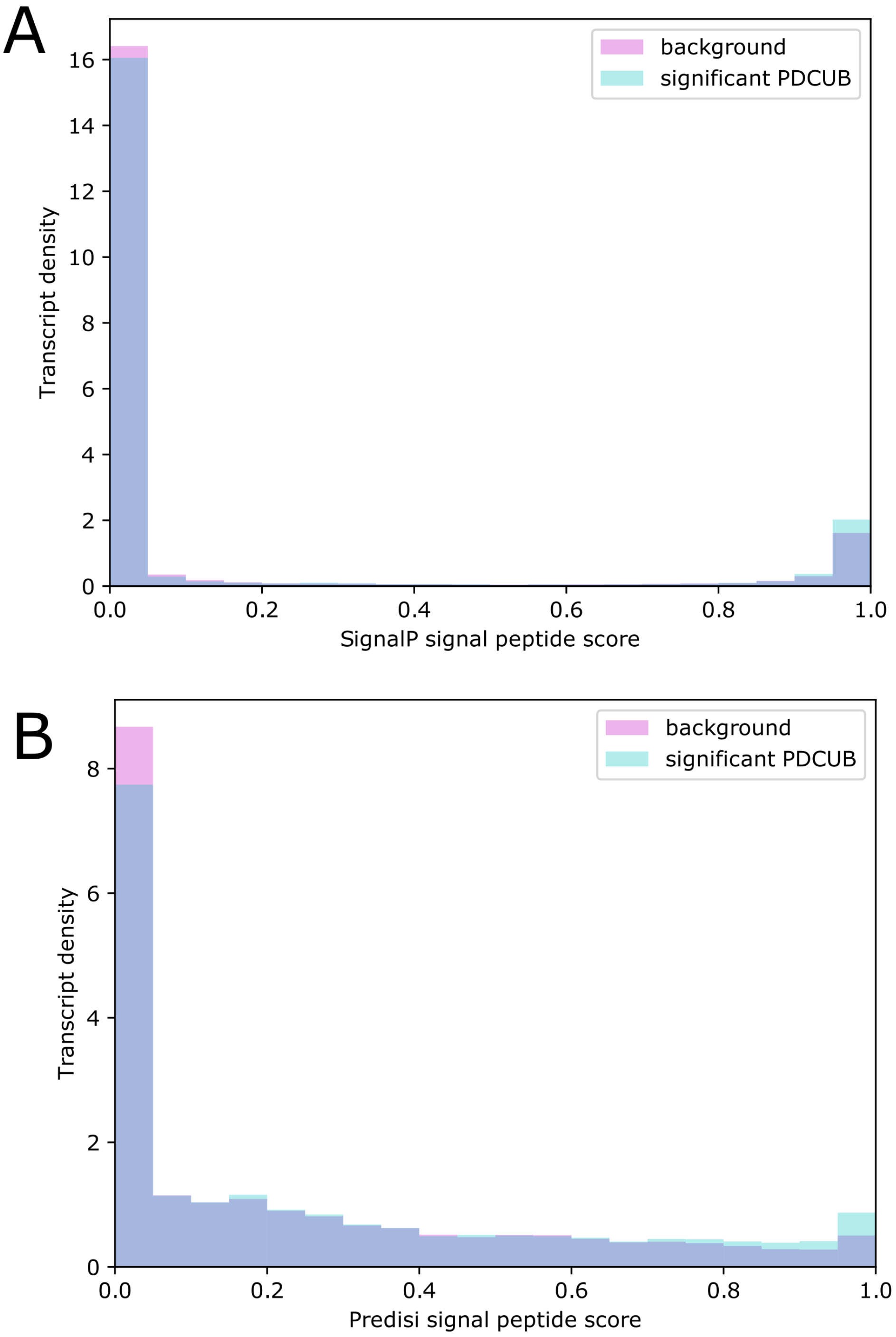


Figure 11: *Scatter plot representing statistically-enriched GO terms associated with transcripts that have statistically significant PDCUB scores compared to the scores of all AUGS within the human transcriptome. Each dot corresponds to an enriched GO term, the x-axis position corresponds to the expected number of occurrences, the y-axis represents the observed frequency within the PDCUB-enriched transcripts. Transcripts with the highest PDCUB scores are enriched. Triangles by the labels indicate enriched (pointed up) and depleted (down) GO terms.*

Signal Peptides

To investigate whether our results reflect the presence of nucleotide sequences coding for signal peptides, we used the signal-peptide prediction tools PrediSi [18] and SignalP [19] to predict which transcripts are likely to contain signal peptides. We performed a statistical test comparing the number of predicted signal peptides in the high-significance PDCUB transcripts to a background predicted rate for a complement set of all protein-coding GENCODE transcripts not contained within the significant PDCUB set. We observe that 15-24% of proteins encoded by high-PDCUB transcripts are predicted to contain signal peptides, amounting to a 1-4% increase in predicted signal peptides among high-significance PDCUB transcripts relative to all protein-coding transcripts in GENCODE (Figure 12). We tested this difference via binomial test and found the increase to be statistically significant, with a p-value of $< 10^{-23}$. A Venn diagram of transcripts predicted by each approach to have a signal peptide, compared to transcripts with a high-significance PDCUB score, is shown in Figure 13. Analysis of individual PDCUB quintiles showed a rough correlation between signal peptide density and quintile number for PrediSi and SignalP alike. In both cases, we found the bottom quintile to have the lowest predicted density, with the third quintile being near the global average density and the fourth quintile having the highest predicted density (Figure 14).



<u>Figure 12</u>**: Histograms comparing the predicted probability of signal peptides for significant PDCUB transcripts against those of the full GENCODE set.* ***A. *SignalP distribution.* ***B.*** *PrediSi Distribution.*

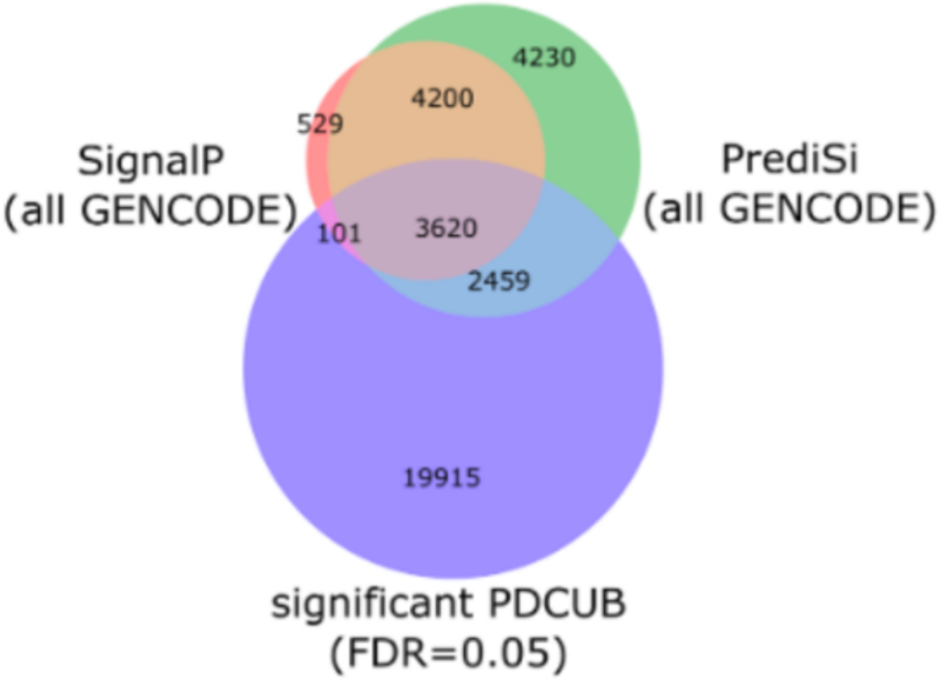
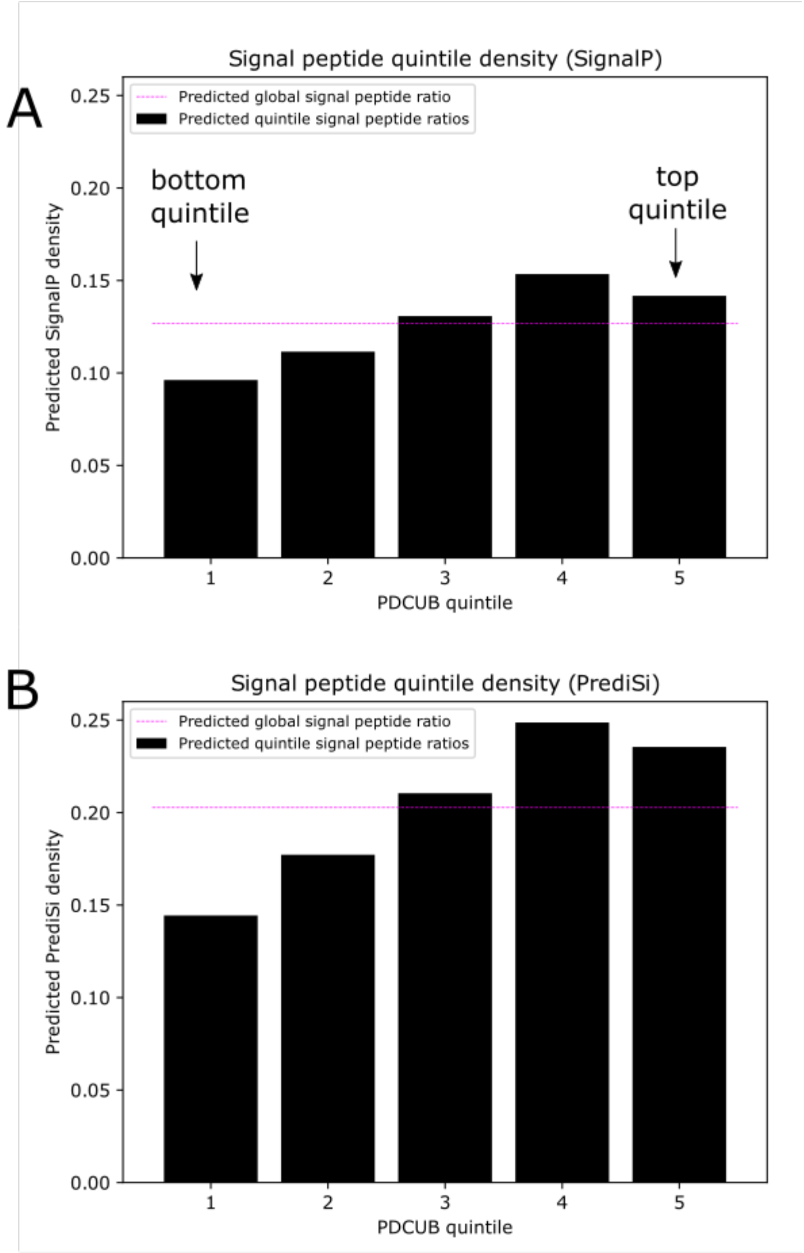


__Figure 13__: Venn diagram of coding transcripts predicted to have signal peptides according to SignalP (top left) and PrediSi (top right), compared to coding transcripts that meet the significance threshold for high PDCUB score (bottom).



<u>Figure 14</u>**: Predicted signal peptide densities for each PDCUB quintile. Magenta line indicates global predicted signal peptide density.* ***A. *SignalP predictions.* ***B.*** *PrediSi predictions.*

Local tAI and CAI

We next sought to contextualize PDCUB against standard measures of translational efficiency, including the discovery of the translation initiation ramp [20]. We confirmed the presence of the translational ramp seen by Tuller et al, through a metagene plot of tAI against CDS position for the first 300 codons of the coding transcriptome (Figure 15). We repeated this process with the transcriptome split into quintiles according to PDCUB scores, and found a clear, nonoverlapping spectrum of ascending average tAI values as we move from the lowest-PDCUB transcripts to the highest-PDCUB transcripts (Figure 16). This PDCUB-defined spectrum exhibits differently shaped local tAI curves, with the fifth quintile (top quintile, highest PDCUB score) having the most pronounced translational ramp and lower quintiles having diminishing ramp size until the second quintile, which is flat with no ramp. More strikingly, the first quintile (bottom quintile, lowest PDCUB score) exhibits a “translational valley”, with tAI descending quickly from codons 0 to 30 before flattening out until codon 100, whereupon we observe a short ramp that is delayed relative to that of the highest PDCUB quintile.

We also compared tAI trends for quintiles defined for CAI computed for the entire coding sequence (global CAI), as well as for specific regions of the coding sequence (regional CAI). Trends for tAI plotted against quintiles defined for global CAI and regional CAI for the first 100 codons showed similar but much weaker trends, with global CAI showing the weakest changes in tAI between the three methods (Figure 17). Quantitative comparison of quintile-specific tAI trajectories between the three methods confirmed that for both PDCUB and regional CAI the greatest change in tAI was seen in the first and fifth quintiles, corresponding respectively to a translational valley and a strong ramp, while global CAI quintiles display a steady monotonic increase from quintiles one through five (Figure 18). Finally, we plotted tAI against PDCUB score for the coding transcriptome and found poor general correlation between the two parameters ($R^2 \approx 0.3$, Figure 19).

We calculated local tAI values for CDS regions corresponding to the first and second windows of length w codons, for $w = 30$ and $w = 50$, for all transcripts. The midpoints of these two sets of paired windows correspond, respectively, to the lower and upper bounds reported for the size of the translation initiation ramp and should thus demark a boundary between lower and higher translational efficiency signals, which should be apparent when plotting the local tAIs for the second window against the first. No such boundary was apparent in our scatter plot, regardless of the PDCUB of the corresponding transcripts (Figure 20). However, we did see a clear gradient of PDCUB scores across the transcript distribution in both codon windows, with low-PDCUB transcripts concentrated at low tAI values, and high-PDCUB transcripts concentrated at high tAI values.

As an alternative approach to understanding how PDCUB compares to translational efficiency, we performed a similar set of analyses with regional CAI scores computed from global and regional relative adaptiveness (Supplementary Figures 21 and 22). This revealed a trend in which transcripts with lower PDCUB scores correspond to a diminished regional CAI in codons 30-50 (the translation initiation ramp) relative to the immediately subsequent 30-50 codons (the translation steady state). We plotted the CAI of each protein-coding transcript against its PDCUB score, and found the relationship between CAI and PDCUB to be similar to (but weaker than) that between tAI and PDCUB (Figure 23).

We further compared the CAI for the first 100 codons to score AUGs for the prediction of start codons in the same analysis as Figure 3A. We found that CAI was not predictive for start codons, and was in fact lower than random expectation for the same window that PDCUB is computed (Figure 24).

Ribo-seq profiles

Given the observation that higher ribosomal density is associated with lower translational speed [21], we examined whether transcripts with higher PDCUB scores show distinct trends in ribosomal occupancy compared to transcripts with lower PDCUB scores. Figure 25A shows the normalized ribosomal occupancy relative to the start codon for each PDCUB quintile. We observed a spectrum of ribosome profiles where the first quintile (lowest PDCUB) has the highest relative occupancy at initiation while the fifth quintile (highest PDCUB) has the lowest relative occupancy. In contrast, a control experiment of ribosomal density plotted against CAI quintiles showed no such clear trend between the two metrics (Figure 25B).

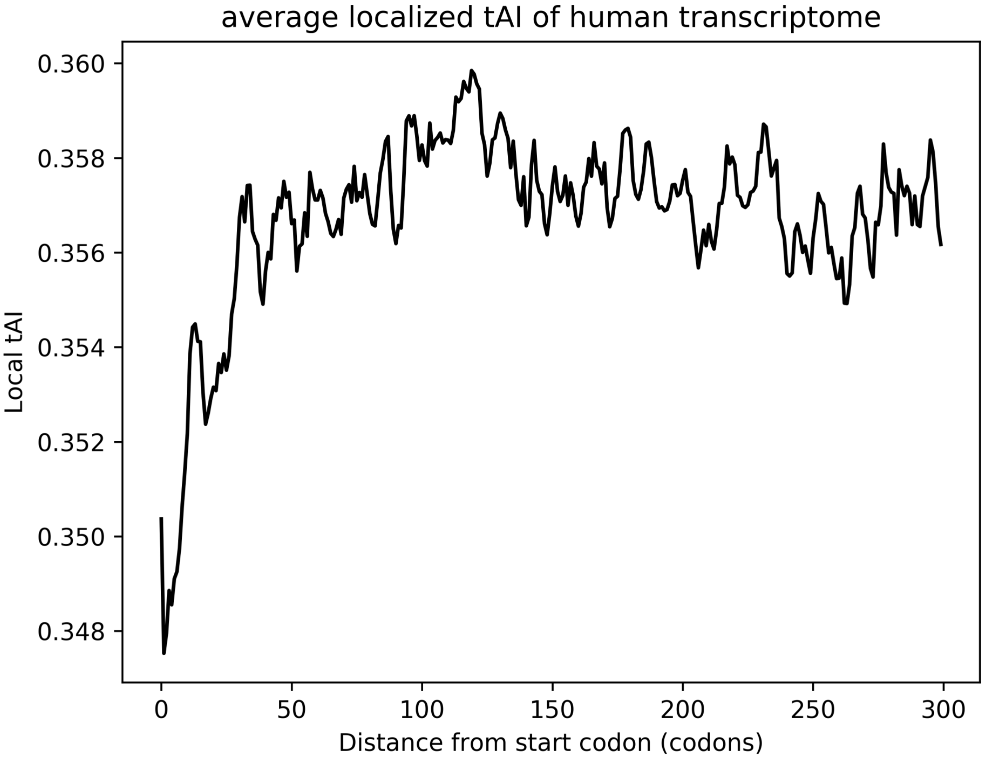


Figure 15: *Average local tAI for first 300 codons of coding transcriptome in humans, using a sliding window of size 5 codons. A distinct ramp is seen from codons 0 to 50.*

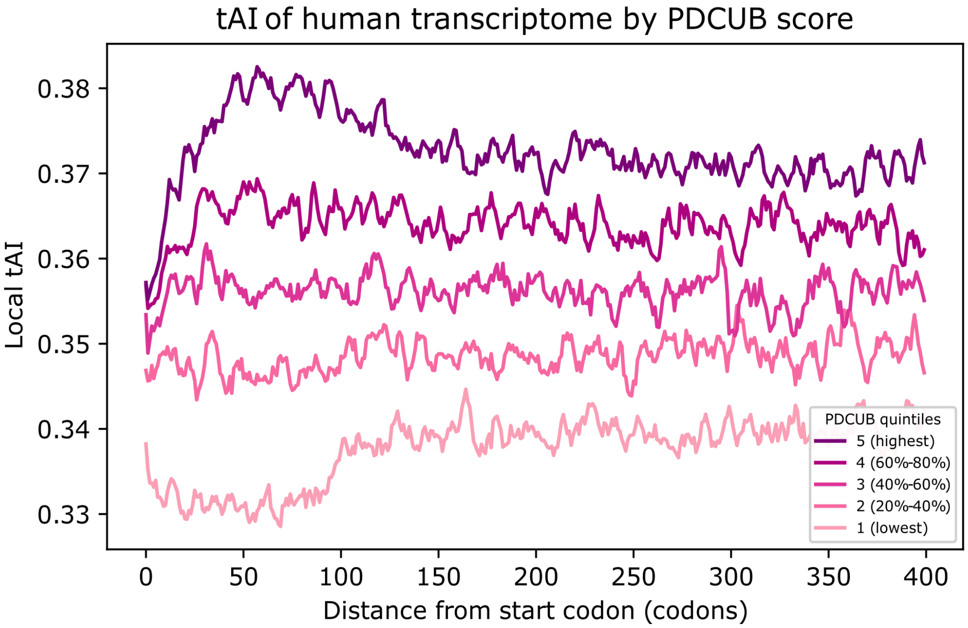
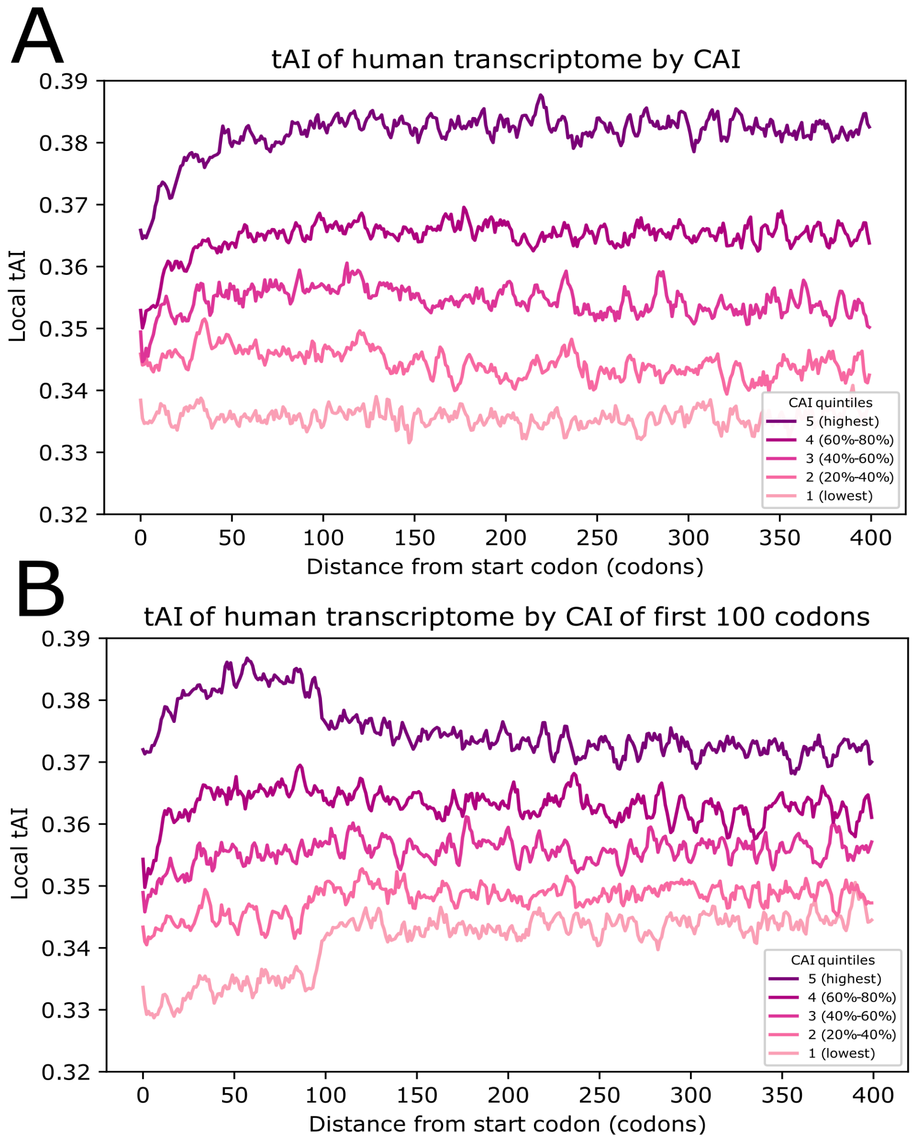


Figure 16: *Local tAI profiles for transcripts in quintiles defined for PDCUB score.*


A tAI of human transcriptome by CAI

B tAI of human transcriptome by CAI of first 100 codons

Figure 17: *Local tAI plotted against CAI quintiles.* ***A.*** *Global CAI.* ***B.*** *Regional CAI.*

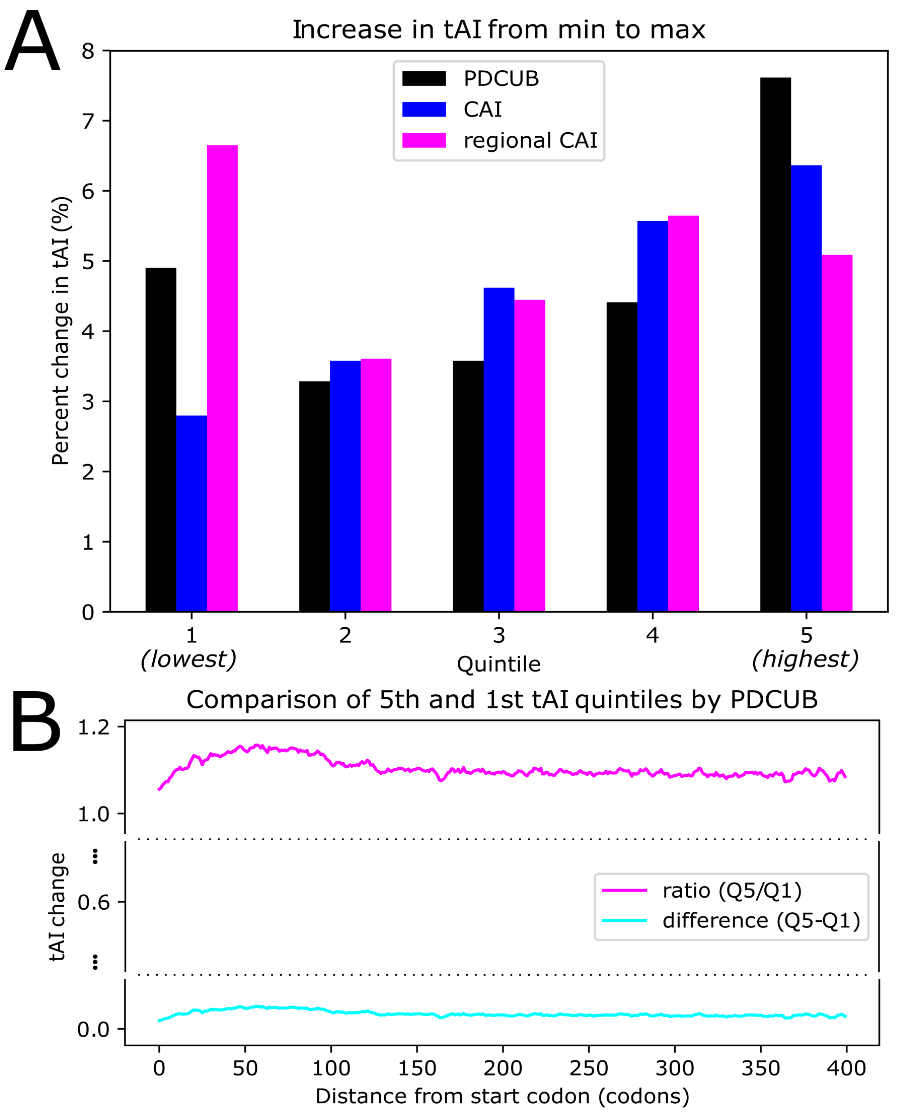


__Figure 18__**: Comparison of increase in tAI by quintile.* ***A. *Side-by-side percent increase in tAI from lowest to highest point for each PDCUB, CAI, and regional CAI quintile.* ***B****. Ratio of, and difference between, tAI for fifth and first PDCUB quintiles.*

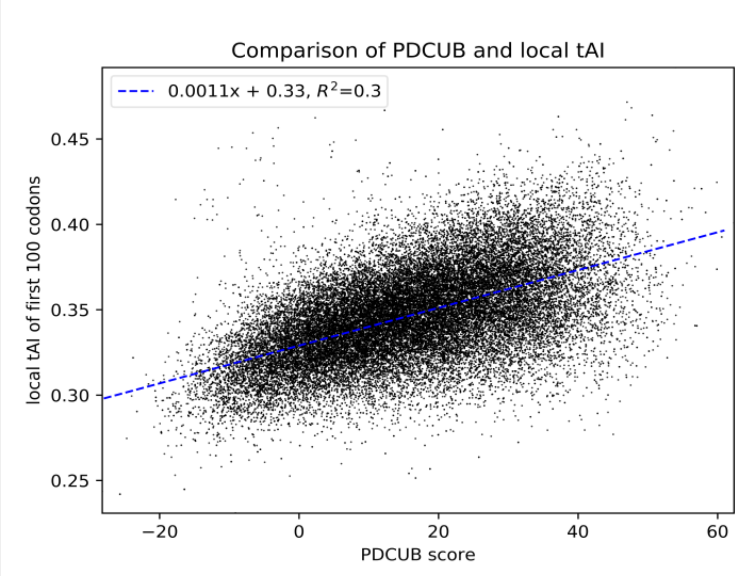
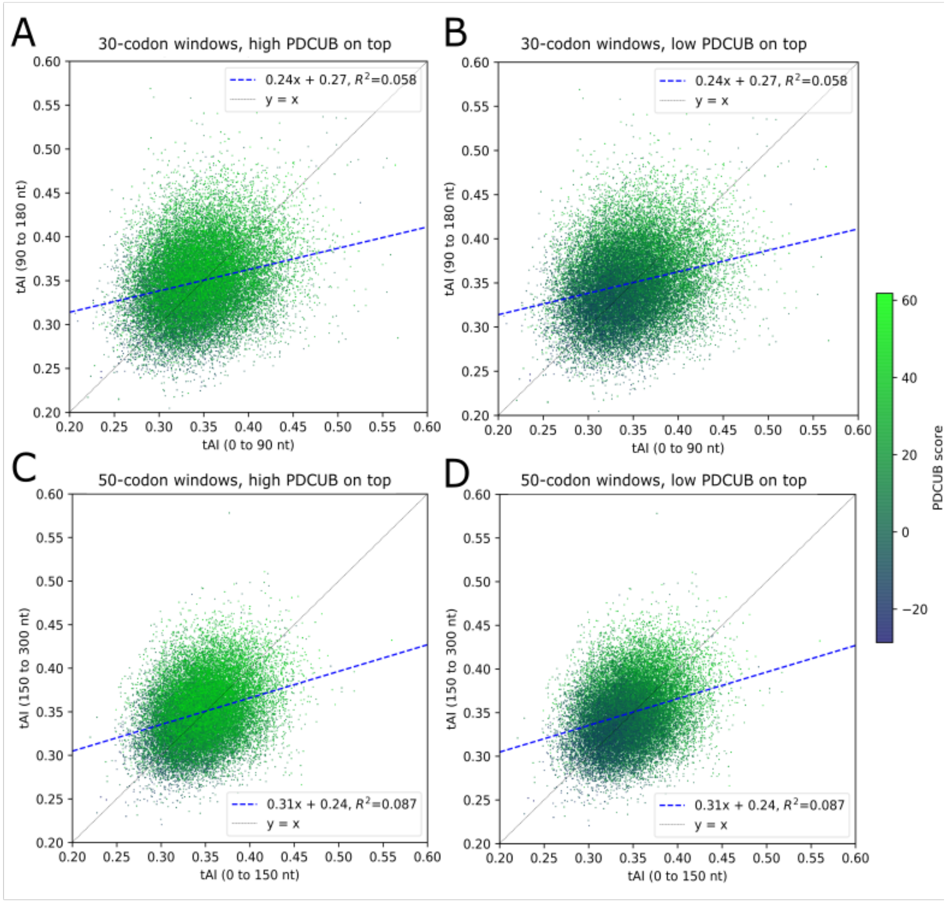


Figure 19: *Scatter plot of local tAI for the first 100 codons of all coding transcripts, plotted against the PDCUB score of the same transcript. Each dot corresponds to a transcript. The dashed blue is the best linear fit to the scatter data.*



*<u>**Figure 20**</u>: Scatterplots of relative local tAI for first and second bins in initial regions of coding transcripts. Each point corresponds to a single CDS, while the colormap indicates PDCUB score.* ***A.*** *Second bin of 30 codons plotted against first bin of 30 codons, with highest PDCUB transcripts on top.* ***B****. Same as A, but with lowest PDCUB transcripts on top.* ***C.*** *Second bin of 50 codons plotted against first bin of 50 codons, with highest PDCUB transcripts on top.* ***D.*** *Same as C, but with lowest PDCUB transcripts on top.*

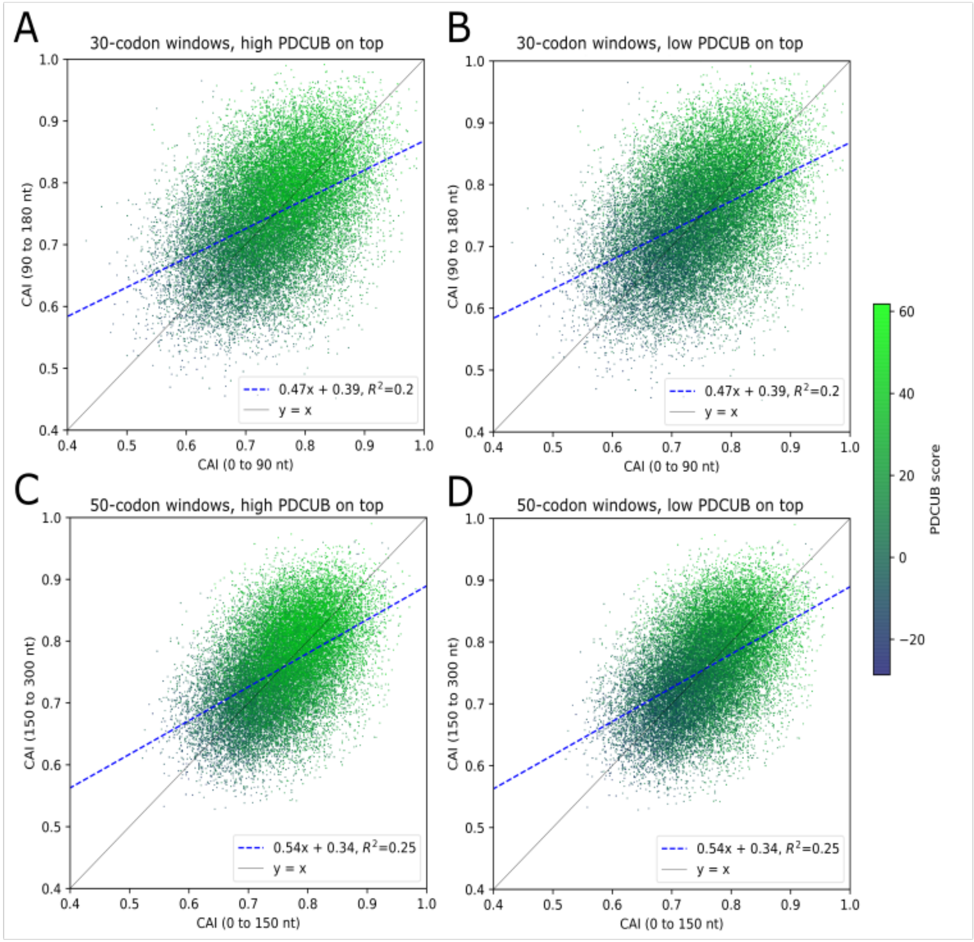


Figure 21*: Scatterplots of relative regional CAI (global RA) for first and second bins in initial regions of coding transcripts. Each point corresponds to a single CDS, while the colormap indicates PDCUB score. Low-scoring PDCUB sequences show the greatest deviation from y=x, while high-PDCUB transcripts correspond to a high regional CAI for both regions of the CDS.* ***A****. Second bin of 30 codons plotted against first bin of 30 codons, with highest-PDCUB transcripts on top.* ***B.*** *Same as A, but with lowest-PDCUB transcripts on top.* ***C.*** *Second bin of 50 codons plotted against first bin of 50 codons, with highest-PDCUB transcripts on top.* ***D.*** *Same as C, but with lowest-PDCUB transcripts on top.*

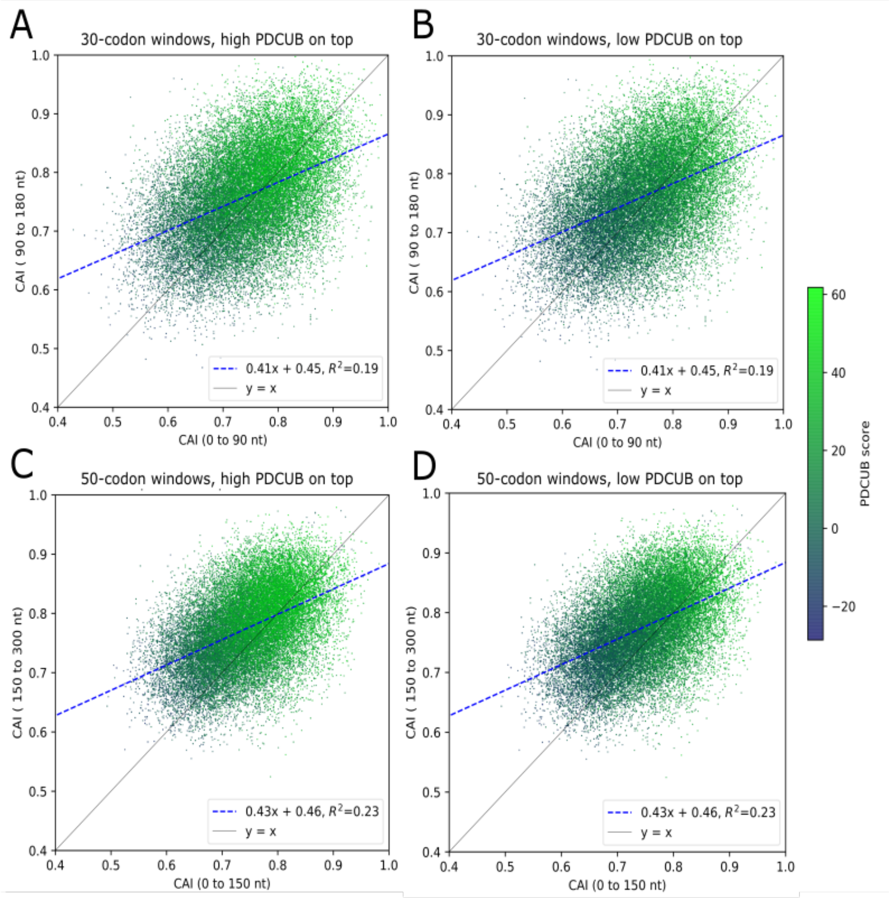


Figure 22*: Scatterplots of relative regional CAI (global RA) for first and second bins in initial regions of coding transcripts. Each point corresponds to a single CDS, while the colormap indicates PDCUB score. Low-scoring PDCUB sequences show the greatest deviation from y=x, while high-PDCUB transcripts correspond to a high regional CAI for both regions of the CDS.* ***A****. Second bin of 30 codons plotted against first bin of 30 codons, with highest-PDCUB transcripts on top.* ***B.*** *Same as A, but with lowest-PDCUB transcripts on top.* ***C.*** *Second bin of 50 codons plotted against first bin of 50 codons, with highest-PDCUB transcripts on top.* ***D.*** *Same as C, but with lowest-PDCUB transcripts on top.*

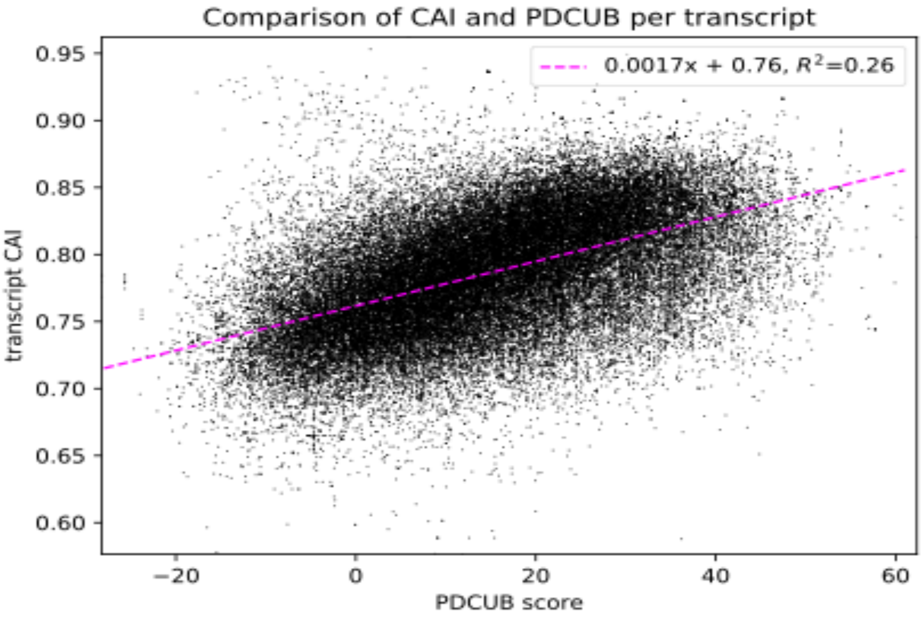


Figure 23: *Scatter plot of CAI for each coding transcript plotted against the PDCUB score of the same transcript. Each dot corresponds to a transcript. The dashed magenta line is the best linear fit to the scattered data.*

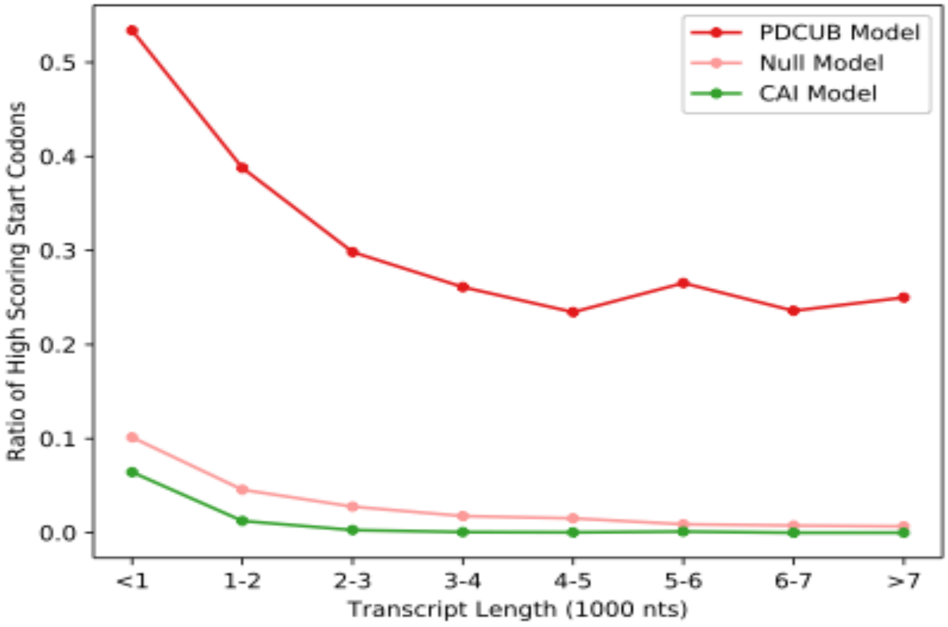
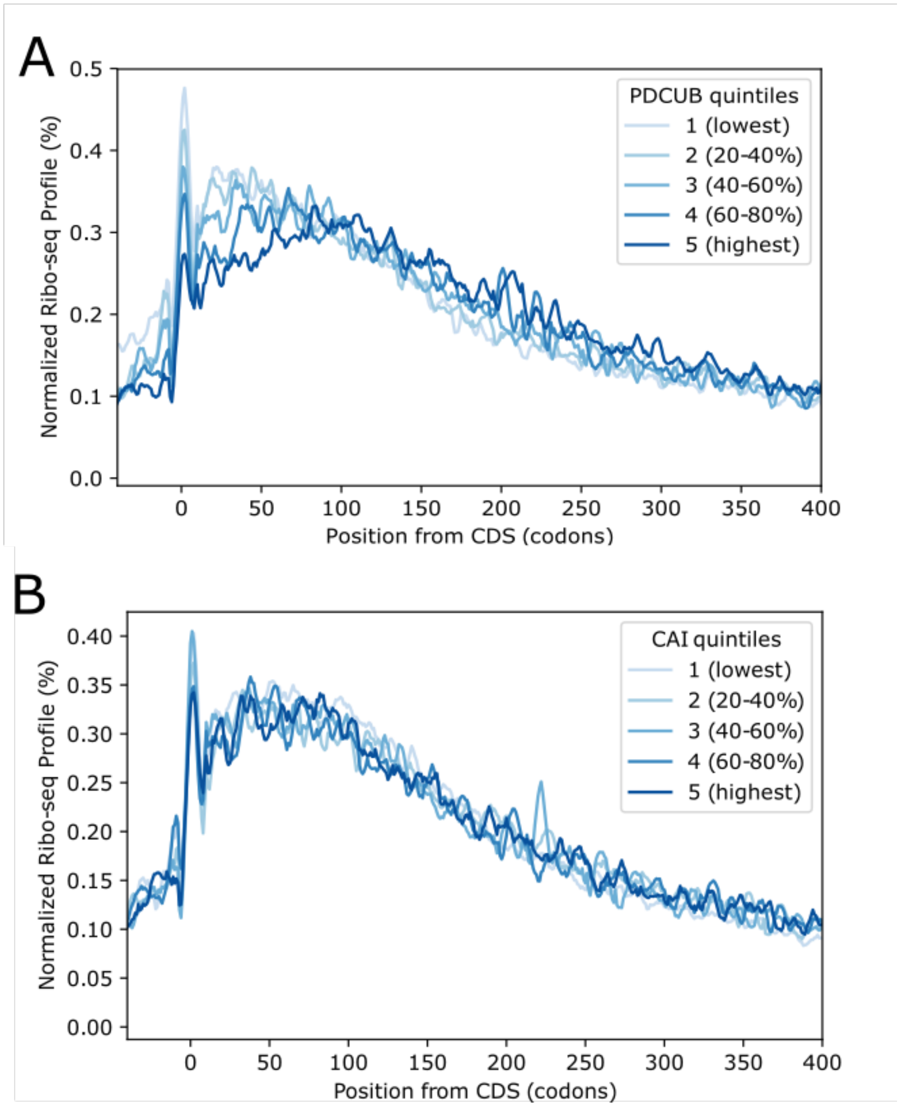


Figure 24: *The rate of start codon identification compared for different models. The curves show what fraction of start codons are identified per transcript, comparing the PDCUB score with other models similar to Figure 3A, but with CAI computed for the first 100 codons also compared.*



***<u>Figure 25</u>**: The shape of ribosomal occupancy, shown as the normalized Ribo-seq profile centered at the start codon at position 0 for each quintile. Read positions are A-shifted by 15-nt, and smoothed using a Savitzky-Galay filter. **A**. PDCUB quintiles. **B.** CAI quintiles.*

DISCUSSION

We observe position-dependent trends when visualizing codon usage across all protein-coding regions in the human transcriptome. Additionally, over one third of ORFs in individual protein-coding transcripts exhibit a statistically significant PDCUB when compared to reading frames that follow non-start AUGs. We have shown that start codons can be distinguished from other AUGs across reading frames solely by using a score that quantifies PDCUB for the 300 nt immediately following the AUG, thus further validating that PDCUB is a distinguishing feature of many start codons. We also observe a substantial difference between the ribosomal occupancy for higher-PDCUB transcripts compared to those with lower PDCUB. These observations are consistent with the hypothesis that human translational machinery, including ribosomes and translation initiation factors, rely broadly upon positionally dependent sequence patterns to direct and regulate multiple aspects of translation.

As discussed in previous studies in yeast [13], we tested whether our observed PDCUB pattern could be explained by signal peptides and found a 1-4% enrichment of signal peptides in proteins encoded by high-significance PDCUB transcripts (Figure 12). Sorting predictions by PDCUB quintile revealed a rough correlation between PDCUB quintile and predicted signal peptide density, but even the most overrepresented quintile showed $< 5\%$ enrichment in signal peptides relative to the global background (Figure 14). Interestingly, despite this correlation, both SignalP and PrediSi predict a maximum overrepresentation of signal peptides in the fourth rather than fifth PDCUB quintile. This might be explained by the fourth quintile being uniquely composed of a single overrepresented GO term related to GPCRs, which are known to have a high (≥ 5%) rate of occurrence of signal peptides [22]. Overall, the low predicted enrichment of signal peptides in all PDCUB quintiles suggests that while sequence features coding for signal peptides may contribute slightly to PDCUB, they are not its driving mechanism. It is also worth noting that while signal peptides are defined by an amino acid sequence, PDCUB describes position-dependent patterns at the nucleotide level in addition to the triplet level, and thus may capture translationally important patterns that position-dependent amino acid signals alone cannot completely resolve.

Previous research has shown that GC-rich codons have greater translational efficiency [23]. We observe that the pattern of PDCUB across transcripts shows enrichment of GC content in codons at the beginning of coding regions relative to downstream segments. Simultaneously, we observe lower ribosomal occupancy over the region in which PDCUB is calculated (1-300 nt downstream of the start codon). Therefore, we speculate that observed PDCUB patterns may lead to greater efficiency of translation initiation and possibly also translation elongation. Selection for GC-rich, translationally efficient coding regions may encourage the development of PDCUB patterns, or vice versa. However, the fact that PDCUB outperforms GC content in the identification of start codons suggests that GC content is not the whole story, and higher-order sequence patterns may be at play.

GC3 codons appear to have a similar effect on PDCUB as GC content. We found that codons that became depleted after shuffling had a ‘G’ or ‘C’ in the 3rd position. We hypothesize that the position-dependent enrichment and depletion of GC3 codons are a driving mechanism behind the PDCUB pattern. However, PDCUB outperforms GC3 content in the identification of start codons, which suggests, much like GC content, GC3 content does not alone account for the PDCUB pattern.

We observed that PDCUB defines a spectrum of translational efficiency, as seen in both ribosome occupancy and local tAI trajectories. Previous studies have been conflicted about the presence of a translational ramp in mammals, with some reporting no ramp [24] and others reporting a distinct ramp [20]. We find that both results are true depending on where one looks within the PDCUB spectrum, with the highest-PDCUB transcripts having a strong initial ramp, intermediate PDCUB transcripts having a weaker ramp or no ramp, and the lowest-PDCUB transcripts exhibiting an initial translational valley that, to our knowledge, has not been described before. In particular, high-PDCUB transcripts seem to exhibit the same translational ramp described elsewhere [20, 25], and this translational ramp corresponds to reduced ribosomal occupancy in the early regions of those same transcripts. In contrast, tAI trajectories plotted against global CAI quintiles show a much weaker ramp and no valley (Figure 17). When we alter the CAI calculation to only include the first 100 codons of each transcript (the same window in which PDCUB is calculated) the altered trajectories display a much weaker valley (more of a delayed ramp) and an initial ramp profile that begins to more closely (but still weakly) resemble PDCUB, while downstream regions display poor separation compared to their PDCUB and global CAI equivalents. The low start codon predictive potential of CAI computed from the first 100 codons is further evidence of the limited ability of CAI to capture the dynamic patterns of codon usage at the start of coding regions, and that CAI captures other features of translational control distinct from PDCUB. All of this indicates that PDCUB dramatically improves on CAI in its ability to detect important early regions of variable translation efficiency, while retaining the ability of

CAI to clearly identify stable downstream regimes of translational efficiency – a balance which modified regional CAI fails to achieve, leaving PDCUB as the best of the three options.

Upon quantifying the absolute growth of these tAI trajectories for PDCUB, CAI, and regional CAI quintiles, we see that distinct trends occur for each measure of coding potential (Supplementary Figure 20). Transcripts in the highest PDCUB quintile show the greatest growth in localized tAI, while the second-greatest growth occurs for the lowest quintile, corresponding to emergence from the translational valley. By contrast, for standard CAI we see a monotonic increase in the height of the localized tAI growth across all quintiles, due to a lack of a translational valley for the lowest CAI quintile. This is further evidence that PDCUB captures important information regarding translational efficiency that is not captured by standard applications of CAI and tAI alone, as reinforced by the low direct correlation between these metrics over the entire protein-coding transcriptome (Figures 19-23).

Importantly, the PDCUB-defined transcript subpopulations correspond to transcript function (Supplementary Table 1, PDCUB sheets). In our analysis of GO terms, we observe that functions involved in growth and development at the system, organ, tissue, cellular, and subcellular levels are highly enriched among high-PDCUB transcripts. Among low-PDCUB transcripts, such terms are depleted. This could be explained by the fact that developmental pathways are part of precise regulatory functions that require exquisite tuning of translational control [26, 27] and precise timing [28, 29], and that PDCUB may serve as a mechanism to enhance translational efficiency within regulatory cascades. This may be true particularly at the translation initiation step, thereby accounting for the pronounced translational ramp found among high-PDCUB transcripts and the enhanced ability of PDCUB to detect start codons.

The only exceptions to the trends described above are for developmental terms involving skin and keratin, which are enriched among low-PDCUB transcripts only. This set of exceptions may be tied to the prominent role of these tissues in human immune function relative to other organs and tissues, as they comprise the body’s first and primary barrier against pathogens. More generally, we observe that transcripts related to immune response, as well as those governing detection and perception of chemical stimuli, are statistically depleted among high-PDCUB transcripts and enriched among low-PDCUB transcripts. Immune response and chemical detection pathways have in common high genetic copy number and sequence variation [30]. Moreover, there is an abundance of pseudogenes observed for human odorant receptors and other genes responsible for *detection* and *perception* of chemical stimuli [31], as well as immune genes [32, 33]. These observations may suggest a lower selective pressure on individual protein-coding genes due to the robustness of these pathways conferred by high copy number and a need to constantly adapt to changing chemical [34] and pathogen [35-37] environments. We posit that genes within these pathways might sample a broader range of codon ordering compared to genes involved in the development of complex vital organs. Meanwhile, genes tied to chemical stimulus *response* and *regulation* show a reversed trend relative to detection and perception, with enrichment among high-PDCUB transcripts and depletion among low-PDCUB transcripts. This trend reversal may stem from a need for precise timing among regulatory and response genes, as described for developmental genes above, while detection and perception genes rely on high copy number and low translational control to account for the complex suite of chemicals that must be adapted to in the human sensory environment.

Comparison of these trends to those for CAI quintiles reveals at best partial qualitative similarity for some of the overarching GO categories identified above, with several interesting and significant deviations (Supplementary Table 1, CAI sheets). Most notably, we see enrichment of immune terms among high-CAI transcripts and depletion among mid- and low-CAI transcripts – exactly opposite of the trend seen for PDCUB quintiles. A similar trend reversal is seen for terms relating to skin, which are enriched exclusively in the highest CAI quintile and the lowest PDCUB quintile. Given the prominence of these trend reversals, and paired with the critical role of immune response in human evolution, we postulate that PDCUB and CAI play functionally distinct roles in translational regulation. For instance, it is possible that CAI governs steady-state baseline levels of background translation, with high CAI reserved for those genes that are more consistently active, while high PDCUB correlates to situationally sensitive genes that require precise timing and a more well-defined translational ramp. While these two categories might overlap (as in the development terms), it is also possible for them to diverge sharply (as in the immune terms). Additional evidence for this view is the presence of biologically cyclic GO terms such as “hair cycle” and “molting cycle” in the results of the CAI quintile analysis, but the total absence of such terms from the PDCUB analysis. Further evidence lies in the split described above for stimulus and sensory GO terms. Among PDCUB quintiles the relationship between these terms is neatly split between low- and high-PDCUB transcripts, with detection and perception showing enrichment among low-PDCUB transcripts and depletion among high-PDCUB transcripts, while regulation and response follow the opposite trend. In contrast, CAI quintiles do not show this split and instead have detection and perception terms uniformly enriched, but response and regulation terms uniformly depleted, across all quintiles in which they appear (the exception being quintile 4, which has no stimulus or sensory terms). These results serve as a basis for understanding the complex relationship between PDCUB and other metrics of codon bias, including the extent to which they might be complementary or competing factors in translational efficiency.

MATERIALS AND METHODS

The dataset used was the GENCODE Release 34 protein-coding transcript sequences FASTA [15]. We filtered out all transcripts in which the length of the labelled coding region was not a multiple of three, lacked an in-frame start codon, or lacked an in-frame stop codon. This dataset was used for all of our analyses.

Codon frequencies relative to transcript sequence position

Codons following the start codon were counted in bins of size 60 (20 nucleotides) for a total of 50 bins. To keep the bin size constant, the probability of finding a given codon in a bin was calculated, and subsequently used to derive a mean and standard deviation for each codon. The z-score of a given sense codon in a given bin in Figure 1A was calculated by

$$z_{c,b} = \frac{p_{c,b} - \mu_c}{\sigma_c}$$

where $p_{c,b}$ is the probability of a codon in a given bin, $\mu_c$is the mean of the codon across all bins, and $\sigma_c$is the standard deviation of the codon across all bins.

Hierarchical clustering of codons based on z-score profiles

To obtain the dendrogram of z-score profiles, a distance matrix was first calculated wherein each row and column corresponds to a single codon. A cell at row *m* and column *n* would thus correspond to the distance between the z-score profiles of codons *m* and *n.* The distance between two codons was calculated based on the sum of squares of the differences between z-scores:

$$D_{m,n} = \sum_{b}^{50} (z_{m,b} - z_{n,b})^2$$

where $z_{m,b}$and $z_{n,b}$ are respectively the z-scores of codons *m* and *n* at bin b.

The dendrogram was computed using python's scipy.cluster.hierarchy.linkage method [38], which performs clustering given a distance matrix such as the one calculated above. The method used for calculating distance between two clusters *u* and *v* was the 'average 'method

$$d(u,v) = \sum_{i,j} \frac{d(u[i],v[j])}{|u|*|v|},$$

which describes the distance between two clusters of codon profiles over positions i and j relative to the cardinalities of clusters u and v.

Scoring models for predicting start codons

Several scoring models were developed to predict the location of start codons, in order to determine the effect of position dependence in the human transcriptome. These models were based on a weight matrix, or position specific-scoring matrix (PSSM), where the rows correspond to codons or nucleotides and the columns correspond to positions after the start codon. In general, the PSSM values were calculated as log likelihoods based on a background model.

PDCUB PSSM

The weights in the PDCUB PSSM were calculating according to the equation

$$W_{cb} = log_2\left(\frac{f_{cb}}{p_c}\right)$$

where f_{cb} is the frequency of codon c at bin b, and p_c refers to the background probability of codon c. Bins b of size w (in codons) with start positions s_b comprise a series $\{[s_1, s_1 + w], [s_2, s_2 + w], \dots [s_N, s_N + w]\}$ up to N bins for each putative CDS. We define a putative CDS region as series of triplets after the AUG, $C = (c_1, c_2, \dots, c_n)$. The PDCUB score $S(C)$ is computed for a given CDS according to

$$\mathrm{S(C)} = \sum_{b=1}^{N} \sum_{i=s_b}^{w} W_{c_i,b}$$

with PDCUB models computed at bin sizes of $w =$ 1, 2, 3, 4, 5, 10, 20, 25, 50, and 100 codons. After examining the percentage of transcripts in which the true start codon was the highest scoring AUG, it was determined that a bin size of 3 nucleotides was optimal.

Kozak consensus sequence

The Kozak consensus sequence is defined as (gcc)gccRccAUGG, where R corresponds to A or G, and position +1 corresponds to the A in the AUG. Positions -3 (the R) and +3 (the G after the AUG) are particularly important, and upper-case indicates highly conserved bases [16]. We gave positions +3 and -3 a weight of 6, and the remaining positions a weight of 1. The score of +6 for each highly conserved position ensures that correctly matching one highly conserved position outweighs any number of less-conserved matches. Transcripts were scored based on the number of mismatches the surrounding nucleotides had with the consensus sequence, with a maximum score being 17 (no mismatches). The leading (gcc) was subtracted from this score, so that codons containing the matching sequences were rewarded, but codons without them were not penalized.

Kozak PSSM

The Kozak PSSM model was implemented in a manner similar to that for PDCUB, with a few key differences. A PSSM was built by computing the frequencies $f_{b,i}$ for each nucleotide b at position i relative to the start codon across all protein-coding transcripts. The PSSM is then computed using this frequency relative to the global single-nucleotide frequencies p_b, as shown below.

$$W_{b,i} = log_2\left(\frac{f_{b,i}}{p_b}\right)$$

Therefore, the Kozak PSSM differs from the PDCUB PSSM because the former was based on nucleotide counts instead of codon counts. The positions in the Kozak PSSM correspond to the ten nucleotides before the start codon, which is the location of the Kozak consensus sequence. All positions in the transcript were thus scored by taking the sum of the weights of ten preceding nucleotides.

GC content PSSM

The GC content PSSM has two rows, one corresponding to ‘S ’nucleotides {G,C} and the other to ‘W ’nucleotides {A,U}. Nucleotides were counted into either of those categories in a manner similar to the nucleotide PSSM used for the Kozak weight matrix.

Statistical Null Model

The null model was calculated to determine the apparent extent of random chance involved in an AUG being the true start codon in a given transcript. All AUG triplets in a transcript were scored, and an AUG was randomly selected as the start codon. If the selected AUG had the highest score, then the model was considered successful for that transcript.

Theoretical Model

If the length of a given transcript is L, then there are $L-2$ possible positions in which a triplet can exist when allowing all three possible reading frames. Moreover, the codon at that given position must be an AUG, of which there is a 1/64 chance. Therefore, the total number of expected AUGs in a transcript is

$$\langle N_{AUG} \rangle = (L-2)(\frac{1}{64})$$

The theoretical probability that an AUG selected a random is the true start codon can thus be modeled by

$$\frac{1}{\langle N_{AUG} \rangle} = \frac{64}{L-2}$$

Amino Acid Usage Bias Model

The amino acid PSSM was implemented in a manner similar to the position-dependent models. Each row in the matrix corresponds to an amino acid. When calculating the score for a transcript, the triplet was translated into its corresponding amino acid.

GC3 Model

The GC3 model differs from the GC model in that at each bin, the number of codons with a ‘W’ nucleotide or a ’S’ nucleotide in the 3rd position were counted. When calculating the score for the a transcript, each triplet was evaluated based only on its 3rd nucleotide.

Cross-validation methods

The dataset of all valid sequences was split into three separate sets. In all cases, 80% of the data (about 55,000 transcripts) were used to obtain PSSMs of the model in question. In most cases, the remaining 20% (about 14,000 transcripts) were scored with the calculated PSSM. In the case of PDCUB, where multiple bin sizes were tested, 10% of the data was used to determine the bin size that produced the highest-scoring AUG triplets, and 10% was left for scoring.

Information content of bins within a transcript

The information content IC of position i in a transcript was calculated as follows:

$$IC(i) = H_f - H_p(i)$$

where H_f is the Shannon entropy of the global triplet frequencies (background model) of the transcript, and $H_p(i)$ is the Shannon entropy of the position-dependent codon frequencies at position i. The background entropy is calculated as

$$H_f = -\sum_c f_c log_2(f_c)$$

where f_c is the global frequency of triplet *c.* The product of the global frequency and the log of the global frequency are computed for all triplets and then added together. The Shannon entropy of the position-dependent codon frequencies is calculated as

$$H_p(i) = -\sum_c p_{c,i} log_2(p_{c,i})$$

where $p_{c,i}$ is the probability of codon c at position i.

Determination of high-significance AUG triplets and PDCUB quintiles

Upon generating a list of transcripts and their highest-scoring AUG triplet based on the PDCUB model, the p-values for each AUG score were calculated. All scores were sorted by p-value and accordingly ranked (the lowest p-value would have a rank of 1, the next lowest a rank of 2, and so on). These values were filtered with a Benjamin-Hochberg multiple test correction to produce a list of statistically significant genes ($\alpha < 0.05$), with corresponding q-values calculated according to

$$q_r = \frac{P_r * N}{r}$$

where *r* is the rank and *N* is the total number of scores. All transcripts with a rank lower than the highest rank r with q_r below the significance level were deemed statistically significant. This set was considered the high-significance PDCUB set (significant PDCUB set), while an equivalently sized transcript set with the highest q_r was considered the low-significance PDCUB set.

A second ranking system was also used, wherein all transcripts were divided into a set of quintiles according to PDCUB score. The 20% of transcripts with the highest PDCUB score comprise the top PDCUB quintile and so on, in Table 1.

Analyzing the contribution of signal peptides

Prediction of signal peptides in human proteome

The mRNA content of the human transcriptome was retrieved from GENCODE, converted to protein sequence format, and fed to the online tools PrediSi [18] and SignalP-5.0 [19] for prediction of signal peptides. Both tools were accessed through their public-facing online interfaces using default program parameters. In both cases, transcripts from the significant PDCUB set were separated from the remainder of the GENCODE set.

Statistical enrichment of signal peptides in significant PDCUB set and quintiles

Predicted signal peptide densities for the significant PDCUB set and each PDCUB quintile were compared to background signal peptide enrichment in two batches: one for the PrediSi scores, and another for the SignalP scores. In both cases, the significant PDCUB set and quintile sets were tested for enrichment of signal peptides relative to the global set. Enrichment significance was determined via binomial test.

Overlap between significant PDCUB and signal peptides

Transcripts from the significant PDCUB set were checked for overlap with all GENCODE protein-coding transcripts predicted to have a signal peptide. This was done for two sets of transcripts: GENCODE transcripts that scored positively for signal peptides using PrediSi, and those that scored positively for signal peptides using SignalP-5.0. All three resulting sets (significant PDCUB, PrediSi GENCODE, and SignalP-5.0 GENCODE) were visualized together as a Venn diagram using matplotlib.

Gene Ontology Enrichment

ENSEMBL IDs for transcripts were analyzed for each PDCUB category (high-significance, top quintile, and so on) using the GO Enrichment Analysis tool, which is part of the web-accessible

Gene Ontology Resource (http://geneontology.org/) [39-41]. Tables of statistically significant GO terms were retrieved for both sets of transcripts. The tables are provided as excel files in Supplementary Table 1. Significant GO terms were those with unexpected enrichment or depletion among a given PDCUB category, when a Benjamini-Hochberg multiple test correction with a false discovery rate of 0.05 was applied.

Calculating regional CAI and local tAI

Sets of local relative adaptiveness (RA) and CAI values were calculated for all protein-coding GENCODE human transcripts. Whereas RA weights are traditionally defined by one unique set of values for a given transcriptome, and are derived from the total codon composition across the full lengths of every CDS, our local RA weights were computed for a particular region r of every CDS. These local RA values were used as input weights w_{rc} to calculate regional CAI values for the same region r and codon c of a single transcript. For example, when examining the regional CAI of the first 50 codons in a given CDS, our local RA was derived exclusively from the first 50 codons of every CDS. The local RA was calculated as

$$w_{rc} = \frac{f_{rc}}{max(f_{rc'})}$$

where f_{rc} is the frequency of codon c in region r and $f_{rc'}$ is the frequency of codon c' in region r, where c and c' are both synonymous codons for the same amino acid. From this, the regional CAI was calculated as

$$CAI_r = \left(\prod_{k=1}^{l_r} w_{rc_{kr}}\right)^{1/l_r}$$

where l_r is the length of the transcript region, and c_{kr} is the codon defined by the k^{th} triplet in that region.

To calculate local tAI, we retrieved human codon tAI weights from Table S1 of Tuller *et al.* [20] and computed a similar geometric mean for the same regions as above. Quintile tAI figures were generated by splitting GENCODE transcripts into five parts sorted by PDCUB score. Unlike regional CAI, which consists of a single average value for the entire region under consideration, average local tAI was plotted for each quintile as a function of codon position using a sliding window five codons in length.

Ribosome profiling

Ribosome profiling of HEK293 cells was downloaded from GSE accession number GSE102720 [42]. Ribo-seq reads were aligned to protein-coding transcripts using Bowtie [43] with the parameters "-v 2 -a -S --strata --best --norc -m 200" after trimming the adaptors with Cutadapt [44]. Read 5′ ends were shifted relative to the start of the CDS and an additional 15 nucleotides to align to the A site [45]. Normalization of each individual transcript was performed by dividing the sum of the reads at each nucleotide position by the total number of reads mapping up to 500 codons from the start codon for that transcript, and converted to a percentage. The sum at each position of these normalized transcripts was divided by the total number of transcripts to create an average profile. The average curves for each quintile were smoothed using a Savitzky-Golay filter [46] with a window size of 19 and a polynomial degree of 4 using the scipy implementation [38].

Project code

Code for this project is available at https://github.com/hendrixlab/PDCUB.

References

1. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S: **The landscape of long noncoding RNAs in the human transcriptome**. *Nature genetics* 2015, **47**(3):199-208.
2. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES: **Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins**. *Cell* 2013, **154**(1):240-251.
3. Bánfai B, Jia H, Khatun J, Wood E, Risk B, Gundling WE, Kundaje A, Gunawardena HP, Yu Y, Xie L: **Long noncoding RNAs are rarely translated in two human cell lines**. *Genome research* 2012, **22**(9):1646-1657.
4. Ji Z, Song R, Regev A, Struhl K: **Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins**. *elife* 2015, **4**:e08890.
5. Li L-J, Leng R-X, Fan Y-G, Pan H-F, Ye D-Q: **Translation of noncoding RNAs: focus on lncRNAs, pri-miRNAs, and circRNAs**. *Experimental cell research* 2017, **361**(1):1-8.
6. Makarewich CA, Olson EN: **Mining for micropeptides**. *Trends in cell biology* 2017, **27**(9):685-696.
7. Anderson DM, Anderson KM, Chang C-L, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R: **A micropeptide encoded by a putative long noncoding RNA regulates muscle performance**. *Cell* 2015, **160**(4):595-606.
8. Zhang H, Zhang L, Li Z, Liu K, Liu B, Mathews DH, Huang L: **LinearDesign: Efficient Algorithms for Optimized mRNA Sequence Design**. *arXiv preprint arXiv:200410177* 2020.

9. Kozak M: **Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs**. *Nucleic acids research* 1984, **12**(2):857-872.

10. Wright G, Rodriguez A, Li J, Clark PL, Milenković T, Emrich SJ: **Analysis of computational codon usage models and their association with translationally slow codons**. *PloS one* 2020, **15**(4):e0232003.

11. Sharp PM, Li W-H: **The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications**. *Nucleic acids research* 1987, **15**(3):1281-1295.

12. Hockenberry AJ, Sirer MI, Amaral LAN, Jewett MC: **Quantifying position-dependent codon usage bias**. *Molecular biology and evolution* 2014, **31**(7):1880-1893.

13. Villada JC, Brustolini OJB, Batista da Silveira W: **Integrated analysis of individual codon contribution to protein biosynthesis reveals a new approach to improving the basis of rational gene design**. *DNA Research* 2017, **24**(4):419-434.

14. Hill ST, Kuintzle R, Teegarden A, Merrill III E, Danaee P, Hendrix DA: **A deep recurrent neural network discovers complex biological rules to decipher RNA protein-coding potential**. *Nucleic acids research* 2018, **46**(16):8105-8113.

15. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J: **GENCODE reference annotation for the human and mouse genomes**. *Nucleic acids research* 2019, **47**(D1):D766-D773.

16. Kozak M: **An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs**. *Nucleic acids research* 1987, **15**(20):8125-8148.

17. Saboory E, Ghasemi M, Mehranfard N: **Norepinephrine, neurodevelopment and behavior**. *Neurochemistry international* 2020, **135**:104706.

18. Hiller K, Grote A, Scheer M, Münch R, Jahn D: **PrediSi: prediction of signal peptides and their cleavage positions**. *Nucleic acids research* 2004, **32**(suppl_2):W375-W379.

19. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H: **SignalP 5.0 improves signal peptide predictions using deep neural networks**. *Nature biotechnology* 2019, **37**(4):420-423.

20. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y: **An evolutionarily conserved mechanism for controlling the efficiency of protein translation**. *Cell* 2010, **141**(2):344-354.

21. Ingolia NT: **Ribosome profiling: new views of translation, from single codons to genome scale**. *Nature reviews genetics* 2014, **15**(3):205-213.

22. Schülein R, Westendorf C, Krause G, Rosenthal W: **Functional significance of cleavable signal peptides of G protein-coupled receptors**. *European journal of cell biology* 2012, **91**(4):294-299.

23. Hia F, Yang SF, Shichino Y, Yoshinaga M, Murakawa Y, Vandenbon A, Fukao A, Fujiwara T, Landthaler M, Natsume T: **Codon bias confers stability to human mRNA s**. *EMBO reports* 2019, **20**(11):e48220.

24. Ingolia NT, Lareau LF, Weissman JS: **Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes**. *Cell* 2011, **147**(4):789-802.

25. Verma M, Choi J, Cottrell KA, Lavagnino Z, Thomas EN, Pavlovic-Djuranovic S, Szczesny P, Piston DW, Zaher HS, Puglisi JD: **A short translational ramp determines the efficiency of protein synthesis**. *Nature communications* 2019, **10**(1):1-15.

26. Fujii K, Shi Z, Zhulyn O, Denans N, Barna M: **Pervasive translational regulation of the cell signalling circuitry underlies mammalian development**. *Nature communications* 2017, **8**(1):1-13.

27. Hershey JW, Sonenberg N, Mathews MB: **Principles of translational control: an overview**. *Cold Spring Harbor perspectives in biology* 2012, **4**(12):a011528.

28. Bolouri H, Davidson EH: **Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics**. *Proceedings of the National Academy of Sciences* 2003, **100**(16):9371-9376.

29. Daou R, Beißbarth T, Wingender E, Gültas M, Haubrock M: **Constructing temporal regulatory cascades in the context of development and cell differentiation**. *PloS one* 2020, **15**(4):e0231326.

30. De Smith A, Walters R, Froguel P, Blakemore A: **Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease**. *Cytogenetic and genome research* 2008, **123**(1-4):17-26.

31. Spehr M, Munger SD: **Olfactory receptors: G protein-coupled receptors and beyond**. *Journal of neurochemistry* 2009, **109**(6):1570-1583.

32. Vargas-Madrazo E, Almagro JC, Lara-Ochoa F: **Structural repertoire in VHpseudogenes of immunoglobulins: comparison with human germline genes and human amino acid sequences**. *Journal of molecular biology* 1995, **246**(1):74-81.

33. Ortutay C, Siermala M, Vihinen M: **Molecular characterization of the immune system: emergence of proteins, processes, and domains**. *Immunogenetics* 2007, **59**(5):333-348.

34. Niimura Y: **Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents**. *Human genomics* 2009, **4**(2):1-12.

35. Broecker F, Moelling K: **Evolution of immune systems from viruses and transposable elements**. *Frontiers in microbiology* 2019, **10**:51.

36. Cooper MD, Alder MN: **The evolution of adaptive immune systems**. *Cell* 2006, **124**(4):815-822.

37. Quintana-Murci L: **Human immunology through the lens of evolutionary genetics**. *Cell* 2019, **177**(1):184-199.

38. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J: **SciPy 1.0: fundamental algorithms for scientific computing in Python**. *Nature methods* 2020, **17**(3):261-272.

39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene ontology: tool for the unification of biology**. *Nature genetics* 2000, **25**(1):25-29.

40. **The Gene Ontology resource: enriching a GOld mine**. *Nucleic Acids Research* 2021, **49**(D1):D325-D334.

41. Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD: **PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API**. *Nucleic Acids Research* 2021, **49**(D1):D394-D403.

42. Iwasaki S, Iwasaki W, Takahashi M, Sakamoto A, Watanabe C, Shichino Y, Floor SN, Fujiwara K, Mito M, Dodo K: **The translation inhibitor rocaglamide targets a bimolecular cavity between eIF4A and polypurine RNA**. *Molecular cell* 2019, **73**(4):738-748. e739.

43. Langmead B: **Aligning short sequencing reads with Bowtie**. *Current protocols in bioinformatics* 2010, **32**(1):11.17. 11-11.17. 14.

44. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads**. *EMBnet journal* 2011, **17**(1):10-12.

45. Guo H, Ingolia NT, Weissman JS, Bartel DP: **Mammalian microRNAs predominantly act to decrease target mRNA levels**. *Nature* 2010, **466**(7308):835-840.

46. Savitzky A, Golay MJ: **Smoothing and differentiation of data by simplified least squares procedures**. *Analytical chemistry* 1964, **36**(8):1627-1639.

Tables:

Table 1: Definition of PDCUB quintiles.

PDCUB quintile	**percentile range (%)**	**quintile name**
5	81-100	fifth quintile, top quintile, highest PDCUB
4	61-80	fourth quintile
3	41-60	third quintile, middle quintile, moderate PDCUB
2	21-40	second quintile

1	01-20	first quintile, bottom quintile, lowest PDCUB

Figure Captions:

Figure 1: **A.** A “bird’s eye view” of position-dependent codon usage bias in the human transcriptome (GENCODE protein-coding transcript models). A heatmap and dendrogram visualizing the z-scores computed from codon occurrence frequency in bins of size 60 nt. **B.** Curves show the z-scores computed from codon occurrence frequencies, shown in clusters corresponding to the number of G/C nucleotides within the codon.

Figure 2: **A.** A heatmap demonstrating the optimal bin size and number of bins within the first 300 nt of CDS regions of the human transcriptome favors 100 3-nt bins. **B.** A codon z-score heatmap weight hierarchical clustering for the 300 nt after the start codon **C.** The information content of all transcripts, computed from position-dependent codon frequencies shows a region of enriched information content relative to the background trinucleotide frequencies. Red line segments indicate some regions of enriched information content. The inset shows a histogram of the information content for the first 100 codons with a tail of enriched information content, shown in red.

Figure 3. **A.** The rate of start codon identification compared for different models. The curves show what fraction of start codons are identified per transcript, comparing the PDCUB score computed from a PSSM compared to the Kozak consensus sequence AUGGCC, a PSSM computed from the positions -10 to +2 relative to the start codon for all transcripts, a PSSM computed from the position-dependent GC content within CDS regions relative to global GC content, a PSSM

computed from single-nucleotide frequencies in CDS regions relative to global nucleotide frequencies, a random selection model. **B.** The PDCUB score plotted as a function of position, starting at each codon, and plotted separately for each reading frame, for an example short (~1000nt) transcript. Blue shaded area beginning at the start codon indicates the CDS region.

Figure 4: **A.** Scatterplot representing statistically enriched GO terms associated with transcripts in the lowest PDCUB score quintile. Each dot corresponds to an enriched GO term, the x-axis position corresponds to the expected number of occurrences, the y-axis represents the observe frequency within the bottom PDCUB quintile. Triangles by the labels indicate enriched (pointed up) and depleted (down) GO terms. **B.** Scatterplot representing statistically enriched GO terms associated with transcripts in the highest PDCUB score quintiles, displayed similarly as A. **C.** Local tAI profiles for transcripts in quintiles defined for PDCUB score. **D.** The shape of ribosomal occupancy, shown as the normalized Ribo-seq profile centered at the start codon at position 0 for quintiles defined for PDCUB. Read positions are A-shifted by 15-nt, and smoothed using a Savitzky-Golay filter.

Supplementary Table 1

Table of significant Gene Ontology terms for all PDCUB quintiles, significant PDCUB transcripts, non-significant PDCUB transcripts, CAI quintiles, and universally expressed trancripts. Colored rows correspond to a group of biological functions marked in the legend.

Legend

🟪	🟦	🟩	🟨	🟧
Signaling and transport	Growth and development	Immune response and defense	Chemical stimulus and sensation (detection and perception)	Chemical stimulus and sensation (response and regulation)

80 – 100 PDCUB Quintile (Highest)

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (2175)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
nephron tubule formation	17	10	1.8	+	5.57	1.14E-04	1.77E-02
regulation of norepinephrine secretion (GO:0014061)	16	9	1.69	+	5.33	3.22E-04	3.77E-02
nucleosome assembly (GO:0006334)	101	38	10.67	+	3.56	1.59E-09	4.15E-06
positive regulation of blood pressure (GO:0045777)	40	15	4.22	+	3.55	1.42E-04	2.08E-02
positive regulation of amine transport (GO:0051954)	38	14	4.01	+	3.49	2.72E-04	3.45E-02
non-canonical Wnt signaling pathway (GO:0035567)	52	18	5.49	+	3.28	7.42E-05	1.32E-02
mesenchyme morphogenesis (GO:0072132)	51	17	5.39	+	3.16	1.69E-04	2.40E-02
neuropeptide signaling pathway (GO:0007218)	109	33	11.51	+	2.87	1.76E-06	8.36E-04
nucleosome organization (GO:0034728)	140	42	14.79	+	2.84	6.29E-08	6.18E-05
chromatin assembly (GO:0031497)	134	40	14.15	+	2.83	1.44E-07	1.19E-04
collagen fibril organization (GO:0030199)	57	17	6.02	+	2.82	4.98E-04	4.98E-02
canonical Wnt signaling pathway (GO:0060070)	79	23	8.34	+	2.76	1.10E-04	1.73E-02
neuron fate commitment (GO:0048663)	62	18	6.55	+	2.75	4.51E-04	4.66E-02
regulation of amine transport (GO:0051952)	92	26	9.72	+	2.68	4.73E-05	9.07E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (2175)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
somite development (GO:0061053)	78	22	8.24	+	2.67	2.00E-04	2.73E-02
potassium ion transmembrane transport (GO:0071805)	158	44	16.69	+	2.64	1.83E-07	1.44E-04
cell fate specification (GO:0001708)	81	22	8.55	+	2.57	2.78E-04	3.49E-02
potassium ion transport (GO:0006813)	167	45	17.64	+	2.55	3.84E-07	2.74E-04
endocrine system development (GO:0035270)	123	33	12.99	+	2.54	1.26E-05	3.80E-03
chromatin assembly or disassembly (GO:0006333)	154	41	16.26	+	2.52	1.73E-06	9.08E-04
metanephros development (GO:0001656)	79	21	8.34	+	2.52	4.72E-04	4.82E-02
bone morphogenesis (GO:0060349)	96	25	10.14	+	2.47	2.39E-04	3.18E-02
adenylate cyclase-activating G protein-coupled receptor signaling pathway (GO:0007189)	136	35	14.36	+	2.44	1.77E-05	4.79E-03
cardiac chamber morphogenesis (GO:0003206)	119	30	12.57	+	2.39	1.08E-04	1.78E-02
protein-DNA complex assembly (GO:0065004)	167	42	17.64	+	2.38	4.15E-06	1.63E-03
DNA packaging (GO:0006323)	180	45	19.01	+	2.37	1.74E-06	8.82E-04
spinal cord development (GO:0021510)	100	25	10.56	+	2.37	3.35E-04	3.90E-02
regulation of tube size (GO:0035150)	139	34	14.68	+	2.32	4.24E-05	8.43E-03
blood vessel diameter maintenance (GO:0097746)	138	33	14.57	+	2.26	1.08E-04	1.74E-02
regulation of tube diameter (GO:0035296)	138	33	14.57	+	2.26	1.08E-04	1.72E-02
digestive tract development (GO:0048565)	126	30	13.31	+	2.25	1.79E-04	2.52E-02
embryonic epithelial tube formation (GO:0001838)	123	29	12.99	+	2.23	2.63E-04	3.38E-02
cell fate commitment (GO:0045165)	238	55	25.13	+	2.19	1.05E-06	5.88E-04
epithelial tube formation (GO:0072175)	131	30	13.83	+	2.17	3.66E-04	4.11E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (2175)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
response to glucocorticoid (GO:0051384)	149	34	15.74	+	2.16	1.90E-04	2.65E-02
digestive system development (GO:0055123)	136	31	14.36	+	2.16	2.92E-04	3.58E-02
connective tissue development (GO:0061448)	202	46	21.33	+	2.16	1.06E-05	3.32E-03
cartilage development (GO:0051216)	152	34	16.05	+	2.12	2.21E-04	3.00E-02
response to estradiol (GO:0032355)	144	32	15.21	+	2.1	4.04E-04	4.44E-02
protein-DNA complex subunit organization (GO:0071824)	208	46	21.97	+	2.09	2.16E-05	5.56E-03
response to corticosteroid (GO:0031960)	168	37	17.74	+	2.09	1.68E-04	2.41E-02
central nervous system neuron differentiation (GO:0021953)	184	40	19.43	+	2.06	1.28E-04	1.94E-02
adenylate cyclase-modulating G protein-coupled receptor signaling pathway (GO:0007188)	223	48	23.55	+	2.04	2.56E-05	5.84E-03
skeletal system morphogenesis (GO:0048705)	223	48	23.55	+	2.04	2.56E-05	5.75E-03
vascular process in circulatory system (GO:0003018)	260	54	27.46	+	1.97	2.53E-05	5.85E-03
muscle organ development (GO:0007517)	284	58	29.99	+	1.93	1.61E-05	4.61E-03
sensory organ morphogenesis (GO:0090596)	262	53	27.67	+	1.92	4.32E-05	8.49E-03
regulation of neuron differentiation (GO:0045664)	193	39	20.38	+	1.91	4.49E-04	4.68E-02
embryonic organ morphogenesis (GO:0048562)	292	59	30.84	+	1.91	2.05E-05	5.36E-03
heart morphogenesis (GO:0003007)	238	48	25.13	+	1.91	1.05E-04	1.74E-02
striated muscle tissue development (GO:0014706)	288	58	30.42	+	1.91	2.71E-05	6.01E-03
muscle tissue development (GO:0060537)	303	60	32	+	1.88	2.84E-05	6.19E-03
DNA conformation change (GO:0071103)	269	53	28.41	+	1.87	7.98E-05	1.38E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (2175)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
mesenchyme development (GO:0060485)	224	44	23.66	+	1.86	3.42E-04	3.93E-02
regionalization (GO:0003002)	316	61	33.37	+	1.83	4.16E-05	8.49E-03
blood circulation (GO:0008015)	399	77	42.14	+	1.83	4.06E-06	1.63E-03
pattern specification process (GO:0007389)	425	81	44.88	+	1.8	3.67E-06	1.56E-03
chromatin remodeling (GO:0006338)	247	47	26.09	+	1.8	4.27E-04	4.54E-02
embryonic organ development (GO:0048568)	433	82	45.73	+	1.79	4.41E-06	1.69E-03
ossification (GO:0001503)	268	50	28.3	+	1.77	4.83E-04	4.90E-02
embryonic morphogenesis (GO:0048598)	568	105	59.99	+	1.75	4.84E-07	3.17E-04
regulation of system process (GO:0044057)	553	102	58.4	+	1.75	7.10E-07	4.46E-04
positive regulation of secretion (GO:0051047)	316	58	33.37	+	1.74	2.45E-04	3.24E-02
regulation of small molecule metabolic process (GO:0062012)	330	60	34.85	+	1.72	2.35E-04	3.15E-02
circulatory system process (GO:0003013)	490	89	51.75	+	1.72	7.66E-06	2.74E-03
sensory organ development (GO:0007423)	558	101	58.93	+	1.71	2.11E-06	9.74E-04
camera-type eye development (GO:0043010)	318	57	33.58	+	1.7	4.99E-04	4.96E-02
forebrain development (GO:0030900)	387	69	40.87	+	1.69	1.39E-04	2.06E-02
muscle structure development (GO:0061061)	477	85	50.38	+	1.69	2.19E-05	5.56E-03
eye development (GO:0001654)	366	65	38.65	+	1.68	2.46E-04	3.22E-02
regulation of ion transmembrane transport (GO:0034765)	481	85	50.8	+	1.67	2.45E-05	5.76E-03
skeletal system development (GO:0001501)	478	84	50.48	+	1.66	4.19E-05	8.45E-03
visual system development (GO:0150063)	370	65	39.08	+	1.66	2.71E-04	3.46E-02
sensory system development (GO:0048880)	376	66	39.71	+	1.66	2.94E-04	3.58E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (2175)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
head development (GO:0060322)	806	139	85.12	+	1.63	2.30E-07	1.72E-04
brain development (GO:0007420)	761	130	80.37	+	1.62	9.09E-07	5.50E-04
animal organ morphogenesis (GO:0009887)	957	163	101.07	+	1.61	3.75E-08	3.93E-05
regulation of secretion (GO:0051046)	622	105	65.69	+	1.6	1.60E-05	4.67E-03
regulation of ion transport (GO:0043269)	685	115	72.34	+	1.59	9.31E-06	2.99E-03
morphogenesis of an epithelium (GO:0002009)	437	73	46.15	+	1.58	4.37E-04	4.61E-02
regulation of transmembrane transport (GO:0034762)	569	95	60.09	+	1.58	7.06E-05	1.29E-02
central nervous system development (GO:0007417)	1002	167	105.82	+	1.58	9.38E-08	8.67E-05
regulation of secretion by cell (GO:1903530)	564	94	59.56	+	1.58	8.71E-05	1.47E-02
tissue morphogenesis (GO:0048729)	542	90	57.24	+	1.57	1.14E-04	1.76E-02
embryo development (GO:0009790)	991	162	104.66	+	1.55	4.22E-07	2.88E-04
regulation of hormone levels (GO:0010817)	518	84	54.71	+	1.54	4.17E-04	4.56E-02
chordate embryonic development (GO:0043009)	613	99	64.74	+	1.53	1.27E-04	1.94E-02
negative regulation of developmental process (GO:0051093)	877	140	92.62	+	1.51	9.06E-06	2.97E-03
embryo development ending in birth or egg hatching (GO:0009792)	633	100	66.85	+	1.5	2.60E-04	3.37E-02
neuron differentiation (GO:0030182)	1008	159	106.45	+	1.49	3.86E-06	1.60E-03
negative regulation of cell differentiation (GO:0045596)	641	101	67.69	+	1.49	2.80E-04	3.49E-02
response to lipid (GO:0033993)	829	130	87.55	+	1.48	3.95E-05	8.16E-03
response to organic cyclic compound (GO:0014070)	896	140	94.62	+	1.48	2.30E-05	5.65E-03
tissue development (GO:0009888)	1597	247	168.66	+	1.46	2.19E-08	2.46E-05
tube development (GO:0035295)	845	130	89.24	+	1.46	9.18E-05	1.54E-02
anatomical structure formation involved in	881	135	93.04	+	1.45	7.99E-05	1.37E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (2175)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
morphogenesis (GO:0048646)							
generation of neurons (GO:0048699)	1237	189	130.64	+	1.45	2.47E-06	1.11E-03
cell-cell signaling (GO:0007267)	1030	157	108.78	+	1.44	2.41E-05	5.75E-03
negative regulation of multicellular organismal process (GO:0051241)	991	151	104.66	+	1.44	3.42E-05	7.26E-03
neurogenesis (GO:0022008)	1361	206	143.73	+	1.43	1.75E-06	8.60E-04
response to oxygen-containing compound (GO:1901700)	1567	233	165.49	+	1.41	9.55E-07	5.56E-04
regulation of multicellular organismal development (GO:2000026)	1294	192	136.66	+	1.4	1.12E-05	3.46E-03
response to organonitrogen compound (GO:0010243)	984	146	103.92	+	1.4	1.46E-04	2.10E-02
anatomical structure morphogenesis (GO:0009653)	2130	315	224.95	+	1.4	1.22E-08	1.92E-05
nervous system development (GO:0007399)	2193	322	231.6	+	1.39	1.51E-08	1.97E-05
animal organ development (GO:0048513)	3079	450	325.17	+	1.38	1.21E-11	9.50E-08
response to endogenous stimulus (GO:0009719)	1342	195	141.73	+	1.38	3.41E-05	7.34E-03
positive regulation of multicellular organismal process (GO:0051240)	1412	201	149.12	+	1.35	7.39E-05	1.34E-02
response to abiotic stimulus (GO:0009628)	1125	160	118.81	+	1.35	4.27E-04	4.56E-02
regulation of multicellular organismal process (GO:0051239)	2578	363	272.26	+	1.33	1.03E-07	8.95E-05
regulation of transport (GO:0051049)	1732	242	182.91	+	1.32	3.56E-05	7.46E-03
cell differentiation (GO:0030154)	3403	469	359.38	+	1.31	7.80E-09	1.36E-05
system development (GO:0048731)	4169	574	440.28	+	1.3	6.28E-11	3.29E-07
cellular developmental process (GO:0048869)	3463	474	365.72	+	1.3	1.39E-08	1.99E-05
regulation of cell population proliferation (GO:0042127)	1650	224	174.25	+	1.29	3.51E-04	3.97E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (2175)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
multicellular organism development (GO:0007275)	4525	610	477.88	+	1.28	3.05E-10	9.60E-07
regulation of developmental process (GO:0050793)	2368	319	250.08	+	1.28	2.30E-05	5.73E-03
response to organic substance (GO:0010033)	2688	360	283.87	+	1.27	8.91E-06	2.98E-03
anatomical structure development (GO:0048856)	5001	668	528.15	+	1.26	1.11E-10	4.38E-07
regulation of localization (GO:0032879)	2709	360	286.09	+	1.26	1.73E-05	4.77E-03
regulation of cell communication (GO:0010646)	3242	421	342.38	+	1.23	1.96E-05	5.23E-03
regulation of signaling (GO:0023051)	3257	422	343.97	+	1.23	2.33E-05	5.63E-03
regulation of biological quality (GO:0065008)	3705	479	391.28	+	1.22	6.28E-06	2.30E-03
developmental process (GO:0032502)	5565	713	587.71	+	1.21	1.79E-08	2.16E-05
regulation of signal transduction (GO:0009966)	2867	366	302.78	+	1.21	2.94E-04	3.56E-02
multicellular organismal process (GO:0032501)	6584	834	695.32	+	1.2	2.70E-09	6.07E-06
negative regulation of cellular process (GO:0048523)	4785	600	505.34	+	1.19	8.16E-06	2.79E-03
negative regulation of biological process (GO:0048519)	5187	649	547.79	+	1.18	3.29E-06	1.44E-03
cellular process (GO:0009987)	15074	1696	1591.94	+	1.07	1.13E-06	6.15E-04
biological_process (GO:0008150)	17698	1935	1869.05	+	1.04	7.43E-05	1.31E-02
Unclassified (UNCLASSIFIED)	2897	240	305.95	-	0.78	7.43E-05	1.30E-02
immune system process (GO:0002376)	2370	183	250.29	-	0.73	7.73E-06	2.70E-03
defense response to other organism (GO:0098542)	989	69	104.45	-	0.66	3.95E-04	4.41E-02
microtubule-based process (GO:0007017)	765	50	80.79	-	0.62	4.21E-04	4.56E-02
regulation of immune response (GO:0050776)	905	58	95.58	-	0.61	6.67E-05	1.25E-02
immune response (GO:0006955)	1678	104	177.21	-	0.59	4.69E-09	9.22E-06
detection of stimulus involved in sensory	558	33	58.93	-	0.56	4.90E-04	4.94E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (2175)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
perception (GO:0050906)							
sensory perception of chemical stimulus (GO:0007606)	542	31	57.24	-	0.54	3.07E-04	3.62E-02
microtubule cytoskeleton organization (GO:0000226)	521	29	55.02	-	0.53	2.98E-04	3.57E-02
detection of chemical stimulus (GO:0009593)	524	27	55.34	-	0.49	6.72E-05	1.24E-02
positive regulation of immune response (GO:0050778)	563	28	59.46	-	0.47	1.39E-05	4.13E-03
detection of chemical stimulus involved in sensory perception (GO:0050907)	488	24	51.54	-	0.47	5.03E-05	9.53E-03
immune response-regulating signaling pathway (GO:0002764)	363	17	38.34	-	0.44	2.83E-04	3.50E-02
sensory perception of smell (GO:0007608)	468	21	49.42	-	0.42	1.63E-05	4.57E-03
regulation of chromosome organization (GO:0033044)	286	12	30.2	-	0.4	4.65E-04	4.78E-02
activation of immune response (GO:0002253)	338	14	35.7	-	0.39	1.08E-04	1.76E-02
immune response-regulating cell surface receptor signaling pathway (GO:0002768)	290	12	30.63	-	0.39	3.50E-04	3.99E-02
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	18	46.78	-	0.38	6.02E-06	2.25E-03
cytokine-mediated signaling pathway (GO:0019221)	353	14	37.28	-	0.38	4.71E-05	9.14E-03
antigen receptor-mediated signaling pathway (GO:0050851)	216	7	22.81	-	0.31	3.36E-04	3.88E-02
cell recognition (GO:0008037)	228	7	24.08	-	0.29	1.30E-04	1.94E-02
B cell mediated immunity (GO:0019724)	183	5	19.33	-	0.26	4.01E-04	4.44E-02
adaptive immune response (GO:0002250)	640	15	67.59	-	0.22	1.18E-13	1.86E-09
immunoglobulin production (GO:0002377)	155	3	16.37	-	0.18	1.97E-04	2.72E-02
humoral immune response mediated by circulating	128	2	13.52	-	0.15	4.44E-04	4.66E-02

GO biological process complete	**Homo sapiens - REFLIST (20595)**	**upload_1 (2175)**	**upload_1 (expected)**	**upload_1 (over/under)**	**upload_1 (fold Enrichment)**	**upload_1 (raw P-value)**	**upload_1 (FDR)**
immunoglobulin (GO:0002455)							
B cell receptor signaling pathway (GO:0050853)	130	2	13.73	-	0.15	3.03E-04	3.60E-02
complement activation (GO:0006956)	139	2	14.68	-	0.14	1.44E-04	2.10E-02
phagocytosis, recognition (GO:0006910)	107	1	11.3	-	0.09	4.26E-04	4.58E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

60-80 PDCUB Quintile

GO biological process complete	**Homo sapiens - REFLIST (20595)**	**upload_1 (1661)**	**upload_1 (expected)**	**upload_1 (over/under)**	**upload_1 (fold Enrichment)**	**upload_1 (raw P-value)**	**upload_1 (FDR)**
G protein-coupled receptor signaling pathway (GO:0007186)	1216	156	98.07	+	1.59	9.68E-08	1.52E-03
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

40-60 PDCUB Quintile

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1479)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
eosinophil chemotaxis (GO:0048245)	17	9	1.22	+	7.37	2.92E-05	1.48E-02
positive regulation of peptidyl-serine phosphorylation of STAT protein (GO:0033141)	20	10	1.44	+	6.96	1.54E-05	1.10E-02
eosinophil migration (GO:0072677)	18	9	1.29	+	6.96	4.12E-05	1.85E-02
natural killer cell activation involved in immune response (GO:0002323)	23	10	1.65	+	6.05	3.94E-05	1.87E-02
regulation of peptidyl-serine phosphorylation of STAT protein (GO:0033139)	23	10	1.65	+	6.05	3.94E-05	1.82E-02
monocyte chemotaxis (GO:0002548)	42	17	3.02	+	5.64	1.88E-07	2.69E-04
lymphocyte chemotaxis (GO:0048247)	50	15	3.59	+	4.18	1.96E-05	1.19E-02
natural killer cell activation (GO:0030101)	58	16	4.17	+	3.84	2.51E-05	1.36E-02
chemokine-mediated signaling pathway (GO:0070098)	80	22	5.75	+	3.83	8.50E-07	1.11E-03
cellular response to chemokine (GO:1990869)	90	22	6.46	+	3.4	4.41E-06	4.62E-03
response to chemokine (GO:1990868)	90	22	6.46	+	3.4	4.41E-06	4.33E-03
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	108	31.81	+	3.39	1.74E-24	2.74E-20
neutrophil chemotaxis (GO:0030593)	78	19	5.6	+	3.39	2.02E-05	1.18E-02
granulocyte chemotaxis (GO:0071621)	83	20	5.96	+	3.36	1.42E-05	1.11E-02
sensory perception of smell (GO:0007608)	468	108	33.61	+	3.21	6.98E-23	3.66E-19
detection of chemical stimulus involved in sensory perception (GO:0050907)	488	112	35.05	+	3.2	1.56E-23	1.23E-19
mononuclear cell migration (GO:0071674)	107	24	7.68	+	3.12	5.89E-06	5.45E-03
neutrophil migration (GO:1990266)	88	19	6.32	+	3.01	8.30E-05	3.35E-02
detection of chemical stimulus (GO:0009593)	524	113	37.63	+	3	7.18E-22	2.82E-18

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1479)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
sensory perception of chemical stimulus (GO:0007606)	542	113	38.92	+	2.9	7.08E-21	2.22E-17
granulocyte migration (GO:0097530)	96	20	6.89	+	2.9	8.38E-05	3.29E-02
detection of stimulus involved in sensory perception (GO:0050906)	558	115	40.07	+	2.87	9.43E-21	2.47E-17
leukocyte chemotaxis (GO:0030595)	145	27	10.41	+	2.59	3.67E-05	1.80E-02
detection of stimulus (GO:0051606)	694	120	49.84	+	2.41	2.21E-16	4.34E-13
G protein-coupled receptor signaling pathway (GO:0007186)	1216	183	87.33	+	2.1	6.71E-19	1.51E-15
sensory perception (GO:0007600)	978	146	70.23	+	2.08	7.30E-15	1.27E-11
cytokine-mediated signaling pathway (GO:0019221)	353	49	25.35	+	1.93	4.71E-05	2.05E-02
nervous system process (GO:0050877)	1418	165	101.83	+	1.62	7.87E-09	1.24E-05
system process (GO:0003008)	2040	206	146.5	+	1.41	2.21E-06	2.67E-03
cellular component organization or biogenesis (GO:0071840)	5496	321	394.69	-	0.81	2.13E-05	1.19E-02
negative regulation of biological process (GO:0048519)	5187	301	372.5	-	0.81	2.57E-05	1.35E-02
localization (GO:0051179)	5081	293	364.88	-	0.8	1.96E-05	1.23E-02
cellular component organization (GO:0016043)	5298	302	380.47	-	0.79	3.96E-06	4.45E-03
establishment of localization (GO:0051234)	3911	221	280.86	-	0.79	9.45E-05	3.62E-02
regulation of response to stimulus (GO:0048583)	3805	206	273.25	-	0.75	7.64E-06	6.67E-03
regulation of signal transduction (GO:0009966)	2867	153	205.89	-	0.74	7.45E-05	3.08E-02
regulation of cell communication (GO:0010646)	3242	172	232.82	-	0.74	1.52E-05	1.14E-02
regulation of signaling (GO:0023051)	3257	171	233.9	-	0.73	7.90E-06	6.53E-03
macromolecule localization (GO:0033036)	2308	119	165.75	-	0.72	1.25E-04	4.66E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1479)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
cellular localization (GO:0051641)	2264	111	162.59	-	0.68	1.63E-05	1.06E-02
regulation of cellular component biogenesis (GO:0044087)	950	35	68.22	-	0.51	1.55E-05	1.06E-02
antigen receptor-mediated signaling pathway (GO:0050851)	216	2	15.51	-	0.13	5.93E-05	2.52E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

20-40 PDCUB quintile

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1346)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
termination of RNA polymerase II transcription (GO:0006369)	13	7	0.85	+	8.24	1.23E-04	3.16E-02
snRNA 3'-end processing (GO:0034472)	31	10	2.03	+	4.94	1.42E-04	3.49E-02
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	130	28.95	+	4.49	3.80E-40	5.98E-36
sensory perception of smell (GO:0007608)	468	132	30.59	+	4.32	3.17E-39	1.24E-35
detection of chemical stimulus involved in sensory perception (GO:0050907)	488	136	31.89	+	4.26	6.17E-40	3.23E-36
detection of chemical stimulus (GO:0009593)	524	141	34.25	+	4.12	5.80E-40	4.55E-36
ncRNA 3'-end processing (GO:0043628)	50	13	3.27	+	3.98	9.67E-05	2.67E-02
sensory perception of chemical stimulus (GO:0007606)	542	139	35.42	+	3.92	1.85E-37	5.80E-34

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1346)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
detection of stimulus involved in sensory perception (GO:0050906)	558	139	36.47	+	3.81	2.68E-36	7.02E-33
detection of stimulus (GO:0051606)	694	151	45.36	+	3.33	1.14E-33	2.55E-30
RNA 3'-end processing (GO:0031123)	106	19	6.93	+	2.74	2.19E-04	4.66E-02
sensory perception (GO:0007600)	978	158	63.92	+	2.47	1.05E-22	1.83E-19
G protein-coupled receptor signaling pathway (GO:0007186)	1216	185	79.47	+	2.33	6.40E-24	1.26E-20
nervous system process (GO:0050877)	1418	176	92.67	+	1.9	1.02E-14	1.61E-11
system process (GO:0003008)	2040	198	133.33	+	1.49	8.05E-08	6.66E-05
response to chemical (GO:0042221)	4055	326	265.02	+	1.23	7.98E-05	2.32E-02
cellular metabolic process (GO:0044237)	7029	391	459.38	-	0.85	1.24E-04	3.14E-02
primary metabolic process (GO:0044238)	6869	378	448.93	-	0.84	6.04E-05	1.86E-02
metabolic process (GO:0008152)	7882	428	515.13	-	0.83	1.92E-06	9.42E-04
developmental process (GO:0032502)	5565	299	363.7	-	0.82	9.14E-05	2.56E-02
macromolecule metabolic process (GO:0043170)	5743	308	375.34	-	0.82	5.53E-05	1.74E-02
organic substance metabolic process (GO:0071704)	7404	396	483.89	-	0.82	1.03E-06	5.61E-04
anatomical structure development (GO:0048856)	5001	260	326.84	-	0.8	2.80E-05	9.35E-03
multicellular organism development (GO:0007275)	4525	230	295.73	-	0.78	1.92E-05	7.02E-03
system development (GO:0048731)	4169	210	272.47	-	0.77	2.74E-05	9.35E-03
cellular component organization or biogenesis (GO:0071840)	5496	267	359.19	-	0.74	1.44E-08	1.62E-05
localization (GO:0051179)	5081	246	332.07	-	0.74	5.82E-08	5.08E-05
regulation of biological quality (GO:0065008)	3705	178	242.14	-	0.74	5.57E-06	2.30E-03
transport (GO:0006810)	3752	180	245.21	-	0.73	4.36E-06	1.90E-03
positive regulation of biological process (GO:0048518)	6062	290	396.19	-	0.73	2.74E-10	3.59E-07

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1346)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of multicellular organismal process (GO:0051239)	2578	123	168.49	-	0.73	1.90E-04	4.15E-02
positive regulation of cellular process (GO:0048522)	5580	264	364.68	-	0.72	7.39E-10	8.93E-07
establishment of localization (GO:0051234)	3911	185	255.61	-	0.72	8.68E-07	5.05E-04
positive regulation of cellular metabolic process (GO:0031325)	3186	149	208.22	-	0.72	7.61E-06	3.07E-03
regulation of developmental process (GO:0050793)	2368	110	154.76	-	0.71	1.28E-04	3.20E-02
positive regulation of metabolic process (GO:0009893)	3675	170	240.18	-	0.71	4.63E-07	3.47E-04
organelle organization (GO:0006996)	3360	155	219.6	-	0.71	1.56E-06	8.18E-04
cellular component organization (GO:0016043)	5298	244	346.25	-	0.7	1.60E-10	2.29E-07
cellular localization (GO:0051641)	2264	104	147.97	-	0.7	1.13E-04	3.07E-02
cellular developmental process (GO:0048869)	3463	158	226.33	-	0.7	5.33E-07	3.64E-04
cell differentiation (GO:0030154)	3403	155	222.41	-	0.7	6.58E-07	4.13E-04
cell surface receptor signaling pathway (GO:0007166)	2030	92	132.67	-	0.69	1.67E-04	3.91E-02
regulation of response to stimulus (GO:0048583)	3805	171	248.68	-	0.69	3.95E-08	3.88E-05
positive regulation of nitrogen compound metabolic process (GO:0051173)	2992	134	195.54	-	0.69	1.58E-06	7.99E-04
regulation of catalytic activity (GO:0050790)	2393	107	156.4	-	0.68	2.06E-05	7.37E-03
positive regulation of macromolecule metabolic process (GO:0010604)	3382	151	221.03	-	0.68	1.81E-07	1.43E-04
regulation of molecular function (GO:0065009)	3060	136	199.99	-	0.68	7.14E-07	4.32E-04
catabolic process (GO:0009056)	1882	83	123	-	0.67	1.43E-04	3.46E-02
regulation of signal transduction (GO:0009966)	2867	126	187.37	-	0.67	1.01E-06	5.65E-04

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1346)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of localization (GO:0032879)	2709	119	177.05	-	0.67	1.99E-06	9.46E-04
regulation of transport (GO:0051049)	1732	76	113.2	-	0.67	2.23E-04	4.67E-02
regulation of cell communication (GO:0010646)	3242	140	211.88	-	0.66	4.51E-08	4.17E-05
regulation of signaling (GO:0023051)	3257	140	212.86	-	0.66	3.06E-08	3.20E-05
positive regulation of molecular function (GO:0044093)	1497	63	97.84	-	0.64	1.92E-04	4.13E-02
cell development (GO:0048468)	1623	68	106.07	-	0.64	9.07E-05	2.59E-02
regulation of cellular component organization (GO:0051128)	2359	97	154.17	-	0.63	5.37E-07	3.52E-04
positive regulation of cellular protein metabolic process (GO:0032270)	1366	55	89.28	-	0.62	1.23E-04	3.21E-02
nervous system development (GO:0007399)	2193	88	143.32	-	0.61	4.70E-07	3.36E-04
vesicle-mediated transport (GO:0016192)	1329	52	86.86	-	0.6	7.70E-05	2.28E-02
generation of neurons (GO:0048699)	1237	48	80.84	-	0.59	1.19E-04	3.16E-02
positive regulation of protein metabolic process (GO:0051247)	1455	56	95.09	-	0.59	1.77E-05	6.62E-03
response to abiotic stimulus (GO:0009628)	1125	43	73.53	-	0.58	1.62E-04	3.86E-02
neurogenesis (GO:0022008)	1361	52	88.95	-	0.58	2.72E-05	9.49E-03
plasma membrane bounded cell projection organization (GO:0120036)	1068	40	69.8	-	0.57	1.82E-04	4.09E-02
cell projection organization (GO:0030030)	1119	41	73.13	-	0.56	6.93E-05	2.10E-02
response to organonitrogen compound (GO:0010243)	984	36	64.31	-	0.56	1.75E-04	4.04E-02
intracellular signal transduction (GO:0035556)	1464	53	95.68	-	0.55	2.25E-06	1.04E-03
cytoskeleton organization (GO:0007010)	1169	40	76.4	-	0.52	8.08E-06	3.17E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1346)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of catabolic process (GO:0009894)	956	32	62.48	-	0.51	3.27E-05	1.07E-02
regulation of anatomical structure morphogenesis (GO:0022603)	908	28	59.34	-	0.47	1.03E-05	3.95E-03
membrane organization (GO:0061024)	783	24	51.17	-	0.47	4.83E-05	1.55E-02
actin filament-based process (GO:0030029)	599	13	39.15	-	0.33	3.13E-06	1.41E-03
actin cytoskeleton organization (GO:0030036)	539	11	35.23	-	0.31	5.28E-06	2.24E-03
epithelial tube morphogenesis (GO:0060562)	305	5	19.93	-	0.25	1.83E-04	4.06E-02
phagocytosis (GO:0006909)	249	3	16.27	-	0.18	1.75E-04	3.99E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

0-20 PDCUB quintile (Lowest)

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
negative regulation of execution phase of apoptosis (GO:1900118)	23	11	1.33	+	8.28	1.10E-06	2.28E-04
detection of chemical stimulus involved in sensory perception of bitter taste (GO:0001580)	36	17	2.08	+	8.17	1.52E-09	8.56E-07
sensory perception of bitter taste (GO:0050913)	41	18	2.37	+	7.6	1.26E-09	7.63E-07

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
detection of chemical stimulus involved in sensory perception of taste (GO:0050912)	42	18	2.43	+	7.42	1.71E-09	9.28E-07
regulation of execution phase of apoptosis (GO:1900117)	31	11	1.79	+	6.14	1.10E-05	1.75E-03
sensory perception of taste (GO:0050909)	68	20	3.93	+	5.09	4.19E-08	1.46E-05
keratinization (GO:0031424)	58	17	3.35	+	5.07	4.47E-07	1.10E-04
negative regulation of signaling receptor activity (GO:2000272)	66	14	3.81	+	3.67	9.97E-05	9.99E-03
keratinocyte differentiation (GO:0030216)	108	21	6.24	+	3.37	6.64E-06	1.16E-03
detection of chemical stimulus involved in sensory perception (GO:0050907)	488	91	28.2	+	3.23	4.77E-20	7.50E-16
detection of chemical stimulus (GO:0009593)	524	92	30.28	+	3.04	9.60E-19	5.03E-15
sensory perception of chemical stimulus (GO:0007606)	542	95	31.32	+	3.03	2.79E-19	2.19E-15
detection of stimulus involved in sensory perception (GO:0050906)	558	93	32.24	+	2.88	1.21E-17	4.77E-14
antimicrobial humoral immune response mediated by antimicrobial peptide (GO:0061844)	102	17	5.89	+	2.88	2.55E-04	2.20E-02
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	73	25.6	+	2.85	7.19E-14	1.41E-10
antimicrobial humoral response (GO:0019730)	141	23	8.15	+	2.82	3.05E-05	3.83E-03
sensory perception of smell (GO:0007608)	468	75	27.04	+	2.77	1.52E-13	2.65E-10
detection of stimulus (GO:0051606)	694	98	40.1	+	2.44	2.61E-14	6.84E-11
G protein-coupled receptor signaling pathway (GO:0007186)	1216	152	70.26	+	2.16	2.69E-17	8.47E-14
skin development (GO:0043588)	232	29	13.41	+	2.16	2.84E-04	2.36E-02
sensory perception (GO:0007600)	978	108	56.51	+	1.91	1.13E-09	7.70E-07
nervous system process (GO:0050877)	1418	119	81.93	+	1.45	1.15E-04	1.14E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
Unclassified (UNCLASSIFIED)	2897	224	167.39	+	1.34	1.12E-05	1.76E-03
biological_process (GO:0008150)	17698	966	1022.61	-	0.94	1.12E-05	1.74E-03
cellular process (GO:0009987)	15074	786	870.99	-	0.9	1.37E-07	3.98E-05
biological regulation (GO:0065007)	12249	620	707.76	-	0.88	6.37E-07	1.45E-04
regulation of biological process (GO:0050789)	11528	572	666.1	-	0.86	1.05E-07	3.19E-05
regulation of cellular process (GO:0050794)	11076	542	639.98	-	0.85	3.63E-08	1.30E-05
macromolecule metabolic process (GO:0043170)	5743	277	331.84	-	0.83	5.20E-04	3.95E-02
primary metabolic process (GO:0044238)	6869	330	396.9	-	0.83	5.60E-05	6.19E-03
organic substance metabolic process (GO:0071704)	7404	354	427.81	-	0.83	1.28E-05	1.86E-03
metabolic process (GO:0008152)	7882	374	455.43	-	0.82	1.85E-06	3.68E-04
cellular metabolic process (GO:0044237)	7029	333	406.14	-	0.82	1.15E-05	1.74E-03
nitrogen compound metabolic process (GO:0006807)	6360	301	367.49	-	0.82	3.90E-05	4.79E-03
protein metabolic process (GO:0019538)	3807	171	219.97	-	0.78	2.82E-04	2.36E-02
organonitrogen compound metabolic process (GO:1901564)	4787	212	276.6	-	0.77	1.07E-05	1.71E-03
developmental process (GO:0032502)	5565	246	321.55	-	0.77	9.75E-07	2.07E-04
establishment of localization (GO:0051234)	3911	169	225.98	-	0.75	2.55E-05	3.32E-03
cellular developmental process (GO:0048869)	3463	148	200.1	-	0.74	5.03E-05	5.73E-03
localization (GO:0051179)	5081	217	293.59	-	0.74	2.56E-07	6.95E-05
cell differentiation (GO:0030154)	3403	145	196.63	-	0.74	5.27E-05	5.92E-03
regulation of macromolecule metabolic process (GO:0060255)	6083	257	351.48	-	0.73	2.12E-09	1.04E-06
regulation of catalytic activity (GO:0050790)	2393	100	138.27	-	0.72	5.19E-04	3.96E-02
regulation of metabolic process (GO:0019222)	6586	275	380.55	-	0.72	6.37E-11	5.27E-08
negative regulation of metabolic process (GO:0009892)	2952	123	170.57	-	0.72	7.93E-05	8.36E-03
animal organ development (GO:0048513)	3079	128	177.91	-	0.72	4.36E-05	5.11E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of molecular function (GO:0065009)	3060	127	176.81	-	0.72	4.18E-05	5.05E-03
transport (GO:0006810)	3752	155	216.79	-	0.71	2.98E-06	5.85E-04
regulation of biological quality (GO:0065008)	3705	153	214.08	-	0.71	3.25E-06	6.30E-04
regulation of cellular protein metabolic process (GO:0032268)	2328	96	134.51	-	0.71	3.72E-04	3.00E-02
cellular response to chemical stimulus (GO:0070887)	2567	105	148.32	-	0.71	1.31E-04	1.27E-02
organelle organization (GO:0006996)	3360	136	194.14	-	0.7	4.22E-06	7.72E-04
regulation of gene expression (GO:0010468)	4749	192	274.4	-	0.7	1.12E-08	4.19E-06
anatomical structure development (GO:0048856)	5001	202	288.96	-	0.7	3.12E-09	1.44E-06
regulation of cellular metabolic process (GO:0031323)	5928	239	342.53	-	0.7	2.55E-11	2.35E-08
positive regulation of biological process (GO:0048518)	6062	244	350.27	-	0.7	1.10E-11	1.33E-08
negative regulation of biological process (GO:0048519)	5187	207	299.71	-	0.69	4.48E-10	3.35E-07
catabolic process (GO:0009056)	1882	75	108.74	-	0.69	6.82E-04	4.94E-02
regulation of primary metabolic process (GO:0080090)	5707	227	329.76	-	0.69	1.86E-11	1.83E-08
regulation of protein metabolic process (GO:0051246)	2492	99	143.99	-	0.69	4.92E-05	5.64E-03
positive regulation of macromolecule metabolic process (GO:0010604)	3382	134	195.42	-	0.69	1.09E-06	2.29E-04
cellular component biogenesis (GO:0044085)	2481	98	143.35	-	0.68	4.68E-05	5.41E-03
regulation of nitrogen compound metabolic process (GO:0051171)	5537	218	319.93	-	0.68	1.84E-11	1.93E-08
multicellular organism development (GO:0007275)	4525	178	261.46	-	0.68	3.17E-09	1.42E-06
negative regulation of cellular metabolic process (GO:0031324)	2546	100	147.11	-	0.68	2.58E-05	3.32E-03
regulation of multicellular organismal process (GO:0051239)	2578	101	148.96	-	0.68	1.91E-05	2.61E-03
regulation of response to stimulus (GO:0048583)	3805	149	219.86	-	0.68	8.64E-08	2.77E-05

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
negative regulation of macromolecule metabolic process (GO:0010605)	2737	107	158.15	-	0.68	9.20E-06	1.52E-03
cellular component assembly (GO:0022607)	2251	88	130.07	-	0.68	7.56E-05	8.03E-03
response to organic substance (GO:0010033)	2688	105	155.32	-	0.68	1.18E-05	1.77E-03
cellular component organization or biogenesis (GO:0071840)	5496	213	317.56	-	0.67	3.96E-12	6.22E-09
cellular component organization (GO:0016043)	5298	205	306.12	-	0.67	1.17E-11	1.32E-08
positive regulation of metabolic process (GO:0009893)	3675	142	212.35	-	0.67	5.78E-08	1.93E-05
positive regulation of response to stimulus (GO:0048584)	2080	80	120.18	-	0.67	9.56E-05	9.69E-03
system development (GO:0048731)	4169	160	240.89	-	0.66	3.18E-09	1.39E-06
phosphorus metabolic process (GO:0006793)	1779	68	102.79	-	0.66	2.65E-04	2.24E-02
positive regulation of nitrogen compound metabolic process (GO:0051173)	2992	114	172.88	-	0.66	8.73E-07	1.88E-04
negative regulation of cellular process (GO:0048523)	4785	182	276.48	-	0.66	4.23E-11	3.69E-08
regulation of cellular biosynthetic process (GO:0031326)	4088	155	236.21	-	0.66	1.91E-09	9.67E-07
regulation of biosynthetic process (GO:0009889)	4152	157	239.91	-	0.65	1.17E-09	7.35E-07
regulation of macromolecule biosynthetic process (GO:0010556)	3935	148	227.37	-	0.65	2.26E-09	1.08E-06
phosphate-containing compound metabolic process (GO:0006796)	1758	66	101.58	-	0.65	1.60E-04	1.54E-02
positive regulation of cellular process (GO:0048522)	5580	209	322.42	-	0.65	6.82E-14	1.53E-10
regulation of nucleobase-containing compound metabolic process (GO:0019219)	3931	147	227.14	-	0.65	1.73E-09	9.04E-07
cellular response to organic substance (GO:0071310)	1980	74	114.41	-	0.65	5.30E-05	5.91E-03
negative regulation of response to stimulus (GO:0048585)	1559	58	90.08	-	0.64	3.30E-04	2.69E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of cellular macromolecule biosynthetic process (GO:2000112)	3906	145	225.69	-	0.64	9.91E-10	7.08E-07
positive regulation of cellular metabolic process (GO:0031325)	3186	118	184.09	-	0.64	5.95E-08	1.95E-05
intracellular signal transduction (GO:0035556)	1464	54	84.59	-	0.64	4.30E-04	3.33E-02
regulation of localization (GO:0032879)	2709	99	156.53	-	0.63	4.54E-07	1.10E-04
regulation of RNA biosynthetic process (GO:2001141)	3421	125	197.67	-	0.63	6.81E-09	2.67E-06
regulation of RNA metabolic process (GO:0051252)	3698	135	213.67	-	0.63	1.13E-09	7.39E-07
regulation of transcription, DNA-templated (GO:0006355)	3415	124	197.32	-	0.63	4.15E-09	1.72E-06
regulation of nucleic acid-templated transcription (GO:1903506)	3416	124	197.38	-	0.63	4.14E-09	1.76E-06
cell development (GO:0048468)	1623	58	93.78	-	0.62	8.70E-05	8.93E-03
cell surface receptor signaling pathway (GO:0007166)	2030	72	117.3	-	0.61	6.48E-06	1.14E-03
positive regulation of biosynthetic process (GO:0009891)	1985	70	114.7	-	0.61	6.64E-06	1.15E-03
regulation of transcription by RNA polymerase II (GO:0006357)	2538	88	146.65	-	0.6	9.68E-08	3.04E-05
positive regulation of macromolecule biosynthetic process (GO:0010557)	1847	64	106.72	-	0.6	7.08E-06	1.21E-03
negative regulation of nitrogen compound metabolic process (GO:0051172)	2380	82	137.52	-	0.6	1.99E-07	5.50E-05
positive regulation of cellular biosynthetic process (GO:0031328)	1949	67	112.62	-	0.59	3.34E-06	6.33E-04
regulation of developmental process (GO:0050793)	2368	81	136.83	-	0.59	1.46E-07	4.16E-05
positive regulation of nucleobase-containing compound metabolic process (GO:0045935)	1902	65	109.9	-	0.59	3.37E-06	6.31E-04

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
positive regulation of RNA biosynthetic process (GO:1902680)	1613	55	93.2	-	0.59	1.93E-05	2.61E-03
positive regulation of RNA metabolic process (GO:0051254)	1737	59	100.37	-	0.59	8.15E-06	1.36E-03
regulation of cell population proliferation (GO:0042127)	1650	56	95.34	-	0.59	1.40E-05	2.00E-03
regulation of protein modification process (GO:0031399)	1483	50	85.69	-	0.58	3.31E-05	4.13E-03
positive regulation of nucleic acid-templated transcription (GO:1903508)	1611	54	93.09	-	0.58	1.14E-05	1.75E-03
positive regulation of transcription, DNA-templated (GO:0045893)	1611	54	93.09	-	0.58	1.14E-05	1.74E-03
regulation of signal transduction (GO:0009966)	2867	96	165.66	-	0.58	1.29E-09	7.49E-07
positive regulation of molecular function (GO:0044093)	1497	50	86.5	-	0.58	2.71E-05	3.46E-03
cell-cell signaling (GO:0007267)	1030	34	59.51	-	0.57	5.14E-04	3.94E-02
movement of cell or subcellular component (GO:0006928)	1413	46	81.64	-	0.56	2.13E-05	2.84E-03
embryo development (GO:0009790)	991	32	57.26	-	0.56	4.05E-04	3.18E-02
regulation of signaling (GO:0023051)	3257	105	188.19	-	0.56	5.71E-12	7.48E-09
regulation of cellular component organization (GO:0051128)	2359	76	136.31	-	0.56	9.03E-09	3.46E-06
positive regulation of signal transduction (GO:0009967)	1460	47	84.36	-	0.56	1.25E-05	1.86E-03
regulation of cell communication (GO:0010646)	3242	104	187.33	-	0.56	5.13E-12	7.34E-09
positive regulation of transcription by RNA polymerase II (GO:0045944)	1185	38	68.47	-	0.55	9.44E-05	9.63E-03
apoptotic process (GO:0006915)	905	29	52.29	-	0.55	6.77E-04	4.93E-02
central nervous system development (GO:0007417)	1002	32	57.9	-	0.55	3.28E-04	2.68E-02
regulation of phosphate metabolic process (GO:0019220)	1379	44	79.68	-	0.55	1.69E-05	2.36E-03

GO biological process complete	**Homo sapiens - REFLIST (20595)**	**upload_1 (1190)**	**upload_1 (expected)**	**upload_1 (over/under)**	**upload_1 (fold Enrichment)**	**upload_1 (raw P-value)**	**upload_1 (FDR)**
regulation of phosphorus metabolic process (GO:0051174)	1379	44	79.68	-	0.55	1.69E-05	2.34E-03
plasma membrane bounded cell projection organization (GO:0120036)	1068	34	61.71	-	0.55	1.72E-04	1.62E-02
regulation of multicellular organismal development (GO:2000026)	1294	41	74.77	-	0.55	2.86E-05	3.62E-03
vesicle-mediated transport (GO:0016192)	1329	42	76.79	-	0.55	2.07E-05	2.79E-03
negative regulation of cellular biosynthetic process (GO:0031327)	1616	51	93.37	-	0.55	1.70E-06	3.42E-04
regulation of protein phosphorylation (GO:0001932)	1086	34	62.75	-	0.54	1.10E-04	1.10E-02
negative regulation of multicellular organismal process (GO:0051241)	991	31	57.26	-	0.54	2.30E-04	2.02E-02
positive regulation of cellular component organization (GO:0051130)	1119	35	64.66	-	0.54	7.99E-05	8.37E-03
cell projection organization (GO:0030030)	1119	35	64.66	-	0.54	7.99E-05	8.31E-03
positive regulation of signaling (GO:0023056)	1639	51	94.7	-	0.54	8.53E-07	1.86E-04
cell migration (GO:0016477)	840	26	48.54	-	0.54	5.74E-04	4.29E-02
negative regulation of biosynthetic process (GO:0009890)	1649	51	95.28	-	0.54	6.74E-07	1.51E-04
positive regulation of cell communication (GO:0010647)	1632	50	94.3	-	0.53	5.89E-07	1.36E-04
regulation of transport (GO:0051049)	1732	53	100.08	-	0.53	2.78E-07	7.41E-05
anatomical structure morphogenesis (GO:0009653)	2130	65	123.07	-	0.53	5.76E-09	2.32E-06
negative regulation of cell communication (GO:0010648)	1301	39	75.17	-	0.52	5.42E-06	9.80E-04
regulation of phosphorylation (GO:0042325)	1236	37	71.42	-	0.52	1.04E-05	1.71E-03
negative regulation of nucleobase-containing compound metabolic process (GO:0045934)	1538	46	88.87	-	0.52	6.87E-07	1.52E-04

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
negative regulation of signal transduction (GO:0009968)	1204	36	69.57	-	0.52	1.46E-05	2.06E-03
circulatory system development (GO:0072359)	870	26	50.27	-	0.52	2.91E-04	2.41E-02
negative regulation of signaling (GO:0023057)	1305	39	75.4	-	0.52	5.44E-06	9.72E-04
negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	1547	46	89.39	-	0.51	5.43E-07	1.29E-04
negative regulation of developmental process (GO:0051093)	877	26	50.67	-	0.51	2.29E-04	2.02E-02
regulation of intracellular signal transduction (GO:1902531)	1655	49	95.63	-	0.51	1.55E-07	4.34E-05
negative regulation of macromolecule biosynthetic process (GO:0010558)	1558	46	90.02	-	0.51	3.31E-07	8.68E-05
animal organ morphogenesis (GO:0009887)	957	28	55.3	-	0.51	7.44E-05	7.95E-03
negative regulation of transcription by RNA polymerase II (GO:0000122)	932	27	53.85	-	0.5	8.07E-05	8.35E-03
negative regulation of cell population proliferation (GO:0008285)	704	20	40.68	-	0.49	6.03E-04	4.45E-02
nervous system development (GO:0007399)	2193	62	126.71	-	0.49	1.09E-10	8.57E-08
positive regulation of intracellular signal transduction (GO:1902533)	960	27	55.47	-	0.49	3.93E-05	4.79E-03
regulation of cell differentiation (GO:0045595)	1496	42	86.44	-	0.49	1.36E-07	4.05E-05
chromatin organization (GO:0006325)	686	19	39.64	-	0.48	5.09E-04	3.92E-02
regulation of anatomical structure morphogenesis (GO:0022603)	908	25	52.47	-	0.48	4.57E-05	5.32E-03
phosphorylation (GO:0016310)	874	24	50.5	-	0.48	6.31E-05	6.93E-03
negative regulation of transcription, DNA-templated (GO:0045892)	1322	36	76.39	-	0.47	3.61E-07	9.30E-05
negative regulation of RNA metabolic process (GO:0051253)	1433	39	82.8	-	0.47	1.04E-07	3.20E-05

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
negative regulation of nucleic acid-templated transcription (GO:1903507)	1324	36	76.5	-	0.47	3.64E-07	9.22E-05
negative regulation of RNA biosynthetic process (GO:1902679)	1326	36	76.62	-	0.47	3.68E-07	9.18E-05
peptidyl-amino acid modification (GO:0018193)	851	23	49.17	-	0.47	5.17E-05	5.85E-03
generation of neurons (GO:0048699)	1237	33	71.48	-	0.46	5.81E-07	1.36E-04
cell morphogenesis (GO:0000902)	675	17	39	-	0.44	1.63E-04	1.55E-02
regulation of cellular response to stress (GO:0080135)	637	16	36.81	-	0.43	2.18E-04	1.94E-02
neurogenesis (GO:0022008)	1361	34	78.64	-	0.43	2.31E-08	8.45E-06
tube morphogenesis (GO:0035239)	644	16	37.21	-	0.43	1.67E-04	1.58E-02
embryonic morphogenesis (GO:0048598)	568	14	32.82	-	0.43	4.06E-04	3.17E-02
regulation of transmembrane transport (GO:0034762)	569	14	32.88	-	0.43	4.07E-04	3.17E-02
neuron projection development (GO:0031175)	651	16	37.62	-	0.43	1.28E-04	1.26E-02
regulation of MAPK cascade (GO:0043408)	658	16	38.02	-	0.42	9.77E-05	9.84E-03
membrane organization (GO:0061024)	783	19	45.24	-	0.42	2.32E-05	3.07E-03
positive regulation of cell death (GO:0010942)	587	14	33.92	-	0.41	2.41E-04	2.10E-02
endocytosis (GO:0006897)	509	12	29.41	-	0.41	5.77E-04	4.29E-02
regulation of ion transport (GO:0043269)	685	16	39.58	-	0.4	4.35E-05	5.14E-03
positive regulation of transport (GO:0051050)	909	21	52.52	-	0.4	1.37E-06	2.79E-04
cellular component morphogenesis (GO:0032989)	569	13	32.88	-	0.4	1.90E-04	1.73E-02
positive regulation of cell differentiation (GO:0045597)	833	19	48.13	-	0.39	3.31E-06	6.34E-04
neuron development (GO:0048666)	810	18	46.8	-	0.38	3.43E-06	6.34E-04
negative regulation of protein modification process (GO:0031400)	500	11	28.89	-	0.38	3.43E-04	2.78E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
vasculature development (GO:0001944)	509	11	29.41	-	0.37	2.62E-04	2.22E-02
negative regulation of transport (GO:0051051)	430	9	24.85	-	0.36	6.08E-04	4.46E-02
neuron differentiation (GO:0030182)	1008	21	58.24	-	0.36	4.53E-08	1.55E-05
regulation of ion transmembrane transport (GO:0034765)	481	10	27.79	-	0.36	2.59E-04	2.21E-02
embryonic organ development (GO:0048568)	433	9	25.02	-	0.36	6.24E-04	4.56E-02
cell part morphogenesis (GO:0032990)	483	10	27.91	-	0.36	1.89E-04	1.74E-02
blood vessel development (GO:0001568)	486	10	28.08	-	0.36	1.90E-04	1.73E-02
regulation of cell projection organization (GO:0031344)	643	13	37.15	-	0.35	1.07E-05	1.73E-03
neuron projection morphogenesis (GO:0048812)	455	9	26.29	-	0.34	2.57E-04	2.21E-02
plasma membrane bounded cell projection morphogenesis (GO:0120039)	459	9	26.52	-	0.34	1.87E-04	1.72E-02
cell projection morphogenesis (GO:0048858)	463	9	26.75	-	0.34	1.92E-04	1.74E-02
regulation of plasma membrane bounded cell projection organization (GO:0120035)	626	12	36.17	-	0.33	7.80E-06	1.32E-03
positive regulation of MAPK cascade (GO:0043410)	471	9	27.21	-	0.33	1.43E-04	1.38E-02
regulation of DNA-binding transcription factor activity (GO:0051090)	424	8	24.5	-	0.33	2.43E-04	2.11E-02
sensory system development (GO:0048880)	376	7	21.73	-	0.32	5.82E-04	4.31E-02
regulation of membrane potential (GO:0042391)	434	8	25.08	-	0.32	1.82E-04	1.70E-02
sensory organ development (GO:0007423)	558	10	32.24	-	0.31	1.32E-05	1.90E-03
cell morphogenesis involved in differentiation (GO:0000904)	517	9	29.87	-	0.3	2.32E-05	3.04E-03
regulation of hemopoiesis (GO:1903706)	353	6	20.4	-	0.29	3.99E-04	3.17E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
eye development (GO:0001654)	366	6	21.15	-	0.28	3.00E-04	2.46E-02
modulation of chemical synaptic transmission (GO:0050804)	431	7	24.9	-	0.28	6.78E-05	7.35E-03
visual system development (GO:0150063)	370	6	21.38	-	0.28	2.13E-04	1.90E-02
regulation of trans-synaptic signaling (GO:0099177)	432	7	24.96	-	0.28	6.87E-05	7.39E-03
cellular response to external stimulus (GO:0071496)	313	5	18.09	-	0.28	6.91E-04	4.98E-02
regulation of neuron projection development (GO:0010975)	443	7	25.6	-	0.27	3.51E-05	4.35E-03
regulation of Wnt signaling pathway (GO:0030111)	325	5	18.78	-	0.27	5.22E-04	3.95E-02
cell surface receptor signaling pathway involved in cell-cell signaling (GO:1905114)	331	5	19.13	-	0.26	3.72E-04	2.98E-02
axonogenesis (GO:0007409)	337	5	19.47	-	0.26	2.66E-04	2.23E-02
cell morphogenesis involved in neuron differentiation (GO:0048667)	408	6	23.57	-	0.25	4.29E-05	5.11E-03
regulation of metal ion transport (GO:0010959)	394	5	22.77	-	0.22	1.87E-05	2.58E-03
regulation of cation transmembrane transport (GO:1904062)	347	4	20.05	-	0.2	4.23E-05	5.08E-03
positive regulation of ion transport (GO:0043270)	278	3	16.06	-	0.19	1.82E-04	1.69E-02
regulation of transporter activity (GO:0032409)	287	3	16.58	-	0.18	1.29E-04	1.26E-02
gliogenesis (GO:0042063)	224	2	12.94	-	0.15	5.66E-04	4.26E-02
chromatin remodeling (GO:0006338)	247	2	14.27	-	0.14	1.96E-04	1.76E-02
regulation of transmembrane transporter activity (GO:0022898)	270	2	15.6	-	0.13	6.77E-05	7.39E-03
regulation of neuron differentiation (GO:0045664)	193	1	11.15	-	0.09	3.74E-04	2.98E-02
regulation of ion transmembrane transporter activity (GO:0032412)	259	1	14.97	-	0.07	1.26E-05	1.85E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1190)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
retina development in camera-type eye (GO:0060041)	147	0	8.49	-	< 0.01	4.02E-04	3.17E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

Significant PDCUB

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (3652)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
cell differentiation in spinal cord (GO:0021515)	49	24	8.69	+	2.76	1.15E-04	2.70E-02
neuropeptide signaling pathway (GO:0007218)	109	52	19.33	+	2.69	4.51E-08	2.36E-04
nucleosome assembly (GO:0006334)	101	48	17.91	+	2.68	1.19E-07	3.73E-04
G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger (GO:0007187)	67	29	11.88	+	2.44	1.48E-04	3.29E-02
regulation of amine transport (GO:0051952)	92	38	16.31	+	2.33	3.53E-05	1.39E-02
chromatin assembly (GO:0031497)	134	52	23.76	+	2.19	6.75E-06	5.31E-03
nucleosome organization (GO:0034728)	140	53	24.83	+	2.13	1.01E-05	7.25E-03
adenylate cyclase-activating G protein-coupled receptor signaling pathway (GO:0007189)	136	50	24.12	+	2.07	3.00E-05	1.35E-02
protein-DNA complex assembly (GO:0065004)	167	61	29.61	+	2.06	4.12E-06	4.31E-03
response to glucocorticoid (GO:0051384)	149	54	26.42	+	2.04	1.83E-05	9.57E-03
response to corticosteroid (GO:0031960)	168	60	29.79	+	2.01	9.66E-06	7.23E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (3652)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
potassium ion transmembrane transport (GO:0071805)	158	56	28.02	+	2	2.91E-05	1.35E-02
potassium ion transport (GO:0006813)	167	58	29.61	+	1.96	3.05E-05	1.33E-02
chromatin assembly or disassembly (GO:0006333)	154	53	27.31	+	1.94	7.87E-05	2.13E-02
adenylate cyclase-modulating G protein-coupled receptor signaling pathway (GO:0007188)	223	76	39.54	+	1.92	2.79E-06	3.37E-03
DNA packaging (GO:0006323)	180	59	31.92	+	1.85	8.36E-05	2.15E-02
protein-DNA complex subunit organization (GO:0071824)	208	66	36.88	+	1.79	8.18E-05	2.18E-02
cell fate commitment (GO:0045165)	238	75	42.2	+	1.78	3.10E-05	1.28E-02
embryonic organ morphogenesis (GO:0048562)	292	85	51.78	+	1.64	1.09E-04	2.65E-02
regionalization (GO:0003002)	316	91	56.03	+	1.62	8.34E-05	2.18E-02
embryonic organ development (GO:0048568)	433	121	76.78	+	1.58	1.82E-05	9.84E-03
blood circulation (GO:0008015)	399	110	70.75	+	1.55	7.38E-05	2.04E-02
pattern specification process (GO:0007389)	425	114	75.36	+	1.51	1.50E-04	3.23E-02
embryonic morphogenesis (GO:0048598)	568	148	100.72	+	1.47	4.22E-05	1.51E-02
sensory organ development (GO:0007423)	558	143	98.95	+	1.45	1.10E-04	2.63E-02
regulation of ion transport (GO:0043269)	685	172	121.47	+	1.42	6.39E-05	1.93E-02
response to lipid (GO:0033993)	829	206	147	+	1.4	1.92E-05	9.71E-03
animal organ morphogenesis (GO:0009887)	957	237	169.7	+	1.4	4.77E-06	4.16E-03
negative regulation of cell differentiation (GO:0045596)	641	158	113.67	+	1.39	2.37E-04	4.65E-02
negative regulation of developmental process (GO:0051093)	877	214	155.51	+	1.38	3.11E-05	1.25E-02
G protein-coupled receptor signaling pathway (GO:0007186)	1216	296	215.63	+	1.37	9.11E-07	1.43E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (3652)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
response to organic cyclic compound (GO:0014070)	896	218	158.88	+	1.37	3.08E-05	1.31E-02
central nervous system development (GO:0007417)	1002	241	177.68	+	1.36	2.22E-05	1.09E-02
head development (GO:0060322)	806	192	142.92	+	1.34	2.47E-04	4.80E-02
embryo development (GO:0009790)	991	236	175.73	+	1.34	4.86E-05	1.63E-02
cell-cell signaling (GO:0007267)	1030	245	182.64	+	1.34	3.95E-05	1.51E-02
negative regulation of multicellular organismal process (GO:0051241)	991	235	175.73	+	1.34	6.83E-05	1.92E-02
response to endogenous stimulus (GO:0009719)	1342	317	237.97	+	1.33	3.64E-06	4.09E-03
response to oxygen-containing compound (GO:1901700)	1567	370	277.87	+	1.33	4.64E-07	8.11E-04
positive regulation of multicellular organismal process (GO:0051240)	1412	315	250.38	+	1.26	1.85E-04	3.77E-02
tissue development (GO:0009888)	1597	355	283.19	+	1.25	8.47E-05	2.15E-02
animal organ development (GO:0048513)	3079	675	545.98	+	1.24	9.29E-08	3.65E-04
regulation of multicellular organismal process (GO:0051239)	2578	560	457.14	+	1.23	4.70E-06	4.35E-03
system process (GO:0003008)	2040	443	361.74	+	1.22	6.31E-05	1.94E-02
response to organic substance (GO:0010033)	2688	582	476.65	+	1.22	4.15E-06	4.08E-03
anatomical structure morphogenesis (GO:0009653)	2130	456	377.7	+	1.21	1.54E-04	3.28E-02
nervous system development (GO:0007399)	2193	466	388.87	+	1.2	2.09E-04	4.16E-02
regulation of developmental process (GO:0050793)	2368	501	419.9	+	1.19	1.73E-04	3.63E-02
system development (GO:0048731)	4169	876	739.27	+	1.18	4.03E-07	9.05E-04
multicellular organism development (GO:0007275)	4525	943	802.39	+	1.18	4.38E-07	8.61E-04
anatomical structure development (GO:0048856)	5001	1037	886.8	+	1.17	1.66E-07	4.34E-04
regulation of biological quality (GO:0065008)	3705	758	656.99	+	1.15	8.97E-05	2.24E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (3652)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
multicellular organismal process (GO:0032501)	6584	1342	1167.51	+	1.15	1.81E-08	1.42E-04
developmental process (GO:0032502)	5565	1111	986.81	+	1.13	2.65E-05	1.26E-02
negative regulation of biological process (GO:0048519)	5187	1031	919.78	+	1.12	1.20E-04	2.78E-02
cellular process (GO:0009987)	15074	2782	2672.99	+	1.04	1.50E-04	3.27E-02
biological_process (GO:0008150)	17698	3229	3138.29	+	1.03	4.57E-05	1.60E-02
Unclassified (UNCLASSIFIED)	2897	423	513.71	-	0.82	4.57E-05	1.56E-02
immune system process (GO:0002376)	2370	339	420.26	-	0.81	6.54E-05	1.90E-02
immune response (GO:0006955)	1678	223	297.55	-	0.75	1.60E-05	9.68E-03
cell cycle (GO:0007049)	1205	157	213.68	-	0.73	1.30E-04	2.96E-02
microtubule-based process (GO:0007017)	765	91	135.65	-	0.67	1.43E-04	3.22E-02
DNA metabolic process (GO:0006259)	714	84	126.61	-	0.66	1.89E-04	3.81E-02
positive regulation of immune response (GO:0050778)	563	59	99.83	-	0.59	4.08E-05	1.53E-02
immune response-regulating signaling pathway (GO:0002764)	363	33	64.37	-	0.51	6.76E-05	1.93E-02
immune response-regulating cell surface receptor signaling pathway (GO:0002768)	290	25	51.42	-	0.49	1.78E-04	3.68E-02
immune response-activating cell surface receptor signaling pathway (GO:0002429)	263	20	46.64	-	0.43	5.31E-05	1.70E-02
immune response-activating signal transduction (GO:0002757)	263	20	46.64	-	0.43	5.31E-05	1.67E-02
activation of immune response (GO:0002253)	338	25	59.94	-	0.42	2.61E-06	3.42E-03
adaptive immune response (GO:0002250)	640	45	113.49	-	0.4	5.54E-12	8.71E-08
cell recognition (GO:0008037)	228	16	40.43	-	0.4	6.49E-05	1.93E-02
regulation of chromosome organization (GO:0033044)	286	20	50.71	-	0.39	5.58E-06	4.62E-03
positive regulation of DNA metabolic process (GO:0051054)	203	13	36	-	0.36	5.16E-05	1.69E-02

GO biological process complete	**Homo sapiens - REFLIST (20595)**	**upload_1 (3652)**	**upload_1 (expected)**	**upload_1 (over/under)**	**upload_1 (fold Enrichment)**	**upload_1 (raw P-value)**	**upload_1 (FDR)**
antigen receptor-mediated signaling pathway (GO:0050851)	216	13	38.3	-	0.34	1.55E-05	9.74E-03
positive regulation of chromosome organization (GO:2001252)	175	9	31.03	-	0.29	1.81E-05	1.02E-02
complement activation (GO:0006956)	139	5	24.65	-	0.2	1.08E-05	7.38E-03
complement activation, classical pathway (GO:0006958)	124	4	21.99	-	0.18	1.78E-05	1.03E-02
humoral immune response mediated by circulating immunoglobulin (GO:0002455)	128	4	22.7	-	0.18	1.26E-05	8.28E-03
phagocytosis, recognition (GO:0006910)	107	3	18.97	-	0.16	4.17E-05	1.52E-02
B cell receptor signaling pathway (GO:0050853)	130	3	23.05	-	0.13	1.30E-06	1.86E-03
centrosome cycle (GO:0007098)	88	2	15.6	-	0.13	9.17E-05	2.25E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

Non-significant PDCUB

GO biological process complete	**Homo sapiens - REFLIST (20595)**	**upload_1 (over/under)**	**upload_1 (fold Enrichment)**	**upload_1 (raw P-value)**	**upload_1 (FDR)**
detection of chemical stimulus involved in sensory perception of bitter taste (GO:0001580)	36	+	5.58	6.02E-09	1.69E-06

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
sensory perception of bitter taste (GO:0050913)	41	+	5.12	1.01E-08	2.64E-06
detection of chemical stimulus involved in sensory perception of taste (GO:0050912)	42	+	5	1.43E-08	3.51E-06
negative regulation of execution phase of apoptosis (GO:1900118)	23	+	4.94	3.23E-05	2.85E-03
detection of chemical stimulus involved in sensory perception (GO:0050907)	488	+	3.85	8.43E-52	1.32E-47
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	+	3.77	6.31E-45	1.98E-41
regulation of execution phase of apoptosis (GO:1900117)	31	+	3.67	3.32E-04	2.04E-02
detection of chemical stimulus (GO:0009593)	524	+	3.65	8.89E-50	6.98E-46
sensory perception of smell (GO:0007608)	468	+	3.62	1.15E-43	3.02E-40
sensory perception of chemical stimulus (GO:0007606)	542	+	3.56	7.85E-49	4.11E-45
sensory perception of taste (GO:0050909)	68	+	3.47	6.31E-07	8.85E-05
detection of stimulus involved in sensory perception (GO:0050906)	558	+	3.43	3.02E-46	1.19E-42
snRNA 3'-end processing (GO:0034472)	31	+	3.38	9.81E-04	4.74E-02
keratinization (GO:0031424)	58	+	3.01	8.74E-05	6.63E-03
detection of stimulus (GO:0051606)	694	+	2.95	1.62E-40	3.63E-37
ncRNA 3'-end processing (GO:0043628)	50	+	2.8	8.41E-04	4.16E-02
killing of cells of other organism (GO:0031640)	86	+	2.54	2.04E-04	1.34E-02
antimicrobial humoral immune response mediated by antimicrobial peptide (GO:0061844)	102	+	2.4	1.26E-04	9.09E-03
antimicrobial humoral response (GO:0019730)	141	+	2.36	1.68E-05	1.58E-03
G protein-coupled receptor signaling pathway (GO:0007186)	1216	+	2.28	4.24E-36	8.34E-33
sensory perception (GO:0007600)	978	+	2.23	3.83E-27	6.69E-24
cell killing (GO:0001906)	125	+	2.1	7.29E-04	3.79E-02
nervous system process (GO:0050877)	1418	+	1.71	1.23E-15	1.20E-12
humoral immune response (GO:0006959)	337	+	1.66	3.38E-04	2.06E-02
system process (GO:0003008)	2040	+	1.39	1.47E-08	3.50E-06
Unclassified (UNCLASSIFIED)	2897	+	1.25	6.07E-06	6.49E-04
biological_process (GO:0008150)	17698	-	0.96	6.07E-06	6.54E-04
biological regulation (GO:0065007)	12249	-	0.93	2.02E-04	1.33E-02
regulation of biological process (GO:0050789)	11528	-	0.93	5.01E-04	2.87E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of cellular process (GO:0050794)	11076	-	0.92	7.94E-05	6.18E-03
cellular process (GO:0009987)	15074	-	0.92	1.22E-09	4.09E-07
cellular macromolecule metabolic process (GO:0044260)	4442	-	0.84	7.32E-05	5.81E-03
cellular metabolic process (GO:0044237)	7029	-	0.83	2.77E-08	5.66E-06
primary metabolic process (GO:0044238)	6869	-	0.83	1.53E-08	3.53E-06
nitrogen compound metabolic process (GO:0006807)	6360	-	0.83	6.76E-08	1.25E-05
macromolecule modification (GO:0043412)	2825	-	0.82	7.30E-04	3.79E-02
metabolic process (GO:0008152)	7882	-	0.82	3.84E-11	2.08E-08
organic substance metabolic process (GO:0071704)	7404	-	0.82	1.53E-10	6.70E-08
macromolecule metabolic process (GO:0043170)	5743	-	0.82	7.82E-08	1.43E-05
negative regulation of macromolecule metabolic process (GO:0010605)	2737	-	0.81	6.89E-04	3.60E-02
regulation of macromolecule metabolic process (GO:0060255)	6083	-	0.81	8.46E-09	2.29E-06
regulation of metabolic process (GO:0019222)	6586	-	0.81	1.05E-09	3.58E-07
cellular protein modification process (GO:0006464)	2610	-	0.81	5.33E-04	3.01E-02
protein modification process (GO:0036211)	2610	-	0.81	5.33E-04	3.00E-02
cellular protein metabolic process (GO:0044267)	3219	-	0.81	8.59E-05	6.55E-03
regulation of gene expression (GO:0010468)	4749	-	0.8	3.85E-07	5.93E-05
regulation of biosynthetic process (GO:0009889)	4152	-	0.8	2.93E-06	3.52E-04
regulation of cellular biosynthetic process (GO:0031326)	4088	-	0.8	3.36E-06	3.89E-04
developmental process (GO:0032502)	5565	-	0.8	1.38E-08	3.44E-06
regulation of RNA biosynthetic process (GO:2001141)	3421	-	0.79	1.26E-05	1.23E-03
regulation of transcription, DNA-templated (GO:0006355)	3415	-	0.79	1.09E-05	1.09E-03
regulation of macromolecule biosynthetic process (GO:0010556)	3935	-	0.79	1.57E-06	2.01E-04
regulation of nucleic acid-templated transcription (GO:1903506)	3416	-	0.79	1.09E-05	1.08E-03
regulation of cellular metabolic process (GO:0031323)	5928	-	0.79	3.62E-10	1.42E-07
regulation of cellular macromolecule biosynthetic process (GO:2000112)	3906	-	0.79	1.68E-06	2.13E-04
organonitrogen compound metabolic process (GO:1901564)	4787	-	0.79	4.24E-08	8.13E-06

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
negative regulation of nitrogen compound metabolic process (GO:0051172)	2380	-	0.79	2.68E-04	1.69E-02
regulation of primary metabolic process (GO:0080090)	5707	-	0.78	2.22E-10	9.19E-08
protein metabolic process (GO:0019538)	3807	-	0.78	7.41E-07	1.01E-04
regulation of nitrogen compound metabolic process (GO:0051171)	5537	-	0.78	1.24E-10	5.58E-08
negative regulation of biological process (GO:0048519)	5187	-	0.77	5.81E-10	2.12E-07
regulation of RNA metabolic process (GO:0051252)	3698	-	0.77	4.70E-07	6.97E-05
regulation of nucleobase-containing compound metabolic process (GO:0019219)	3931	-	0.77	1.47E-07	2.46E-05
macromolecule localization (GO:0033036)	2308	-	0.77	9.43E-05	7.09E-03
protein localization (GO:0008104)	1901	-	0.77	4.69E-04	2.71E-02
negative regulation of cellular process (GO:0048523)	4785	-	0.76	3.64E-10	1.40E-07
negative regulation of biosynthetic process (GO:0009890)	1649	-	0.76	7.55E-04	3.88E-02
negative regulation of cellular biosynthetic process (GO:0031327)	1616	-	0.76	7.84E-04	3.95E-02
anatomical structure development (GO:0048856)	5001	-	0.76	4.56E-11	2.39E-08
cellular response to chemical stimulus (GO:0070887)	2567	-	0.75	6.76E-06	7.13E-04
negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	1547	-	0.75	6.05E-04	3.32E-02
localization (GO:0051179)	5081	-	0.74	4.05E-12	2.45E-09
animal organ development (GO:0048513)	3079	-	0.74	2.96E-07	4.66E-05
regulation of cellular protein metabolic process (GO:0032268)	2328	-	0.74	1.36E-05	1.32E-03
cellular localization (GO:0051641)	2264	-	0.74	1.47E-05	1.41E-03
negative regulation of macromolecule biosynthetic process (GO:0010558)	1558	-	0.74	3.96E-04	2.36E-02
cellular component biogenesis (GO:0044085)	2481	-	0.74	4.17E-06	4.68E-04
multicellular organism development (GO:0007275)	4525	-	0.74	3.67E-11	2.06E-08
establishment of localization (GO:0051234)	3911	-	0.74	1.34E-09	4.40E-07
phosphorus metabolic process (GO:0006793)	1779	-	0.74	1.20E-04	8.71E-03
phosphate-containing compound metabolic process (GO:0006796)	1758	-	0.74	1.29E-04	9.25E-03
transport (GO:0006810)	3752	-	0.73	1.02E-09	3.55E-07
cellular response to organic substance (GO:0071310)	1980	-	0.73	2.35E-05	2.18E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
negative regulation of transcription, DNA-templated (GO:0045892)	1322	-	0.73	6.27E-04	3.43E-02
negative regulation of nucleic acid-templated transcription (GO:1903507)	1324	-	0.73	6.27E-04	3.42E-02
cellular macromolecule localization (GO:0070727)	1386	-	0.73	4.26E-04	2.50E-02
negative regulation of RNA biosynthetic process (GO:1902679)	1326	-	0.73	6.30E-04	3.42E-02
cellular protein localization (GO:0034613)	1376	-	0.72	4.09E-04	2.41E-02
cellular developmental process (GO:0048869)	3463	-	0.72	2.19E-09	6.76E-07
negative regulation of nucleobase-containing compound metabolic process (GO:0045934)	1538	-	0.72	1.58E-04	1.09E-02
system development (GO:0048731)	4169	-	0.72	1.86E-11	1.08E-08
cellular response to stress (GO:0033554)	1519	-	0.72	1.45E-04	1.02E-02
cell differentiation (GO:0030154)	3403	-	0.72	2.04E-09	6.41E-07
response to organic substance (GO:0010033)	2688	-	0.72	1.63E-07	2.69E-05
regulation of protein metabolic process (GO:0051246)	2492	-	0.72	3.90E-07	5.95E-05
protein-containing complex assembly (GO:0065003)	1185	-	0.72	7.63E-04	3.88E-02
organic substance catabolic process (GO:1901575)	1587	-	0.71	5.91E-05	4.83E-03
positive regulation of response to stimulus (GO:0048584)	2080	-	0.71	3.25E-06	3.81E-04
regulation of biological quality (GO:0065008)	3705	-	0.71	7.20E-11	3.54E-08
regulation of cell death (GO:0010941)	1593	-	0.71	5.06E-05	4.25E-03
protein-containing complex subunit organization (GO:0043933)	1337	-	0.71	1.98E-04	1.32E-02
regulation of multicellular organismal process (GO:0051239)	2578	-	0.71	7.86E-08	1.42E-05
positive regulation of gene expression (GO:0010628)	1128	-	0.71	6.75E-04	3.54E-02
regulation of apoptotic process (GO:0042981)	1426	-	0.7	1.08E-04	7.92E-03
cellular component assembly (GO:0022607)	2251	-	0.7	5.97E-07	8.54E-05
cellular response to endogenous stimulus (GO:0071495)	1070	-	0.7	8.49E-04	4.18E-02
negative regulation of RNA metabolic process (GO:0051253)	1433	-	0.7	9.23E-05	6.97E-03
positive regulation of biological process (GO:0048518)	6062	-	0.7	8.84E-21	1.26E-17
regulation of programmed cell death (GO:0043067)	1456	-	0.7	6.02E-05	4.90E-03
cellular component organization or biogenesis (GO:0071840)	5496	-	0.7	1.45E-18	1.75E-15

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of catalytic activity (GO:0050790)	2393	-	0.69	9.05E-08	1.62E-05
cellular catabolic process (GO:0044248)	1615	-	0.69	1.57E-05	1.49E-03
organelle organization (GO:0006996)	3360	-	0.69	5.94E-11	3.01E-08
negative regulation of multicellular organismal process (GO:0051241)	991	-	0.69	7.77E-04	3.93E-02
positive regulation of multicellular organismal process (GO:0051240)	1412	-	0.69	4.66E-05	3.94E-03
response to endogenous stimulus (GO:0009719)	1342	-	0.68	5.50E-05	4.55E-03
catabolic process (GO:0009056)	1882	-	0.68	1.10E-06	1.48E-04
regulation of molecular function (GO:0065009)	3060	-	0.68	1.02E-10	4.86E-08
regulation of locomotion (GO:0040012)	990	-	0.68	5.28E-04	2.99E-02
cellular component organization (GO:0016043)	5298	-	0.68	1.03E-19	1.35E-16
proteolysis (GO:0006508)	1216	-	0.68	9.62E-05	7.17E-03
regulation of response to stimulus (GO:0048583)	3805	-	0.68	8.70E-14	6.84E-11
regulation of cell cycle (GO:0051726)	1001	-	0.67	3.78E-04	2.28E-02
organonitrogen compound catabolic process (GO:1901565)	1029	-	0.67	3.05E-04	1.89E-02
positive regulation of cellular process (GO:0048522)	5580	-	0.67	3.59E-22	5.64E-19
positive regulation of metabolic process (GO:0009893)	3675	-	0.67	7.54E-14	6.24E-11
programmed cell death (GO:0012501)	942	-	0.67	4.65E-04	2.70E-02
negative regulation of developmental process (GO:0051093)	877	-	0.67	8.12E-04	4.06E-02
positive regulation of macromolecule metabolic process (GO:0010604)	3382	-	0.67	7.91E-13	5.40E-10
positive regulation of transcription by RNA polymerase II (GO:0045944)	1185	-	0.66	6.38E-05	5.14E-03
circulatory system development (GO:0072359)	870	-	0.66	6.37E-04	3.44E-02
regulation of cell motility (GO:2000145)	950	-	0.66	3.29E-04	2.03E-02
regulation of cellular component biogenesis (GO:0044087)	950	-	0.66	3.29E-04	2.02E-02
cell death (GO:0008219)	978	-	0.66	2.64E-04	1.67E-02
regulation of transferase activity (GO:0051338)	873	-	0.66	6.43E-04	3.44E-02
response to oxygen-containing compound (GO:1901700)	1567	-	0.66	1.89E-06	2.38E-04
negative regulation of gene expression (GO:0010629)	879	-	0.66	4.48E-04	2.62E-02
regulation of cell migration (GO:0030334)	894	-	0.66	4.03E-04	2.39E-02
positive regulation of cellular metabolic process (GO:0031325)	3186	-	0.65	5.06E-13	3.79E-10

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
response to nitrogen compound (GO:1901698)	1071	-	0.65	7.98E-05	6.18E-03
positive regulation of biosynthetic process (GO:0009891)	1985	-	0.65	3.36E-08	6.77E-06
cell surface receptor signaling pathway (GO:0007166)	2030	-	0.65	1.61E-08	3.62E-06
positive regulation of cellular biosynthetic process (GO:0031328)	1949	-	0.65	2.56E-08	5.29E-06
cell-cell signaling (GO:0007267)	1030	-	0.65	7.23E-05	5.77E-03
movement of cell or subcellular component (GO:0006928)	1413	-	0.64	2.64E-06	3.24E-04
cell adhesion (GO:0007155)	952	-	0.64	1.41E-04	9.93E-03
positive regulation of nitrogen compound metabolic process (GO:0051173)	2992	-	0.64	7.51E-13	5.37E-10
regulation of protein modification process (GO:0031399)	1483	-	0.64	1.23E-06	1.64E-04
negative regulation of cellular component organization (GO:0051129)	722	-	0.64	9.11E-04	4.46E-02
positive regulation of phosphorus metabolic process (GO:0010562)	886	-	0.64	2.46E-04	1.59E-02
positive regulation of phosphate metabolic process (GO:0045937)	886	-	0.64	2.46E-04	1.58E-02
positive regulation of cellular protein metabolic process (GO:0032270)	1366	-	0.64	2.84E-06	3.44E-04
biological adhesion (GO:0022610)	958	-	0.64	9.73E-05	7.21E-03
regulation of developmental process (GO:0050793)	2368	-	0.64	2.00E-10	8.49E-08
positive regulation of nucleobase-containing compound metabolic process (GO:0045935)	1902	-	0.63	1.13E-08	2.91E-06
positive regulation of macromolecule biosynthetic process (GO:0010557)	1847	-	0.63	1.81E-08	4.01E-06
positive regulation of RNA biosynthetic process (GO:1902680)	1613	-	0.63	1.93E-07	3.13E-05
negative regulation of response to stimulus (GO:0048585)	1559	-	0.63	3.05E-07	4.75E-05
response to lipid (GO:0033993)	829	-	0.63	2.42E-04	1.58E-02
regulation of localization (GO:0032879)	2709	-	0.63	2.76E-12	1.74E-09
regulation of cellular component movement (GO:0051270)	1025	-	0.63	3.57E-05	3.10E-03
positive regulation of nucleic acid-templated transcription (GO:1903508)	1611	-	0.63	1.21E-07	2.10E-05
positive regulation of transcription, DNA-templated (GO:0045893)	1611	-	0.63	1.21E-07	2.08E-05
positive regulation of RNA metabolic process (GO:0051254)	1737	-	0.63	3.53E-08	7.03E-06
positive regulation of protein phosphorylation (GO:0001934)	723	-	0.63	5.82E-04	3.25E-02
positive regulation of signal transduction (GO:0009967)	1460	-	0.63	5.19E-07	7.56E-05

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
positive regulation of phosphorylation (GO:0042327)	807	-	0.63	2.50E-04	1.60E-02
positive regulation of developmental process (GO:0051094)	1267	-	0.63	3.05E-06	3.63E-04
apoptotic process (GO:0006915)	905	-	0.63	9.61E-05	7.19E-03
anatomical structure formation involved in morphogenesis (GO:0048646)	881	-	0.63	1.20E-04	8.72E-03
regulation of organelle organization (GO:0033043)	1189	-	0.62	6.00E-06	6.51E-04
anatomical structure morphogenesis (GO:0009653)	2130	-	0.62	3.61E-10	1.45E-07
response to hormone (GO:0009725)	772	-	0.62	2.87E-04	1.80E-02
positive regulation of protein modification process (GO:0031401)	970	-	0.62	4.26E-05	3.64E-03
regulation of multicellular organismal development (GO:2000026)	1294	-	0.62	1.53E-06	1.97E-04
cell development (GO:0048468)	1623	-	0.62	5.20E-08	9.85E-06
regulation of signal transduction (GO:0009966)	2867	-	0.62	6.23E-14	5.44E-11
positive regulation of protein metabolic process (GO:0051247)	1455	-	0.62	2.47E-07	3.96E-05
positive regulation of catalytic activity (GO:0043085)	1134	-	0.62	5.83E-06	6.41E-04
regulation of phosphate metabolic process (GO:0019220)	1379	-	0.61	4.00E-07	6.04E-05
regulation of phosphorus metabolic process (GO:0051174)	1379	-	0.61	4.00E-07	5.98E-05
protein phosphorylation (GO:0006468)	683	-	0.61	5.06E-04	2.89E-02
cytoskeleton organization (GO:0007010)	1169	-	0.61	3.16E-06	3.73E-04
regulation of protein phosphorylation (GO:0001932)	1086	-	0.61	7.49E-06	7.79E-04
regulation of MAPK cascade (GO:0043408)	658	-	0.61	6.32E-04	3.43E-02
positive regulation of intracellular signal transduction (GO:1902533)	960	-	0.61	2.43E-05	2.23E-03
regulation of phosphorylation (GO:0042325)	1236	-	0.61	1.12E-06	1.50E-04
positive regulation of signaling (GO:0023056)	1639	-	0.61	1.43E-08	3.46E-06
positive regulation of cell population proliferation (GO:0008284)	935	-	0.61	3.05E-05	2.70E-03
positive regulation of cell communication (GO:0010647)	1632	-	0.61	1.32E-08	3.36E-06
regulation of protein localization (GO:0032880)	853	-	0.6	5.78E-05	4.76E-03
regulation of cellular component organization (GO:0051128)	2359	-	0.6	2.30E-12	1.51E-09
regulation of cellular localization (GO:0060341)	820	-	0.6	5.33E-05	4.46E-03
negative regulation of cell population proliferation (GO:0008285)	704	-	0.6	1.87E-04	1.26E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of cell communication (GO:0010646)	3242	-	0.6	7.78E-18	8.15E-15
regulation of signaling (GO:0023051)	3257	-	0.6	6.78E-18	7.61E-15
positive regulation of cell death (GO:0010942)	587	-	0.6	8.12E-04	4.05E-02
response to organonitrogen compound (GO:0010243)	984	-	0.6	9.04E-06	9.17E-04
positive regulation of cellular component organization (GO:0051130)	1119	-	0.59	1.50E-06	1.96E-04
regulation of cell population proliferation (GO:0042127)	1650	-	0.59	2.82E-09	8.51E-07
intracellular signal transduction (GO:0035556)	1464	-	0.59	2.28E-08	4.85E-06
embryo development (GO:0009790)	991	-	0.59	5.96E-06	6.50E-04
regulation of transport (GO:0051049)	1732	-	0.59	8.45E-10	3.02E-07
regulation of cell differentiation (GO:0045595)	1496	-	0.59	1.49E-08	3.48E-06
head development (GO:0060322)	806	-	0.59	3.57E-05	3.12E-03
peptidyl-amino acid modification (GO:0018193)	851	-	0.59	2.12E-05	1.97E-03
response to abiotic stimulus (GO:0009628)	1125	-	0.58	7.30E-07	1.02E-04
regulation of cellular catabolic process (GO:0031329)	813	-	0.58	2.91E-05	2.63E-03
phosphorylation (GO:0016310)	874	-	0.58	1.26E-05	1.23E-03
positive regulation of hydrolase activity (GO:0051345)	588	-	0.58	3.92E-04	2.34E-02
behavior (GO:0007610)	558	-	0.58	5.96E-04	3.30E-02
tube development (GO:0035295)	845	-	0.58	1.93E-05	1.80E-03
vesicle-mediated transport (GO:0016192)	1329	-	0.58	4.08E-08	7.92E-06
positive regulation of molecular function (GO:0044093)	1497	-	0.58	5.25E-09	1.50E-06
regulation of cytoskeleton organization (GO:0051493)	531	-	0.58	7.33E-04	3.79E-02
central nervous system development (GO:0007417)	1002	-	0.58	1.89E-06	2.36E-04
regulation of intracellular signal transduction (GO:1902531)	1655	-	0.58	4.67E-10	1.75E-07
animal organ morphogenesis (GO:0009887)	957	-	0.58	4.02E-06	4.54E-04
brain development (GO:0007420)	761	-	0.57	3.57E-05	3.13E-03
negative regulation of cell communication (GO:0010648)	1301	-	0.56	9.85E-09	2.62E-06
plasma membrane bounded cell projection organization (GO:0120036)	1068	-	0.56	2.78E-07	4.41E-05
negative regulation of signaling (GO:0023057)	1305	-	0.56	7.89E-09	2.18E-06
positive regulation of organelle organization (GO:0010638)	583	-	0.55	1.64E-04	1.13E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of anatomical structure size (GO:0090066)	521	-	0.55	3.79E-04	2.27E-02
negative regulation of signal transduction (GO:0009968)	1204	-	0.55	2.56E-08	5.35E-06
regulation of ion transport (GO:0043269)	685	-	0.55	3.03E-05	2.72E-03
tissue morphogenesis (GO:0048729)	542	-	0.55	2.80E-04	1.76E-02
positive regulation of transport (GO:0051050)	909	-	0.55	1.29E-06	1.71E-04
chromatin organization (GO:0006325)	686	-	0.55	3.03E-05	2.71E-03
cell projection organization (GO:0030030)	1119	-	0.55	5.90E-08	1.10E-05
nervous system development (GO:0007399)	2193	-	0.55	5.36E-15	4.96E-12
response to peptide (GO:1901652)	450	-	0.54	7.71E-04	3.91E-02
generation of neurons (GO:0048699)	1237	-	0.54	4.16E-09	1.21E-06
positive regulation of cellular catabolic process (GO:0031331)	425	-	0.53	9.42E-04	4.57E-02
regulation of catabolic process (GO:0009894)	956	-	0.53	1.91E-07	3.13E-05
regulation of system process (GO:0044057)	553	-	0.52	7.82E-05	6.18E-03
neurogenesis (GO:0022008)	1361	-	0.52	1.20E-10	5.55E-08
morphogenesis of an epithelium (GO:0002009)	437	-	0.52	4.69E-04	2.70E-02
regulation of cellular component size (GO:0032535)	389	-	0.52	9.88E-04	4.76E-02
tube morphogenesis (GO:0035239)	644	-	0.52	1.23E-05	1.21E-03
positive regulation of cell differentiation (GO:0045597)	833	-	0.51	5.58E-07	8.05E-05
positive regulation of catabolic process (GO:0009896)	494	-	0.51	1.41E-04	9.95E-03
regulation of ion transmembrane transport (GO:0034765)	481	-	0.51	1.54E-04	1.07E-02
eye development (GO:0001654)	366	-	0.5	9.25E-04	4.50E-02
neuron development (GO:0048666)	810	-	0.49	1.45E-07	2.45E-05
neuron projection development (GO:0031175)	651	-	0.48	2.74E-06	3.34E-04
regulation of cation transmembrane transport (GO:1904062)	347	-	0.48	6.63E-04	3.51E-02
supramolecular fiber organization (GO:0097435)	512	-	0.48	2.82E-05	2.56E-03
regulation of apoptotic signaling pathway (GO:2001233)	348	-	0.48	6.67E-04	3.52E-02
neuron differentiation (GO:0030182)	1008	-	0.48	1.54E-09	4.94E-07
cell junction organization (GO:0034330)	477	-	0.48	5.40E-05	4.49E-03
cellular component morphogenesis (GO:0032989)	569	-	0.48	8.28E-06	8.57E-04
positive regulation of cell projection organization (GO:0031346)	350	-	0.47	5.14E-04	2.93E-02
response to wounding (GO:0009611)	428	-	0.47	1.11E-04	8.11E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of cellular response to stress (GO:0080135)	637	-	0.47	1.51E-06	1.95E-04
modulation of chemical synaptic transmission (GO:0050804)	431	-	0.47	8.52E-05	6.57E-03
cell morphogenesis (GO:0000902)	675	-	0.47	6.00E-07	8.49E-05
regulation of trans-synaptic signaling (GO:0099177)	432	-	0.47	8.59E-05	6.58E-03
regulation of anatomical structure morphogenesis (GO:0022603)	908	-	0.46	3.32E-09	9.85E-07
response to decreased oxygen levels (GO:0036293)	307	-	0.46	8.22E-04	4.09E-02
sensory organ development (GO:0007423)	558	-	0.45	3.42E-06	3.92E-04
regulation of body fluid levels (GO:0050878)	367	-	0.45	1.84E-04	1.26E-02
membrane organization (GO:0061024)	783	-	0.45	1.59E-08	3.61E-06
embryonic morphogenesis (GO:0048598)	568	-	0.45	2.05E-06	2.53E-04
regulation of hemopoiesis (GO:1903706)	353	-	0.45	2.62E-04	1.67E-02
response to oxygen levels (GO:0070482)	334	-	0.44	3.55E-04	2.15E-02
cellular response to peptide (GO:1901653)	297	-	0.44	6.70E-04	3.52E-02
regulation of leukocyte differentiation (GO:1902105)	279	-	0.44	9.24E-04	4.51E-02
negative regulation of intracellular signal transduction (GO:1902532)	501	-	0.44	5.64E-06	6.24E-04
regulation of DNA-binding transcription factor activity (GO:0051090)	424	-	0.43	2.60E-05	2.38E-03
endocytosis (GO:0006897)	509	-	0.43	3.30E-06	3.84E-04
embryonic organ development (GO:0048568)	433	-	0.42	1.52E-05	1.45E-03
neuron projection morphogenesis (GO:0048812)	455	-	0.42	8.47E-06	8.71E-04
regulation of metal ion transport (GO:0010959)	394	-	0.42	3.76E-05	3.25E-03
regulation of cell projection organization (GO:0031344)	643	-	0.42	1.00E-07	1.77E-05
muscle system process (GO:0003012)	271	-	0.42	7.49E-04	3.86E-02
plasma membrane bounded cell projection morphogenesis (GO:0120039)	459	-	0.42	6.46E-06	6.86E-04
regulation of plasma membrane bounded cell projection organization (GO:0120035)	626	-	0.42	1.02E-07	1.78E-05
autophagy (GO:0006914)	272	-	0.42	7.58E-04	3.88E-02
process utilizing autophagic mechanism (GO:0061919)	272	-	0.42	7.58E-04	3.87E-02
cell part morphogenesis (GO:0032990)	483	-	0.42	3.64E-06	4.15E-04
cell projection morphogenesis (GO:0048858)	463	-	0.42	4.91E-06	5.48E-04

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of neuron projection development (GO:0010975)	443	-	0.41	8.94E-06	9.12E-04
cognition (GO:0050890)	299	-	0.41	2.37E-04	1.55E-02
actin cytoskeleton organization (GO:0030036)	539	-	0.41	4.90E-07	7.20E-05
cell-cell signaling by wnt (GO:0198738)	260	-	0.4	5.86E-04	3.27E-02
Wnt signaling pathway (GO:0016055)	260	-	0.4	5.86E-04	3.26E-02
regulation of Wnt signaling pathway (GO:0030111)	325	-	0.4	1.36E-04	9.65E-03
regulation of membrane potential (GO:0042391)	434	-	0.4	7.07E-06	7.41E-04
regulation of transporter activity (GO:0032409)	287	-	0.4	2.45E-04	1.59E-02
embryonic organ morphogenesis (GO:0048562)	292	-	0.39	1.86E-04	1.26E-02
actin filament organization (GO:0007015)	271	-	0.39	3.41E-04	2.07E-02
neuron projection guidance (GO:0097485)	228	-	0.38	8.22E-04	4.08E-02
actin filament-based process (GO:0030029)	599	-	0.38	2.15E-08	4.62E-06
axon development (GO:0061564)	375	-	0.37	1.02E-05	1.03E-03
epithelial tube morphogenesis (GO:0060562)	305	-	0.37	7.93E-05	6.20E-03
small GTPase mediated signal transduction (GO:0007264)	260	-	0.37	2.55E-04	1.63E-02
cellular response to extracellular stimulus (GO:0031668)	237	-	0.37	4.60E-04	2.68E-02
regulation of cell morphogenesis (GO:0022604)	310	-	0.37	6.02E-05	4.88E-03
cellular response to external stimulus (GO:0071496)	313	-	0.36	4.42E-05	3.75E-03
synapse organization (GO:0050808)	289	-	0.36	7.85E-05	6.17E-03
regulation of transmembrane transporter activity (GO:0022898)	270	-	0.36	1.45E-04	1.01E-02
response to mechanical stimulus (GO:0009612)	222	-	0.35	6.37E-04	3.43E-02
phagocytosis (GO:0006909)	249	-	0.35	1.89E-04	1.27E-02
regulation of canonical Wnt signaling pathway (GO:0060828)	250	-	0.35	1.90E-04	1.27E-02
cell surface receptor signaling pathway involved in cell-cell signaling (GO:1905114)	331	-	0.34	1.39E-05	1.34E-03
cell morphogenesis involved in differentiation (GO:0000904)	517	-	0.34	2.10E-08	4.59E-06
regulation of ion transmembrane transporter activity (GO:0032412)	259	-	0.34	1.04E-04	7.70E-03
antigen receptor-mediated signaling pathway (GO:0050851)	216	-	0.32	3.40E-04	2.07E-02
regulation of neuron differentiation (GO:0045664)	193	-	0.32	6.02E-04	3.32E-02
gliogenesis (GO:0042063)	224	-	0.31	1.82E-04	1.25E-02

GO biological process complete	**Homo sapiens - REFLIST (20595)**	**upload_1 (over/under)**	**upload_1 (fold Enrichment)**	**upload_1 (raw P-value)**	**upload_1 (FDR)**
cell recognition (GO:0008037)	228	-	0.31	1.33E-04	9.52E-03
learning or memory (GO:0007611)	257	-	0.31	3.96E-05	3.40E-03
axonogenesis (GO:0007409)	337	-	0.29	7.36E-07	1.01E-04
cell morphogenesis involved in neuron differentiation (GO:0048667)	408	-	0.28	3.62E-08	7.11E-06
muscle contraction (GO:0006936)	222	-	0.28	6.52E-05	5.23E-03
learning (GO:0007612)	147	-	0.24	6.51E-04	3.47E-02
membrane invagination (GO:0010324)	148	-	0.24	6.61E-04	3.51E-02
B cell receptor signaling pathway (GO:0050853)	130	-	0.2	7.93E-04	3.98E-02
phagocytosis, engulfment (GO:0006911)	132	-	0.2	5.44E-04	3.05E-02
response to hydrogen peroxide (GO:0042542)	121	-	0.14	2.88E-04	1.80E-02
nephron development (GO:0072006)	126	-	0.14	1.99E-04	1.32E-02
actin filament-based movement (GO:0030048)	93	-	0.09	9.04E-04	4.44E-02
nephron epithelium development (GO:0072009)	98	-	0.09	6.41E-04	3.44E-02
regulation of potassium ion transport (GO:0043266)	101	-	0.09	4.20E-04	2.47E-02
phagocytosis, recognition (GO:0006910)	107	-	0.08	1.85E-04	1.26E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)				
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18				
Analyzed List:	upload_1 (Homo sapiens)				
Reference List:	Homo sapiens (all genes in database)				
Test Type:	FISHER				
Correction:	FDR				

80-100 CAI quintile (Highest)

GO biological process complete	**Homo sapiens - REFLIST (20595)**	**upload_1 (1780)**	**upload_1 (expected)**	**upload_1 (over/under)**	**upload_1 (fold Enrichment)**	**upload_1 (raw P-value)**	**upload_1 (FDR)**
oxygen transport (GO:0015671)	14	9	1.21	+	7.44	3.66E-05	1.98E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1780)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
detoxification of copper ion (GO:0010273)	16	9	1.38	+	6.51	7.90E-05	2.76E-02
stress response to copper ion (GO:1990169)	16	9	1.38	+	6.51	7.90E-05	2.70E-02
gas transport (GO:0015669)	19	10	1.64	+	6.09	4.85E-05	2.31E-02
keratinization (GO:0031424)	58	22	5.01	+	4.39	1.71E-07	8.96E-04
mucosal immune response (GO:0002385)	32	12	2.77	+	4.34	1.18E-04	3.79E-02
keratinocyte differentiation (GO:0030216)	108	28	9.33	+	3	2.59E-06	3.71E-03
nucleosome assembly (GO:0006334)	101	25	8.73	+	2.86	1.72E-05	1.28E-02
antimicrobial humoral immune response mediated by antimicrobial peptide (GO:0061844)	102	25	8.82	+	2.84	1.98E-05	1.35E-02
hair cycle (GO:0042633)	91	22	7.87	+	2.8	7.24E-05	2.92E-02
molting cycle (GO:0042303)	91	22	7.87	+	2.8	7.24E-05	2.84E-02
antimicrobial humoral response (GO:0019730)	141	34	12.19	+	2.79	1.42E-06	2.48E-03
epidermis development (GO:0008544)	279	61	24.11	+	2.53	1.72E-09	2.70E-05
chromatin assembly (GO:0031497)	134	29	11.58	+	2.5	5.29E-05	2.45E-02
nucleosome organization (GO:0034728)	140	30	12.1	+	2.48	4.02E-05	2.04E-02
DNA packaging (GO:0006323)	180	36	15.56	+	2.31	2.10E-05	1.32E-02
epidermal cell differentiation (GO:0009913)	161	32	13.92	+	2.3	7.34E-05	2.75E-02
skin development (GO:0043588)	232	45	20.05	+	2.24	5.01E-06	6.06E-03
sensory perception of light stimulus (GO:0050953)	220	42	19.01	+	2.21	1.45E-05	1.20E-02
visual perception (GO:0007601)	217	41	18.76	+	2.19	2.17E-05	1.31E-02
muscle contraction (GO:0006936)	222	39	19.19	+	2.03	1.23E-04	3.87E-02
adenylate cyclase-modulating G protein-coupled receptor signaling pathway (GO:0007188)	223	39	19.27	+	2.02	1.30E-04	4.02E-02
negative regulation of growth (GO:0045926)	247	43	21.35	+	2.01	6.91E-05	2.94E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1780)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
tissue development (GO:0009888)	1597	201	138.03	+	1.46	5.27E-07	1.38E-03
response to oxygen-containing compound (GO:1901700)	1567	183	135.43	+	1.35	1.07E-04	3.52E-02
animal organ development (GO:0048513)	3079	335	266.11	+	1.26	2.32E-05	1.35E-02
anatomical structure development (GO:0048856)	5001	515	432.23	+	1.19	1.69E-05	1.33E-02
multicellular organismal process (GO:0032501)	6584	663	569.05	+	1.17	6.27E-06	7.04E-03
regulation of macromolecule metabolic process (GO:0060255)	6083	447	525.75	-	0.85	7.20E-05	2.98E-02
regulation of nitrogen compound metabolic process (GO:0051171)	5537	400	478.56	-	0.84	4.36E-05	2.14E-02
regulation of gene expression (GO:0010468)	4749	337	410.45	-	0.82	5.33E-05	2.39E-02
regulation of biosynthetic process (GO:0009889)	4152	290	358.85	-	0.81	7.25E-05	2.78E-02
regulation of cellular biosynthetic process (GO:0031326)	4088	284	353.32	-	0.8	5.75E-05	2.51E-02
regulation of transcription, DNA-templated (GO:0006355)	3415	226	295.15	-	0.77	1.33E-05	1.23E-02
regulation of nucleic acid-templated transcription (GO:1903506)	3416	226	295.24	-	0.77	1.33E-05	1.16E-02
regulation of RNA biosynthetic process (GO:2001141)	3421	226	295.67	-	0.76	1.15E-05	1.13E-02
organic cyclic compound metabolic process (GO:1901360)	2922	191	252.54	-	0.76	3.40E-05	1.91E-02
regulation of macromolecule biosynthetic process (GO:0010556)	3935	256	340.1	-	0.75	5.43E-07	1.22E-03
regulation of cellular macromolecule biosynthetic process (GO:2000112)	3906	253	337.59	-	0.75	4.25E-07	1.34E-03
regulation of RNA metabolic process (GO:0051252)	3698	236	319.61	-	0.74	2.82E-07	1.11E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1780)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
cellular aromatic compound metabolic process (GO:0006725)	2688	171	232.32	-	0.74	1.73E-05	1.23E-02
regulation of nucleobase-containing compound metabolic process (GO:0019219)	3931	250	339.75	-	0.74	8.06E-08	6.34E-04
heterocycle metabolic process (GO:0046483)	2625	164	226.88	-	0.72	8.30E-06	8.70E-03
positive regulation of nucleobase-containing compound metabolic process (GO:0045935)	1902	117	164.39	-	0.71	1.02E-04	3.43E-02
nucleobase-containing compound metabolic process (GO:0006139)	2474	150	213.82	-	0.7	3.24E-06	4.24E-03
nucleic acid metabolic process (GO:0090304)	2007	120	173.46	-	0.69	2.01E-05	1.32E-02
regulation of DNA metabolic process (GO:0051052)	358	10	30.94	-	0.32	3.71E-05	1.94E-02
sensory perception of smell (GO:0007608)	468	13	40.45	-	0.32	1.47E-06	2.32E-03
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	11	38.29	-	0.29	6.89E-07	1.35E-03
regulation of chromosome organization (GO:0033044)	286	7	24.72	-	0.28	7.71E-05	2.82E-02
positive regulation of DNA metabolic process (GO:0051054)	203	3	17.55	-	0.17	7.73E-05	2.76E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

60-80 CAI quintile

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1702)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger (GO:0007187)	67	22	5.54	+	3.97	6.32E-07	3.31E-03
potassium ion transmembrane transport (GO:0071805)	158	31	13.06	+	2.37	4.67E-05	4.90E-02
G protein-coupled receptor signaling pathway (GO:0007186)	1216	164	100.49	+	1.63	9.43E-09	1.48E-04
regulation of ion transport (GO:0043269)	685	92	56.61	+	1.63	2.46E-05	3.87E-02
response to lipid (GO:0033993)	829	106	68.51	+	1.55	3.97E-05	4.80E-02
regulation of transcription, DNA-templated (GO:0006355)	3415	349	282.22	+	1.24	5.23E-05	5.14E-02
regulation of nucleic acid-templated transcription (GO:1903506)	3416	349	282.3	+	1.24	5.25E-05	4.85E-02
cellular component organization or biogenesis (GO:0071840)	5496	376	454.2	-	0.83	2.87E-05	3.76E-02
cellular component organization (GO:0016043)	5298	355	437.83	-	0.81	6.88E-06	1.35E-02
organelle organization (GO:0006996)	3360	198	277.68	-	0.71	1.88E-07	1.48E-03
macromolecule localization (GO:0033036)	2308	136	190.74	-	0.71	2.61E-05	3.73E-02
cellular localization (GO:0051641)	2264	125	187.1	-	0.67	1.15E-06	4.54E-03
protein localization (GO:0008104)	1901	101	157.1	-	0.64	1.79E-06	5.61E-03
cellular macromolecule localization (GO:0070727)	1386	73	114.54	-	0.64	4.34E-05	4.87E-02
establishment of localization in cell (GO:0051649)	1618	82	133.71	-	0.61	1.88E-06	4.92E-03
intracellular transport (GO:0046907)	1247	62	103.05	-	0.6	2.11E-05	3.68E-02
cell cycle (GO:0007049)	1205	57	99.58	-	0.57	5.51E-06	1.24E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1702)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

40-60 CAI quintile

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1687)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	135	36.29	+	3.72	3.21E-33	5.04E-29
sensory perception of smell (GO:0007608)	468	135	38.34	+	3.52	3.35E-31	1.75E-27
detection of chemical stimulus involved in sensory perception (GO:0050907)	488	139	39.97	+	3.48	1.23E-31	9.69E-28
detection of chemical stimulus (GO:0009593)	524	143	42.92	+	3.33	6.88E-31	2.71E-27
sensory perception of chemical stimulus (GO:0007606)	542	139	44.4	+	3.13	9.35E-28	2.10E-24
detection of stimulus involved in sensory perception (GO:0050906)	558	143	45.71	+	3.13	1.64E-28	5.16E-25
detection of stimulus (GO:0051606)	694	156	56.85	+	2.74	8.63E-26	1.70E-22
G protein-coupled receptor signaling pathway (GO:0007186)	1216	229	99.61	+	2.3	4.55E-28	1.19E-24
sensory perception (GO:0007600)	978	168	80.11	+	2.1	4.75E-17	8.29E-14
nervous system process (GO:0050877)	1418	195	116.15	+	1.68	3.05E-11	4.36E-08
system process (GO:0003008)	2040	232	167.1	+	1.39	1.41E-06	6.31E-04
regulation of transcription by RNA polymerase II (GO:0006357)	2538	275	207.9	+	1.32	4.59E-06	1.50E-03
regulation of transcription, DNA-templated (GO:0006355)	3415	356	279.73	+	1.27	3.60E-06	1.31E-03
regulation of nucleic acid-templated transcription (GO:1903506)	3416	356	279.82	+	1.27	3.61E-06	1.29E-03
regulation of RNA biosynthetic process (GO:2001141)	3421	356	280.22	+	1.27	4.29E-06	1.43E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1687)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of RNA metabolic process (GO:0051252)	3698	367	302.91	+	1.21	1.41E-04	2.41E-02
regulation of cellular macromolecule biosynthetic process (GO:2000112)	3906	383	319.95	+	1.2	2.48E-04	3.82E-02
cellular process (GO:0009987)	15074	1159	1234.76	-	0.94	8.43E-05	1.79E-02
positive regulation of biological process (GO:0048518)	6062	426	496.56	-	0.86	2.59E-04	3.95E-02
positive regulation of cellular process (GO:0048522)	5580	385	457.08	-	0.84	1.26E-04	2.22E-02
organonitrogen compound metabolic process (GO:1901564)	4787	324	392.12	-	0.83	1.29E-04	2.26E-02
cellular macromolecule metabolic process (GO:0044260)	4442	300	363.86	-	0.82	2.29E-04	3.60E-02
nitrogen compound metabolic process (GO:0006807)	6360	425	520.97	-	0.82	8.19E-07	3.78E-04
anatomical structure development (GO:0048856)	5001	334	409.65	-	0.82	2.49E-05	6.20E-03
macromolecule metabolic process (GO:0043170)	5743	383	470.43	-	0.81	3.69E-06	1.29E-03
developmental process (GO:0032502)	5565	371	455.85	-	0.81	5.16E-06	1.62E-03
metabolic process (GO:0008152)	7882	524	645.64	-	0.81	2.89E-09	2.39E-06
system development (GO:0048731)	4169	276	341.5	-	0.81	9.62E-05	1.96E-02
cellular metabolic process (GO:0044237)	7029	465	575.77	-	0.81	2.78E-08	1.90E-05
primary metabolic process (GO:0044238)	6869	451	562.66	-	0.8	1.68E-08	1.20E-05
multicellular organism development (GO:0007275)	4525	295	370.66	-	0.8	1.23E-05	3.46E-03
organic substance metabolic process (GO:0071704)	7404	480	606.48	-	0.79	3.83E-10	3.77E-07
response to stress (GO:0006950)	3336	214	273.26	-	0.78	1.21E-04	2.18E-02
macromolecule modification (GO:0043412)	2825	180	231.4	-	0.78	3.20E-04	4.75E-02
cellular component organization or biogenesis (GO:0071840)	5496	349	450.19	-	0.78	4.31E-08	2.71E-05

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1687)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
cellular nitrogen compound metabolic process (GO:0034641)	3193	200	261.55	-	0.76	4.25E-05	9.97E-03
biosynthetic process (GO:0009058)	2420	150	198.23	-	0.76	3.05E-04	4.56E-02
organic cyclic compound metabolic process (GO:1901360)	2922	180	239.35	-	0.75	4.09E-05	9.88E-03
cellular component organization (GO:0016043)	5298	326	433.98	-	0.75	2.58E-09	2.25E-06
nucleobase-containing compound metabolic process (GO:0006139)	2474	151	202.65	-	0.75	1.17E-04	2.16E-02
heterocycle metabolic process (GO:0046483)	2625	160	215.02	-	0.74	6.67E-05	1.52E-02
cellular component biogenesis (GO:0044085)	2481	151	203.23	-	0.74	1.01E-04	1.98E-02
cellular aromatic compound metabolic process (GO:0006725)	2688	163	220.18	-	0.74	4.12E-05	9.82E-03
organelle organization (GO:0006996)	3360	200	275.23	-	0.73	7.42E-07	3.54E-04
regulation of localization (GO:0032879)	2709	161	221.9	-	0.73	1.13E-05	3.24E-03
regulation of biological quality (GO:0065008)	3705	220	303.49	-	0.72	1.42E-07	7.70E-05
localization (GO:0051179)	5081	298	416.2	-	0.72	3.05E-11	4.79E-08
regulation of response to stimulus (GO:0048583)	3805	218	311.68	-	0.7	4.25E-09	3.34E-06
regulation of protein metabolic process (GO:0051246)	2492	142	204.13	-	0.7	3.51E-06	1.31E-03
regulation of molecular function (GO:0065009)	3060	173	250.65	-	0.69	9.61E-08	5.40E-05
nitrogen compound transport (GO:0071705)	1578	89	129.26	-	0.69	2.07E-04	3.28E-02
regulation of cellular protein metabolic process (GO:0032268)	2328	131	190.69	-	0.69	3.90E-06	1.33E-03
regulation of signal transduction (GO:0009966)	2867	161	234.84	-	0.69	1.76E-07	9.25E-05
regulation of transport (GO:0051049)	1732	97	141.87	-	0.68	8.48E-05	1.78E-02
establishment of localization (GO:0051234)	3911	219	320.36	-	0.68	2.35E-10	3.08E-07
negative regulation of response to stimulus (GO:0048585)	1559	87	127.7	-	0.68	1.86E-04	2.99E-02
positive regulation of response to stimulus (GO:0048584)	2080	116	170.38	-	0.68	9.54E-06	2.78E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1687)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
positive regulation of cell communication (GO:0010647)	1632	91	133.68	-	0.68	1.19E-04	2.17E-02
cellular component assembly (GO:0022607)	2251	125	184.39	-	0.68	2.80E-06	1.07E-03
positive regulation of signaling (GO:0023056)	1639	91	134.26	-	0.68	1.00E-04	2.00E-02
transport (GO:0006810)	3752	208	307.34	-	0.68	2.78E-10	3.12E-07
regulation of cell communication (GO:0010646)	3242	179	265.56	-	0.67	5.08E-09	3.80E-06
cellular localization (GO:0051641)	2264	125	185.45	-	0.67	1.96E-06	8.35E-04
small molecule metabolic process (GO:0044281)	1547	85	126.72	-	0.67	9.84E-05	1.98E-02
regulation of catalytic activity (GO:0050790)	2393	131	196.02	-	0.67	5.61E-07	2.76E-04
regulation of signaling (GO:0023051)	3257	178	266.79	-	0.67	2.26E-09	2.09E-06
intracellular signal transduction (GO:0035556)	1464	80	119.92	-	0.67	1.45E-04	2.43E-02
organic substance transport (GO:0071702)	1927	105	157.85	-	0.67	8.44E-06	2.50E-03
macromolecule localization (GO:0033036)	2308	123	189.06	-	0.65	2.31E-07	1.17E-04
protein localization (GO:0008104)	1901	100	155.72	-	0.64	2.01E-06	8.08E-04
cellular response to stress (GO:0033554)	1519	79	124.43	-	0.63	1.85E-05	5.00E-03
regulation of cellular component organization (GO:0051128)	2359	122	193.23	-	0.63	3.14E-08	2.06E-05
regulation of organelle organization (GO:0033043)	1189	61	97.39	-	0.63	1.12E-04	2.10E-02
regulation of intracellular signal transduction (GO:1902531)	1655	84	135.57	-	0.62	2.74E-06	1.08E-03
cell projection organization (GO:0030030)	1119	56	91.66	-	0.61	8.92E-05	1.85E-02
plasma membrane bounded cell projection organization (GO:0120036)	1068	53	87.48	-	0.61	1.24E-04	2.22E-02
regulation of cellular component biogenesis (GO:0044087)	950	46	77.82	-	0.59	1.75E-04	2.84E-02
positive regulation of cellular component organization (GO:0051130)	1119	54	91.66	-	0.59	3.29E-05	8.08E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1687)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of anatomical structure morphogenesis (GO:0022603)	908	42	74.38	-	0.56	7.45E-05	1.63E-02
cell cycle process (GO:0022402)	802	37	65.69	-	0.56	2.37E-04	3.70E-02
cytoskeleton organization (GO:0007010)	1169	52	95.76	-	0.54	2.00E-06	8.29E-04
regulation of cellular localization (GO:0060341)	820	36	67.17	-	0.54	5.50E-05	1.27E-02
phosphorylation (GO:0016310)	874	38	71.59	-	0.53	2.31E-05	5.86E-03
vesicle-mediated transport (GO:0016192)	1329	57	108.86	-	0.52	7.69E-08	4.47E-05
phosphorus metabolic process (GO:0006793)	1779	76	145.72	-	0.52	2.70E-10	3.27E-07
phosphate-containing compound metabolic process (GO:0006796)	1758	75	144	-	0.52	2.96E-10	3.10E-07
protein phosphorylation (GO:0006468)	683	28	55.95	-	0.5	6.92E-05	1.55E-02
organophosphate metabolic process (GO:0019637)	833	33	68.23	-	0.48	5.36E-06	1.65E-03
regulation of cellular response to stress (GO:0080135)	637	25	52.18	-	0.48	6.93E-05	1.53E-02
peptidyl-amino acid modification (GO:0018193)	851	33	69.71	-	0.47	1.94E-06	8.48E-04
regulation of cell projection organization (GO:0031344)	643	24	52.67	-	0.46	2.07E-05	5.51E-03
positive regulation of cellular component biogenesis (GO:0044089)	514	19	42.1	-	0.45	1.72E-04	2.81E-02
regulation of plasma membrane bounded cell projection organization (GO:0120035)	626	23	51.28	-	0.45	2.20E-05	5.75E-03
negative regulation of intracellular signal transduction (GO:1902532)	501	18	41.04	-	0.44	1.40E-04	2.42E-02
endocytosis (GO:0006897)	509	18	41.69	-	0.43	8.11E-05	1.75E-02
adaptive immune response (GO:0002250)	640	22	52.42	-	0.42	4.69E-06	1.50E-03
microtubule-based process (GO:0007017)	765	24	62.66	-	0.38	6.91E-08	4.18E-05
microtubule-based movement (GO:0007018)	356	11	29.16	-	0.38	3.03E-04	4.59E-02
microtubule cytoskeleton organization (GO:0000226)	521	16	42.68	-	0.37	7.11E-06	2.15E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1687)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
organophosphate biosynthetic process (GO:0090407)	490	15	40.14	-	0.37	1.73E-05	4.77E-03
regulation of neuron projection development (GO:0010975)	443	13	36.29	-	0.36	2.25E-05	5.79E-03
regulation of apoptotic signaling pathway (GO:2001233)	348	10	28.51	-	0.35	1.69E-04	2.80E-02
cell recognition (GO:0008037)	228	4	18.68	-	0.21	1.43E-04	2.42E-02
immunoglobulin production (GO:0002377)	155	1	12.7	-	0.08	1.05E-04	2.03E-02
regulation of axonogenesis (GO:0050770)	156	1	12.78	-	0.08	1.05E-04	2.00E-02
positive regulation of apoptotic signaling pathway (GO:2001235)	122	0	9.99	-	< 0.01	1.05E-04	2.01E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

20-40 CAI quintile

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1468)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	120	31.58	+	3.8	6.64E-31	1.04E-26
sensory perception of smell (GO:0007608)	468	122	33.36	+	3.66	4.25E-30	2.22E-26
detection of chemical stimulus involved in	488	126	34.78	+	3.62	9.90E-31	7.78E-27

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1468)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
sensory perception (GO:0050907)							
detection of chemical stimulus (GO:0009593)	524	128	37.35	+	3.43	3.01E-29	1.18E-25
sensory perception of chemical stimulus (GO:0007606)	542	129	38.63	+	3.34	1.48E-28	4.67E-25
detection of stimulus involved in sensory perception (GO:0050906)	558	130	39.77	+	3.27	5.12E-28	1.34E-24
detection of stimulus (GO:0051606)	694	136	49.47	+	2.75	4.59E-23	9.01E-20
G protein-coupled receptor signaling pathway (GO:0007186)	1216	197	86.68	+	2.27	5.48E-24	1.23E-20
sensory perception (GO:0007600)	978	152	69.71	+	2.18	3.54E-17	6.18E-14
nervous system process (GO:0050877)	1418	180	101.07	+	1.78	1.25E-12	1.97E-09
system process (GO:0003008)	2040	221	145.41	+	1.52	2.66E-09	3.80E-06
positive regulation of biological process (GO:0048518)	6062	349	432.1	-	0.81	2.93E-06	2.42E-03
positive regulation of cellular process (GO:0048522)	5580	320	397.74	-	0.8	7.09E-06	4.64E-03
cellular component organization or biogenesis (GO:0071840)	5496	312	391.75	-	0.8	3.54E-06	2.65E-03
localization (GO:0051179)	5081	286	362.17	-	0.79	5.64E-06	3.85E-03
transport (GO:0006810)	3752	206	267.44	-	0.77	3.73E-05	1.63E-02
establishment of localization (GO:0051234)	3911	213	278.77	-	0.76	1.44E-05	8.40E-03
regulation of response to stimulus (GO:0048583)	3805	207	271.22	-	0.76	1.88E-05	9.85E-03
cellular component organization (GO:0016043)	5298	287	377.64	-	0.76	8.66E-08	1.13E-04
regulation of cellular protein metabolic process (GO:0032268)	2328	120	165.94	-	0.72	1.51E-04	4.15E-02
organelle organization (GO:0006996)	3360	173	239.5	-	0.72	2.77E-06	2.42E-03
immune system process (GO:0002376)	2370	118	168.93	-	0.7	2.76E-05	1.35E-02
cell surface receptor signaling pathway (GO:0007166)	2030	101	144.7	-	0.7	1.19E-04	3.67E-02
regulation of molecular function (GO:0065009)	3060	149	218.12	-	0.68	3.24E-07	3.64E-04
cellular localization (GO:0051641)	2264	110	161.38	-	0.68	1.52E-05	8.56E-03

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1468)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of catalytic activity (GO:0050790)	2393	116	170.57	-	0.68	7.52E-06	4.73E-03
regulation of cellular component organization (GO:0051128)	2359	113	168.15	-	0.67	4.42E-06	3.16E-03
immune response (GO:0006955)	1678	80	119.61	-	0.67	1.44E-04	4.12E-02
regulation of localization (GO:0032879)	2709	129	193.1	-	0.67	5.29E-07	5.54E-04
establishment of localization in cell (GO:0051649)	1618	75	115.33	-	0.65	7.24E-05	2.71E-02
positive regulation of molecular function (GO:0044093)	1497	69	106.71	-	0.65	1.18E-04	3.72E-02
regulation of protein modification process (GO:0031399)	1483	68	105.71	-	0.64	1.10E-04	3.69E-02
regulation of phosphate metabolic process (GO:0019220)	1379	62	98.29	-	0.63	1.24E-04	3.73E-02
regulation of phosphorus metabolic process (GO:0051174)	1379	62	98.29	-	0.63	1.24E-04	3.66E-02
vesicle-mediated transport (GO:0016192)	1329	59	94.73	-	0.62	1.17E-04	3.77E-02
positive regulation of protein metabolic process (GO:0051247)	1455	64	103.71	-	0.62	3.76E-05	1.60E-02
positive regulation of cellular protein metabolic process (GO:0032270)	1366	58	97.37	-	0.6	2.17E-05	1.10E-02
regulation of phosphorylation (GO:0042325)	1236	52	88.1	-	0.59	5.14E-05	2.07E-02
regulation of transport (GO:0051049)	1732	72	123.46	-	0.58	5.79E-07	5.69E-04
positive regulation of catalytic activity (GO:0043085)	1134	47	80.83	-	0.58	7.02E-05	2.69E-02
positive regulation of protein modification process (GO:0031401)	970	39	69.14	-	0.56	1.28E-04	3.74E-02
regulation of protein phosphorylation (GO:0001932)	1086	43	77.41	-	0.56	2.90E-05	1.38E-02
programmed cell death (GO:0012501)	942	36	67.15	-	0.54	5.74E-05	2.25E-02
cytoskeleton organization (GO:0007010)	1169	44	83.33	-	0.53	3.44E-06	2.70E-03
cell death (GO:0008219)	978	36	69.71	-	0.52	1.42E-05	8.60E-03
apoptotic process (GO:0006915)	905	33	64.51	-	0.51	2.99E-05	1.34E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1468)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of organelle organization (GO:0033043)	1189	43	84.75	-	0.51	9.15E-07	8.46E-04
positive regulation of protein phosphorylation (GO:0001934)	723	26	51.54	-	0.5	1.76E-04	4.77E-02
microtubule-based process (GO:0007017)	765	27	54.53	-	0.5	7.97E-05	2.78E-02
membrane organization (GO:0061024)	783	27	55.81	-	0.48	3.80E-05	1.57E-02
regulation of cellular localization (GO:0060341)	820	28	58.45	-	0.48	1.67E-05	9.03E-03
adaptive immune response (GO:0002250)	640	21	45.62	-	0.46	9.39E-05	3.21E-02
regulation of transmembrane transport (GO:0034762)	569	18	40.56	-	0.44	1.44E-04	4.05E-02
regulation of ion transmembrane transport (GO:0034765)	481	13	34.29	-	0.38	7.56E-05	2.76E-02
endocytosis (GO:0006897)	509	9	36.28	-	0.25	2.41E-07	2.91E-04
regulation of ion transmembrane transporter activity (GO:0032412)	259	3	18.46	-	0.16	2.91E-05	1.34E-02
plasma membrane invagination (GO:0099024)	141	0	10.05	-	< 0.01	1.15E-04	3.77E-02
membrane invagination (GO:0010324)	148	0	10.55	-	< 0.01	7.94E-05	2.84E-02
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

0-20 CAI quintile (Lowest)

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1173)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
NLS-bearing protein import into nucleus (GO:0006607)	19	8	1.08	+	7.39	6.14E-05	2.35E-02
negative regulation of execution phase of apoptosis (GO:1900118)	23	9	1.31	+	6.87	3.38E-05	1.47E-02

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1173)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
detection of chemical stimulus involved in sensory perception of bitter taste (GO:0001580)	36	13	2.05	+	6.34	1.29E-06	1.35E-03
sensory perception of bitter taste (GO:0050913)	41	13	2.34	+	5.57	4.23E-06	3.92E-03
detection of chemical stimulus involved in sensory perception of taste (GO:0050912)	42	13	2.39	+	5.43	5.27E-06	3.95E-03
mitochondrial transmembrane transport (GO:1990542)	97	24	5.52	+	4.34	2.53E-08	5.68E-05
transcription initiation from RNA polymerase II promoter (GO:0006367)	54	13	3.08	+	4.23	5.05E-05	2.09E-02
sensory perception of taste (GO:0050909)	68	15	3.87	+	3.87	3.30E-05	1.48E-02
DNA-templated transcription, initiation (GO:0006352)	88	18	5.01	+	3.59	1.37E-05	7.41E-03
protein localization to mitochondrion (GO:0070585)	79	15	4.5	+	3.33	1.45E-04	3.99E-02
mitochondrial transport (GO:0006839)	207	37	11.79	+	3.14	1.20E-08	3.76E-05
protein import (GO:0017038)	150	25	8.54	+	2.93	8.02E-06	5.25E-03
detection of chemical stimulus involved in sensory perception (GO:0050907)	488	70	27.79	+	2.52	5.20E-11	8.18E-07
detection of chemical stimulus (GO:0009593)	524	72	29.84	+	2.41	1.35E-10	7.09E-07
meiotic cell cycle (GO:0051321)	233	32	13.27	+	2.41	2.34E-05	1.08E-02
sensory perception of chemical stimulus (GO:0007606)	542	74	30.87	+	2.4	8.83E-11	6.94E-07
detection of stimulus involved in sensory perception (GO:0050906)	558	73	31.78	+	2.3	8.53E-10	3.35E-06
establishment of protein localization to organelle (GO:0072594)	300	39	17.09	+	2.28	8.24E-06	5.18E-03
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	443	56	25.23	+	2.22	1.95E-07	3.41E-04
sensory perception of smell (GO:0007608)	468	58	26.66	+	2.18	2.29E-07	3.60E-04
detection of stimulus (GO:0051606)	694	80	39.53	+	2.02	2.23E-08	5.84E-05

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1173)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
G protein-coupled receptor signaling pathway (GO:0007186)	1216	109	69.26	+	1.57	9.90E-06	5.76E-03
heterocycle metabolic process (GO:0046483)	2625	197	149.51	+	1.32	9.89E-05	3.24E-02
positive regulation of cellular process (GO:0048522)	5580	255	317.81	-	0.8	4.68E-05	1.99E-02
positive regulation of biological process (GO:0048518)	6062	277	345.26	-	0.8	1.48E-05	7.74E-03
anatomical structure development (GO:0048856)	5001	228	284.83	-	0.8	1.26E-04	3.68E-02
multicellular organism development (GO:0007275)	4525	197	257.72	-	0.76	1.98E-05	9.73E-03
regulation of RNA metabolic process (GO:0051252)	3698	160	210.62	-	0.76	1.15E-04	3.42E-02
cellular developmental process (GO:0048869)	3463	147	197.24	-	0.75	8.83E-05	3.02E-02
regulation of RNA biosynthetic process (GO:2001141)	3421	145	194.85	-	0.74	9.63E-05	3.22E-02
system development (GO:0048731)	4169	176	237.45	-	0.74	7.61E-06	5.20E-03
regulation of transcription, DNA-templated (GO:0006355)	3415	144	194.5	-	0.74	6.82E-05	2.49E-02
regulation of nucleic acid-templated transcription (GO:1903506)	3416	144	194.56	-	0.74	6.82E-05	2.55E-02
cell differentiation (GO:0030154)	3403	140	193.82	-	0.72	2.26E-05	1.08E-02
regulation of response to stimulus (GO:0048583)	3805	156	216.72	-	0.72	4.24E-06	3.70E-03
regulation of multicellular organismal process (GO:0051239)	2578	103	146.83	-	0.7	1.00E-04	3.21E-02
immune system process (GO:0002376)	2370	94	134.98	-	0.7	1.47E-04	3.98E-02
nervous system development (GO:0007399)	2193	85	124.9	-	0.68	1.27E-04	3.63E-02
regulation of signal transduction (GO:0009966)	2867	110	163.29	-	0.67	5.23E-06	4.11E-03
regulation of signaling (GO:0023051)	3257	123	185.5	-	0.66	3.23E-07	4.62E-04
regulation of cell communication (GO:0010646)	3242	121	184.65	-	0.66	1.93E-07	3.78E-04

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1173)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
regulation of developmental process (GO:0050793)	2368	88	134.87	-	0.65	1.21E-05	6.77E-03
regulation of intracellular signal transduction (GO:1902531)	1655	58	94.26	-	0.62	7.13E-05	2.55E-02
negative regulation of response to stimulus (GO:0048585)	1559	49	88.79	-	0.55	5.22E-06	4.32E-03
neuron differentiation (GO:0030182)	1008	30	57.41	-	0.52	1.00E-04	3.16E-02
positive regulation of developmental process (GO:0051094)	1267	36	72.16	-	0.5	3.42E-06	3.36E-03
negative regulation of cell communication (GO:0010648)	1301	36	74.1	-	0.49	1.28E-06	1.44E-03
positive regulation of cell differentiation (GO:0045597)	833	23	47.44	-	0.48	1.41E-04	3.96E-02
negative regulation of signaling (GO:0023057)	1305	36	74.33	-	0.48	9.91E-07	1.20E-03
regulation of anatomical structure morphogenesis (GO:0022603)	908	25	51.72	-	0.48	5.79E-05	2.33E-02
negative regulation of signal transduction (GO:0009968)	1204	31	68.57	-	0.45	5.05E-07	6.61E-04
cell adhesion (GO:0007155)	952	24	54.22	-	0.44	8.25E-06	4.98E-03
biological adhesion (GO:0022610)	958	24	54.56	-	0.44	6.25E-06	4.47E-03
adaptive immune response (GO:0002250)	640	15	36.45	-	0.41	1.01E-04	3.13E-02
regulation of MAPK cascade (GO:0043408)	658	15	37.48	-	0.4	5.86E-05	2.30E-02
actin filament-based process (GO:0030029)	599	13	34.12	-	0.38	8.13E-05	2.84E-02
supramolecular fiber organization (GO:0097435)	512	10	29.16	-	0.34	1.09E-04	3.30E-02
negative regulation of intracellular signal transduction (GO:1902532)	501	8	28.53	-	0.28	1.52E-05	7.70E-03
Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)						

GO biological process complete	Homo sapiens - REFLIST (20595)	upload_1 (1173)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						

Universally Expressed Transcripts

GO biological process complete	Homo sapiens - REFLIST (20589)	upload_1 (20)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
protection from natural killer cell mediated cytotoxicity (GO:0042270)	6	2	0.01	+	> 100	2.50E-05	2.45E-02
antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent (GO:0002486)	8	2	0.01	+	> 100	4.01E-05	3.49E-02
antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway (GO:0002484)	8	2	0.01	+	> 100	4.01E-05	3.31E-02
cytoplasmic translation (GO:0002181)	119	8	0.12	+	69.21	1.97E-13	3.09E-09
ribosome assembly (GO:0042255)	62	3	0.06	+	49.81	3.29E-05	3.03E-02
translation (GO:0006412)	351	9	0.34	+	26.4	1.94E-11	1.52E-07
peptide biosynthetic process (GO:0043043)	378	9	0.37	+	24.51	3.70E-11	1.94E-07
amide biosynthetic process (GO:0043604)	494	9	0.48	+	18.76	3.79E-10	1.48E-06
peptide metabolic process (GO:0006518)	511	9	0.5	+	18.13	5.08E-10	1.59E-06
cellular amide metabolic process (GO:0043603)	773	10	0.75	+	13.32	7.74E-10	2.02E-06
organonitrogen compound biosynthetic process (GO:1901566)	1252	11	1.22	+	9.04	4.41E-09	9.88E-06

GO biological process complete	Homo sapiens - REFLIST (20589)	upload_1 (20)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
cellular nitrogen compound biosynthetic process (GO:0044271)	1287	10	1.25	+	8	9.72E-08	1.91E-04
cellular macromolecule biosynthetic process (GO:0034645)	1384	9	1.34	+	6.69	2.44E-06	3.82E-03
macromolecule biosynthetic process (GO:0009059)	1412	9	1.37	+	6.56	2.87E-06	4.10E-03
gene expression (GO:0010467)	2014	10	1.96	+	5.11	5.98E-06	6.69E-03
cellular biosynthetic process (GO:0044249)	2318	11	2.25	+	4.89	2.38E-06	4.14E-03
organic substance biosynthetic process (GO:1901576)	2387	11	2.32	+	4.74	3.18E-06	4.16E-03
biosynthetic process (GO:0009058)	2453	11	2.38	+	4.62	4.17E-06	5.02E-03
cellular nitrogen compound metabolic process (GO:0034641)	3203	12	3.11	+	3.86	7.48E-06	7.82E-03
Analysis Type:	PANTHER Overrepresentation Test (Released 20220202)						
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5725227 Released 2020-11-01						
Analyzed List:	upload_1 (Homo sapiens)						
Reference List:	Homo sapiens (all genes in database)						
Test Type:	FISHER						
Correction:	FDR						