**Complete Motif Analysis of Sequence Requirements for Translation Initiation at Non-Aug Start Codons**

*Nucleic acids research*

A. J. Diaz de Arce, W. L. Noderer and C. L. Wang

Jan 25, 2018

The initiation of mRNA translation from start codons other than AUG was previously believed to be rare and of relatively low impact. More recently, evidence has suggested that as much as half of all translation initiation utilizes non-AUG start codons, codons that deviate from AUG by a single base. Furthermore, non-AUG start codons have been shown to be involved in regulation of expression and disease etiology. Yet the ability to gauge expression based on the sequence of a translation initiation site (start codon and its flanking bases) has been limited. Here we have performed a comprehensive analysis of translation initiation sites that utilize non-AUG start codons. By combining genetic-reporter, cell-sorting, and high-throughput sequencing technologies, we have analyzed the expression associated with all possible variants of the -4 to +4 positions of non-AUG translation initiation site motifs. This complete motif analysis revealed that 1) with the right sequence context, certain non-AUG start codons can generate expression comparable to that of AUG start codons, 2) sequence context affects each non-AUG start codon differently, and 3) initiation at non-AUG start codons is highly sensitive to changes in the flanking sequences. Complete motif analysis has the potential to be a key tool for experimental and diagnostic genomics.

<http://www.ncbi.nlm.nih.gov/pubmed/29228265>

**Brca-Deficient Mouse Mammary Tumor Organoids to Study Cancer-Drug Resistance**

*Nature methods*

A. A. Duarte, E. Gogola, N. Sachs, M. Barazas, S. Annunziato, R. d. R. J, A. Velds, S. Blatter, J. M. Houthuijzen, M. van de Ven, H. Clevers, P. Borst, J. Jonkers and S. Rottenberg

Feb, 2018

Poly(ADP-ribose) polymerase inhibition (PARPi) is a promising new therapeutic approach for the treatment of cancers that show homologous recombination deficiency (HRD). Despite the success of PARPi in targeting HRD in tumors that lack the tumor suppressor function of BRCA1 or BRCA2, drug resistance poses a major obstacle. We developed three-dimensional cancer organoids derived from genetically engineered mouse models (GEMMs) for BRCA1- and BRCA2-deficient cancers. Unlike conventional cell lines or mammospheres, organoid cultures can be efficiently derived and rapidly expanded in vitro. Orthotopically transplanted organoids give rise to mammary tumors that recapitulate the epithelial morphology and preserve the drug response of the original tumor. Notably, GEMM-tumor-derived organoids can be easily genetically modified, making them a powerful tool for genetic studies of tumor biology and drug resistance.

<http://www.ncbi.nlm.nih.gov/pubmed/29256493>

**Editing Activity for Eliminating Mischarged Trnas Is Essential in Mammalian Mitochondria**

*Nucleic acids research*

T. Hilander, X. L. Zhou, S. Konovalova, F. P. Zhang, L. Euro, D. Chilov, M. Poutanen, J. Chihade, E. D. Wang and H. Tyynismaa

Jan 25, 2018

Accuracy of protein synthesis is enabled by the selection of amino acids for tRNA charging by aminoacyl-tRNA synthetases (ARSs), and further enhanced by the proofreading functions of some of these enzymes for eliminating tRNAs mischarged with noncognate amino acids. Mouse models of editing-defective cytoplasmic alanyl-tRNA synthetase (AlaRS) have previously demonstrated the importance of proofreading for cytoplasmic protein synthesis, with embryonic lethal and progressive neurodegeneration phenotypes. Mammalian mitochondria import their own set of nuclear-encoded ARSs for translating critical polypeptides of the oxidative phosphorylation system, but the importance of editing by the mitochondrial ARSs for mitochondrial proteostasis has not been known. We demonstrate here that the human mitochondrial AlaRS is capable of editing mischarged tRNAs in vitro, and that loss of the proofreading activity causes embryonic lethality in mice. These results indicate that tRNA proofreading is essential in mammalian mitochondria, and cannot be overcome by other quality control mechanisms.

<http://www.ncbi.nlm.nih.gov/pubmed/29228266>

**Interactome Insider: A Structural Interactome Browser for Genomic Studies**

*Nature methods*

M. J. Meyer, J. F. Beltran, S. Liang, R. Fragoza, A. Rumack, J. Liang, X. Wei and H. Yu

Jan 1, 2018

We present Interactome INSIDER, a tool to link genomic variant information with structural protein-protein interactomes. Underlying this tool is the application of machine learning to predict protein interaction interfaces for 185,957 protein interactions with previously unresolved interfaces in human and seven model organisms, including the entire experimentally determined human binary interactome. Predicted interfaces exhibit functional properties similar to those of known interfaces, including enrichment for disease mutations and recurrent cancer mutations. Through 2,164 de novo mutagenesis experiments, we show that mutations of predicted and known interface residues disrupt interactions at a similar rate and much more frequently than mutations outside of predicted interfaces. To spur functional genomic studies, Interactome INSIDER (<http://interactomeinsider.yulab.org>) enables users to identify whether variants or disease mutations are enriched in known and predicted interaction interfaces at various resolutions. Users may explore known population variants, disease mutations, and somatic cancer mutations, or they may upload their own set of mutations for this purpose.

<http://www.ncbi.nlm.nih.gov/pubmed/29355848>

**Detecting Hierarchical Genome Folding with Network Modularity**

*Nature methods*

H. K. Norton, D. J. Emerson, H. Huang, J. Kim, K. R. Titus, S. Gu, D. S. Bassett and J. E. Phillips-Cremins

Feb, 2018

Mammalian genomes are folded in a hierarchy of compartments, topologically associating domains (TADs), subTADs and looping interactions. Here, we describe 3DNetMod, a graph theory-based method for sensitive and accurate detection of chromatin domains across length scales in Hi-C data. We identify nested, partially overlapping TADs and subTADs genome wide by optimizing network modularity and varying a single resolution parameter. 3DNetMod can be applied broadly to understand genome reconfiguration in development and disease.

<http://www.ncbi.nlm.nih.gov/pubmed/29334377>

**Alternative Start and Termination Sites of Transcription Drive Most Transcript Isoform Differences across Human Tissues**

*Nucleic acids research*

A. Reyes and W. Huber

Jan 25, 2018

Most human genes generate multiple transcript isoforms. The differential expression of these isoforms can help specify cell types. Diverse transcript isoforms arise from the use of alternative transcription start sites, polyadenylation sites and splice sites; however, the relative contribution of these processes to isoform diversity in normal human physiology is unclear. To address this question, we investigated cell type-dependent differences in exon usage of over 18 000 protein-coding genes in 23 cell types from 798 samples of the Genotype-Tissue Expression Project. We found that about half of the expressed genes displayed tissue-dependent transcript isoforms. Alternative transcription start and termination sites, rather than alternative splicing, accounted for the majority of tissue-dependent exon usage. We confirmed the widespread tissue-dependent use of alternative transcription start sites in a second, independent dataset, Cap Analysis of Gene Expression data from the FANTOM consortium. Moreover, our results indicate that most tissue-dependent splicing involves untranslated exons and therefore may not increase proteome complexity. Thus, alternative transcription start and termination sites are the principal drivers of transcript isoform diversity across tissues, and may underlie the majority of cell type specific proteomes and functions.

<http://www.ncbi.nlm.nih.gov/pubmed/29202200>

**Improving Crispr-Cas Specificity with Chemical Modifications in Single-Guide Rnas**

*Nucleic acids research*

D. E. Ryan, D. Taussig, I. Steinfeld, S. M. Phadnis, B. D. Lunstad, M. Singh, X. Vuong, K. D. Okochi, R. McCaffrey, M. Olesiak, S. Roy, C. W. Yung, B. Curry, J. R. Sampson, L. Bruhn and D. J. Dellinger

Jan 25, 2018

CRISPR systems have emerged as transformative tools for altering genomes in living cells with unprecedented ease, inspiring keen interest in increasing their specificity for perfectly matched targets. We have developed a novel approach for improving specificity by incorporating chemical modifications in guide RNAs (gRNAs) at specific sites in their DNA recognition sequence ('guide sequence') and systematically evaluating their on-target and off-target activities in biochemical DNA cleavage assays and cell-based assays. Our results show that a chemical modification (2'-O-methyl-3'-phosphonoacetate, or 'MP') incorporated at select sites in the ribose-phosphate backbone of gRNAs can dramatically reduce off-target cleavage activities while maintaining high on-target performance, as demonstrated in clinically relevant genes. These findings reveal a unique method for enhancing specificity by chemically modifying the guide sequence in gRNAs. Our approach introduces a versatile tool for augmenting the performance of CRISPR systems for research, industrial and therapeutic applications.

<http://www.ncbi.nlm.nih.gov/pubmed/29216382>

**DNA Knots Occur in Intracellular Chromatin**

*Nucleic acids research*

A. Valdes, J. Segura, S. Dyson, B. Martinez-Garcia and J. Roca

Jan 25, 2018

In vivo DNA molecules are narrowly folded within chromatin fibers and self-interacting chromatin domains. Therefore, intra-molecular DNA entanglements (knots) might occur via DNA strand passage activity of topoisomerase II. Here, we assessed the presence of such DNA knots in a variety of yeast circular minichromosomes. We found that small steady state fractions of DNA knots are common in intracellular chromatin. These knots occur irrespective of DNA replication and cell proliferation, though their abundance is reduced during DNA transcription. We found also that in vivo DNA knotting probability does not scale proportionately with chromatin length: it reaches a value of approximately 0.025 in domains of approximately 20 nucleosomes but tends to level off in longer chromatin fibers. These figures suggest that, while high flexibility of nucleosomal fibers and clustering of nearby nucleosomes facilitate DNA knotting locally, some mechanism minimizes the scaling of DNA knot formation throughout intracellular chromatin. We postulate that regulation of topoisomerase II activity and the fractal architecture of chromatin might be crucial to prevent a potentially massive and harmful self-entanglement of DNA molecules in vivo.

<http://www.ncbi.nlm.nih.gov/pubmed/29149297>

**Freepsi: An Alignment-Free Approach to Estimating Exon-Inclusion Ratios without a Reference Transcriptome - check for rna workgroup**

*Nucleic acids research*

J. Zhou, S. Ma, D. Wang, J. Zeng and T. Jiang

Jan 25, 2018

Alternative splicing plays an important role in many cellular processes of eukaryotic organisms. The exon-inclusion ratio, also known as percent spliced in, is often regarded as one of the most effective measures of alternative splicing events. The existing methods for estimating exon-inclusion ratios at the genome scale all require the existence of a reference transcriptome. In this paper, we propose an alignment-free method, FreePSI, to perform genome-wide estimation of exon-inclusion ratios from RNA-Seq data without relying on the guidance of a reference transcriptome. It uses a novel probabilistic generative model based on k-mer profiles to quantify the exon-inclusion ratios at the genome scale and an efficient expectation-maximization algorithm based on a divide-and-conquer strategy and ultrafast conjugate gradient projection descent method to solve the model. We compare FreePSI with the existing methods on simulated and real RNA-seq data in terms of both accuracy and efficiency and show that it is able to achieve very good performance even though a reference transcriptome is not provided. Our results suggest that FreePSI may have important applications in performing alternative splicing analysis for organisms that do not have quality reference transcriptomes. FreePSI is implemented in C++ and freely available to the public on GitHub.

<http://www.ncbi.nlm.nih.gov/pubmed/29136203>