**DNA Mismatch Repair Preferentially Protects Genes from Mutation**

*Genome Res*

E. J. Belfield, Z. J. Ding, F. J. C. Jamieson, A. M. Visscher, S. J. Zheng, A. Mithani and N. P. Harberd

Jan, 2018

Mutation is the source of genetic variation and fuels biological evolution. Many mutations first arise as DNA replication errors. These errors subsequently evade correction by cellular DNA repair, for example, by the well-known DNA mismatch repair (MMR) mechanism. Here, we determine the genome-wide effects of MMR on mutation. We first identify almost 9000 mutations accumulated over five generations in eight MMR-deficient mutation accumulation (MA) lines of the model plant species, Arabidopsis thaliana We then show that MMR deficiency greatly increases the frequency of both smaller-scale insertions and deletions (indels) and of single-nucleotide variant (SNV) mutations. Most indels involve A or T nucleotides and occur preferentially in homopolymeric (poly A or poly T) genomic stretches. In addition, we find that the likelihood of occurrence of indels in homopolymeric stretches is strongly related to stretch length, and that this relationship causes ultrahigh localized mutation rates in specific homopolymeric stretch regions. For SNVs, we show that MMR deficiency both increases their frequency and changes their molecular mutational spectrum, causing further enhancement of the GC to AT bias characteristic of organisms with normal MMR function. Our final genome-wide analyses show that MMR deficiency disproportionately increases the numbers of SNVs in genes, rather than in nongenic regions of the genome. This latter observation indicates that MMR preferentially protects genes from mutation and has important consequences for understanding the evolution of genomes during both natural selection and human tumor growth.

<https://www.ncbi.nlm.nih.gov/pubmed/29233924>

**Gridss: Sensitive and Specific Genomic Rearrangement Detection Using Positional De Bruijn Graph Assembly**

*Genome Res*

D. L. Cameron, J. Schroder, J. S. Penington, H. Do, R. Molania, A. Dobrovic, T. P. Speed and A. T. Papenfuss

Dec, 2017

The identification of genomic rearrangements with high sensitivity and specificity using massively parallel sequencing remains a major challenge, particularly in precision medicine and cancer research. Here, we describe a new method for detecting rearrangements, GRIDSS (Genome Rearrangement IDentification Software Suite). GRIDSS is a multithreaded structural variant (SV) caller that performs efficient genome-wide break-end assembly prior to variant calling using a novel positional de Bruijn graph-based assembler. By combining assembly, split read, and read pair evidence using a probabilistic scoring, GRIDSS achieves high sensitivity and specificity on simulated, cell line, and patient tumor data, recently winning SV subchallenge #5 of the ICGC-TCGA DREAM8.5 Somatic Mutation Calling Challenge. On human cell line data, GRIDSS halves the false discovery rate compared to other recent methods while matching or exceeding their sensitivity. GRIDSS identifies nontemplate sequence insertions, microhomologies, and large imperfect homologies, estimates a quality score for each breakpoint, stratifies calls into high or low confidence, and supports multisample analysis.

<https://www.ncbi.nlm.nih.gov/pubmed/29097403>

**Deep Learning of the Regulatory Grammar of Yeast 5' Untranslated Regions from 500,000 Random Sequences**

*Genome Res*

J. T. Cuperus, B. Groves, A. Kuchina, A. B. Rosenberg, N. Jojic, S. Fields and G. Seelig

Dec, 2017

Our ability to predict protein expression from DNA sequence alone remains poor, reflecting our limited understanding of cis-regulatory grammar and hampering the design of engineered genes for synthetic biology applications. Here, we generate a model that predicts the protein expression of the 5' untranslated region (UTR) of mRNAs in the yeast Saccharomyces cerevisiae. We constructed a library of half a million 50-nucleotide-long random 5' UTRs and assayed their activity in a massively parallel growth selection experiment. The resulting data allow us to quantify the impact on protein expression of Kozak sequence composition, upstream open reading frames (uORFs), and secondary structure. We trained a convolutional neural network (CNN) on the random library and showed that it performs well at predicting the protein expression of both a held-out set of the random 5' UTRs as well as native S. cerevisiae 5' UTRs. The model additionally was used to computationally evolve highly active 5' UTRs. We confirmed experimentally that the great majority of the evolved sequences led to higher protein expression rates than the starting sequences, demonstrating the predictive power of this model.

<https://www.ncbi.nlm.nih.gov/pubmed/29097404>

**Sidr: Simultaneous Isolation and Parallel Sequencing of Genomic DNA and Total Rna from Single Cells**

*Genome Res*

K. Y. Han, K. T. Kim, J. G. Joung, D. S. Son, Y. J. Kim, A. Jo, H. J. Jeon, H. S. Moon, C. E. Yoo, W. Chung, H. H. Eum, S. Kim, H. K. Kim, J. E. Lee, M. J. Ahn, H. O. Lee, D. Park and W. Y. Park

Jan, 2018

Simultaneous sequencing of the genome and transcriptome at the single-cell level is a powerful tool for characterizing genomic and transcriptomic variation and revealing correlative relationships. However, it remains technically challenging to analyze both the genome and transcriptome in the same cell. Here, we report a novel method for simultaneous isolation of genomic DNA and total RNA (SIDR) from single cells, achieving high recovery rates with minimal cross-contamination, as is crucial for accurate description and integration of the single-cell genome and transcriptome. For reliable and efficient separation of genomic DNA and total RNA from single cells, the method uses hypotonic lysis to preserve nuclear lamina integrity and subsequently captures the cell lysate using antibody-conjugated magnetic microbeads. Evaluating the performance of this method using real-time PCR demonstrated that it efficiently recovered genomic DNA and total RNA. Thorough data quality assessments showed that DNA and RNA simultaneously fractionated by the SIDR method were suitable for genome and transcriptome sequencing analysis at the single-cell level. The integration of single-cell genome and transcriptome sequencing by SIDR (SIDR-seq) showed that genetic alterations, such as copy-number and single-nucleotide variations, were more accurately captured by single-cell SIDR-seq compared with conventional single-cell RNA-seq, although copy-number variations positively correlated with the corresponding gene expression levels. These results suggest that SIDR-seq is potentially a powerful tool to reveal genetic heterogeneity and phenotypic information inferred from gene expression patterns at the single-cell level.

<https://www.ncbi.nlm.nih.gov/pubmed/29208629>

**Error-Gated Hebbian Rule: A Local Learning Rule for Principal and Independent Component Analysis**

*Scientific Reports*

T. Isomura and T. Toyoizumi

2018/01/30, 2018

We developed a biologically plausible unsupervised learning algorithm, error-gated Hebbian rule (EGHR)-β, that performs principal component analysis (PCA) and independent component analysis (ICA) in a single-layer feedforward neural network. If parameter β = 1, it can extract the subspace that major principal components span similarly to Oja’s subspace rule for PCA. If β = 0, it can separate independent sources similarly to Bell-Sejnowski’s ICA rule but without requiring the same number of input and output neurons. Unlike these engineering rules, the EGHR-β can be easily implemented in a biological or neuromorphic circuit because it only uses local information available at each synapse. We analytically and numerically demonstrate the reliability of the EGHR-β in extracting and separating major sources given high-dimensional input. By adjusting β, the EGHR-β can extract sources that are missed by the conventional engineering approach that first applies PCA and then ICA. Namely, the proposed rule can successfully extract hidden natural images even in the presence of dominant or non-Gaussian noise components. The results highlight the reliability and utility of the EGHR-β for large-scale parallel computation of PCA and ICA and its future implementation in a neuromorphic hardware.

<https://doi.org/10.1038/s41598-018-20082-0>

**The Landscape of Mirna Editing in Animals and Its Impact on Mirna Biogenesis and Targeting**

*Genome Res*

L. Li, Y. Song, X. Shi, J. Liu, S. Xiong, W. Chen, Q. Fu, Z. Huang, N. Gu and R. Zhang

Jan, 2018

Adenosine-to-inosine (A-to-I) RNA editing regulates miRNA biogenesis and function. To date, fewer than 160 miRNA editing sites have been identified. Here, we present a quantitative atlas of miRNA A-to-I editing through the profiling of 201 pri-miRNA samples and 4694 mature miRNA samples in human, mouse, and Drosophila. We identified 4162 sites present in approximately 80% of the pri-miRNAs and 574 sites in mature miRNAs. miRNA editing is prevalent in many tissue types in human. However, high-level editing is mostly found in neuronal tissues in mouse and Drosophila Interestingly, the edited miRNAs in neuronal and non-neuronal tissues in human gain two distinct sets of new targets, which are significantly associated with cognitive and organ developmental functions, respectively. Furthermore, we reveal that miRNA editing profoundly affects asymmetric strand selection. Altogether, these data provide insight into the impact of RNA editing on miRNA biology and suggest that miRNA editing has recently gained non-neuronal functions in human.

<https://www.ncbi.nlm.nih.gov/pubmed/29233923>

**A Novel Approach for Data Integration and Disease Subtyping**

*Genome Res*

T. Nguyen, R. Tagett, D. Diaz and S. Draghici

Dec, 2017

Advances in high-throughput technologies allow for measurements of many types of omics data, yet the meaningful integration of several different data types remains a significant challenge. Another important and difficult problem is the discovery of molecular disease subtypes characterized by relevant clinical differences, such as survival. Here we present a novel approach, called perturbation clustering for data integration and disease subtyping (PINS), which is able to address both challenges. The framework has been validated on thousands of cancer samples, using gene expression, DNA methylation, noncoding microRNA, and copy number variation data available from the Gene Expression Omnibus, the Broad Institute, The Cancer Genome Atlas (TCGA), and the European Genome-Phenome Archive. This simultaneous subtyping approach accurately identifies known cancer subtypes and novel subgroups of patients with significantly different survival profiles. The results were obtained from genome-scale molecular data without any other type of prior knowledge. The approach is sufficiently general to replace existing unsupervised clustering approaches outside the scope of bio-medical research, with the additional ability to integrate multiple types of data.

<https://www.ncbi.nlm.nih.gov/pubmed/29066617>

**Neuroprotective Drug for Nerve Trauma Revealed Using Artificial Intelligence**

*Scientific Reports*

D. Romeo-Guitart, J. Forés, M. Herrando-Grabulosa, R. Valls, T. Leiva-Rodríguez, E. Galea, F. González-Pérez, X. Navarro, V. Petegnief, A. Bosch, M. Coma, J. M. Mas and C. Casas

2018/01/30, 2018

Here we used a systems biology approach and artificial intelligence to identify a neuroprotective agent for the treatment of peripheral nerve root avulsion. Based on accumulated knowledge of the neurodegenerative and neuroprotective processes that occur in motoneurons after root avulsion, we built up protein networks and converted them into mathematical models. Unbiased proteomic data from our preclinical models were used for machine learning algorithms and for restrictions to be imposed on mathematical solutions. Solutions allowed us to identify combinations of repurposed drugs as potential neuroprotective agents and we validated them in our preclinical models. The best one, NeuroHeal, neuroprotected motoneurons, exerted anti-inflammatory properties and promoted functional locomotor recovery. NeuroHeal endorsed the activation of Sirtuin 1, which was essential for its neuroprotective effect. These results support the value of network-centric approaches for drug discovery and demonstrate the efficacy of NeuroHeal as adjuvant treatment with surgical repair for nervous system trauma.

<https://doi.org/10.1038/s41598-018-19767-3>

**Inhibition of the Checkpoint Protein Pd-1 by the Therapeutic Antibody Pembrolizumab Outlined by Quantum Chemistry**

*Scientific Reports*

A. B. M. L. A. Tavares, J. X. Lima Neto, U. L. Fulco and E. L. Albuquerque

2018/01/30, 2018

Much of the recent excitement in the cancer immunotherapy approach has been generated by the recognition that immune checkpoint proteins, like the receptor PD-1, can be blocked by antibody-based drugs with profound effects. Promising clinical data have already been released pointing to the efficiency of the drug pembrolizumab to block the PD-1 pathway, triggering the T-lymphocytes to destroy the cancer cells. Thus, a deep understanding of this drug/receptor complex is essential for the improvement of new drugs targeting the protein PD-1. In this context, by employing quantum chemistry methods based on the Density Functional Theory (DFT), we investigate in silico the binding energy features of the receptor PD-1 in complex with its drug inhibitor. Our computational results give a better understanding of the binding mechanisms, being also an efficient alternative towards the development of antibody-based drugs, pointing to new treatments for cancer therapy.

<https://doi.org/10.1038/s41598-018-20325-0>

**Sex-Biased Microrna Expression in Mammals and Birds Reveals Underlying Regulatory Mechanisms and a Role in Dosage Compensation**

*Genome Res*

M. Warnefors, K. Mossinger, J. Halbert, T. Studer, J. L. VandeBerg, I. Lindgren, A. Fallahshahroudi, P. Jensen and H. Kaessmann

Dec, 2017

Sexual dimorphism depends on sex-biased gene expression, but the contributions of microRNAs (miRNAs) have not been globally assessed. We therefore produced an extensive small RNA sequencing data set to analyze male and female miRNA expression profiles in mouse, opossum, and chicken. Our analyses uncovered numerous cases of somatic sex-biased miRNA expression, with the largest proportion found in the mouse heart and liver. Sex-biased expression is explained by miRNA-specific regulation, including sex-biased chromatin accessibility at promoters, rather than piggybacking of intronic miRNAs on sex-biased protein-coding genes. In mouse, but not opossum and chicken, sex bias is coordinated across tissues such that autosomal testis-biased miRNAs tend to be somatically male-biased, whereas autosomal ovary-biased miRNAs are female-biased, possibly due to broad hormonal control. In chicken, which has a Z/W sex chromosome system, expression output of genes on the Z Chromosome is expected to be male-biased, since there is no global dosage compensation mechanism that restores expression in ZW females after almost all genes on the W Chromosome decayed. Nevertheless, we found that the dominant liver miRNA, miR-122-5p, is Z-linked but expressed in an unbiased manner, due to the unusual retention of a W-linked copy. Another Z-linked miRNA, the male-biased miR-2954-3p, shows conserved preference for dosage-sensitive genes on the Z Chromosome, based on computational and experimental data from chicken and zebra finch, and acts to equalize male-to-female expression ratios of its targets. Unexpectedly, our findings thus establish miRNA regulation as a novel gene-specific dosage compensation mechanism.

<https://www.ncbi.nlm.nih.gov/pubmed/29079676>