Computational approaches for the transcriptomics, proteomics and epigenetic analysis in adult stem cells


Jordan

King Jordan; Email: king.jordan@biology.gatech.edu

How will the topic you presented help move this field of research forward?

I presented a series of examples illustrating how the computational approaches we are developing in my lab can be applied to the analysis of genome sequence, chromatin and gene expression dynamics. This was all done in the context of a collaborative study with the group of Dr. Victoria Lunyak on the changes in genome structure, chromatin and gene expression that accompany the aging of adult human stem cells.

How are systems biology approaches being currently applied in the field of regenerative medicine?

The paradigm of systems biology is that biological systems are integrated wholes that cannot be adequately understood through a reductionist approach alone. In other words, in addition to understanding the individual parts of a given biological system you must also appreciate how it is that the parts act together to encode the functional whole, whether that whole is a macromolecular protein complex, an entire cell or even an entire tissue or organ system. With respect to regenerative medicine, it is of course critically important to gain an understanding of the biochemical and molecular biology phenomena that underlie the processes of tissue aging and decay. However, the classic experimental approaches that yield such knowledge should also be scaled up to the entire genome and then integrated across levels of biological organization. Thus, traditional molecular biology must be complemented with a higher-level integrated systems approach in order to truly understand what goes on at the level of tissues and organ systems. Once such an integrated level of understanding has been achieved, then the probability of successfully intervening in, and possibly even reversing, the process of tissue degeneration is greatly increased. In short, a systems level understanding of how tissue function declines with aging and disease will help to facilitate medical interventions aimed at tissue regeneration.

The protein coding genes represent only about 2% of the genome. Are we computationally ready for exploring the poorly annotated genomic "dark matter" or "junk DNA," and if so, can we computationally predict its function?

Analysis and understanding of the functional relevance of the non-coding portion of the genome, particularly transposable element-derived repetitive DNA sequences, has been and continues to be one of the great open challenges in genomics research. Because it was so difficult to understand what these kinds of sequences may be doing in the genome, they were initially dismissed as "junk DNA" with little or no functional relevance. Fortunately, thanks to the work of numerous labs worldwide, it is now abundantly clear that transposable element derived repetitive sequences can and do play a variety of critical functional roles for their host genomes, i.e., they are no longer considered as merely "junk DNA" but rather as legitimate players in the regulation of the genome.

Nevertheless, these transposable element-derived sequences still pose a fundamental challenge for computational analysis based largely on the fact that they are repetitive. All of the high-throughput array-based and sequence-based functional genomics assays require specific hybridization between probe and target sequences. Probe sequences can be oligonucleotides laid down on an array or short sequence tags characterized with next-generation sequencing techniques, while target sequences are the messenger RNA or genomic DNA sequences to which these probes map. Repetitive sequences often violate this principle and are thus discarded and simply ignored in subsequent analytical steps. My lab has been developing algorithms that "rescue" multi-mapping sequence probes and, in so doing, allow the transposable element-derived repetitive fraction of the genome to be included in genome-level and systems biology analyses.

With respect to the notion of a computational prediction of function for transposable element-derived sequences, I feel the need to be cautious and circumspect in terms of the power of the computational approaches that we use. To whatever extent possible, everything we do in my bioinformatics is based on the analysis of real data rather than ab initio predictions. In other words, even though we use computers in our research, we are empiricists, not theorists, and we always let the data guide any conclusions that we may make. Of course, there is an important place for prediction in these computational research efforts, but we try to be cautious to ground our predictions in empirical data and, whenever possible, to experimentally validate any predictions that we make. Computational analysis is absolutely essential to high-throughput and systems biology approaches to regenerative medicine, but the experimental approach has been and will remain, the gold standard for generating knowledge in biological research.