Barash and Wang *Genome Biology* 2010, **11**:307
http://genomebiology.com/2010/11/8/307

Genome **Biology**

MEETING REPORT

An illuminated view of molecular biology

Yoseph Barash[1,2]* and Xinchen Wang[2]

Abstract

A report on the 18th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) and the 7th Special Interest Group meeting on Alternative Splicing, Boston, USA, 9-13 July 2010.

Everything is Illuminated, Liev Schreiber's 2005 directorial debut, is a charming offbeat movie about a young man on a self-driven quest in a foreign land. Through his journey, he finds connections between past and present and between things that initially appear completely unrelated. Both the title and the unifying theme of the movie match current trends and advances in molecular biology well. Like the character in the film, researchers meticulously collect information using newly developed high-throughput and high-resolution technologies. And as they work to make sense of these new findings, unifying themes begin to emerge. At the recent International Conference on Intelligent Systems for Molecular Biology and the associated Special Interest Group meeting on Alternative Splicing in Boston, attendees were treated to talks covering diverse topics that, when taken together, offered a glimpse of the interconnectedness of traditionally separate fields in molecular biology and the ever-improving tools available to study them. Here we present a few of the highlights of the meetings.

A unified view of the transcriptome

At the level of the transcriptome, evidence is accumulating about the physical and regulatory coupling between RNA splicing and nucleosome positioning, histone modifications and non-coding RNAs. Nature does not seem to follow the human-contrived definitions that separate biological processes or research fields, and building a unified model of the transcriptome requires combining information from different fields. Addressing the question of the control of alternative RNA splicing, Reini Luco (National Cancer Institute, National Institutes of Health, Bethesda, USA) described how the trimethylation of lysine 36 on histone H3 (H3K36me3) can change the pattern of alternative exon inclusion. By studying alternative exons within the human fibroblast growth factor receptor 2 (FGFR2) and several other human genes, Luco and colleagues found that modulation of the expression levels of the H3K36 methyltransferase SETD2 and the H3K36me3-binding protein MRG15 affect the splicing patterns of a set of exons regulated by the protein PTB, an RNA-binding splicing factor. They then used high-throughput RNA sequencing to measure splicing changes in cells depleted of SETD2, MRG15 or PTB, and found significant overlap between differentially spliced exons in all three groups. Luco showed that the splicing changes correlated with the strength of the nearby PTB-binding sites. Investigating the role that other modifications may play in splicing, Luco presented preliminary chromatin immunoprecipitation and sequencing (ChIP-seq) and RNA sequencing (RNA-seq) data on the genome-wide association between dozens of different histone marks and alternatively spliced exons. Christian Muchardt (CNRS Institut Pasteur, Paris, France) described findings suggesting increased H3K9me3 marks near alternative exons in the human CD44 gene, and showed that the chromodomain-containing protein HP1γ, which binds H3K9me3 marks, can regulate alternative exon inclusion.

Several speakers at the AS-SIG also presented data showing connections between transcriptional gene silencing by small RNAs and the regulation of RNA splicing. In particular, Mariano Alló (Universidad de Buenos Aires, Argentina) described work on the ability of small RNAs to affect alternative splicing patterns. He showed that small interfering RNAs targeted to regions near an alternative exon in the human fibronectin 1 gene can affect exon inclusion in an Argonaute 1 and 2-dependent manner, and that the effect also depended on the presence of nearby histone marks. A genome-wide scan yielded a set of splicing events with similar adjacent features that could also be regulated by this mechanism, suggesting that it may be used more generally in the cell.

John Rinn (Broad Institute, Cambridge, USA) presented his group's investigations on the role of large

*Correspondence: yoseph@psi.utoronto.ca
[1]Biomedical Engineering, Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto M5S 3G4, Canada
[2]Banting and Best Department of Medical Research and Department of Molecular Genetics, Donnelly Centre, University of Toronto, 160 College Street, Toronto M5S 3E1, Canada

BioMed Central

intergenic non-coding RNAs (lincRNAs). A genome-wide screen in mouse identifying potentially non-coding loci enriched with transcriptionally active histone modifications led to the discovery of several new lincRNAs. Using a guilt-by-association technique to ascribe potential functions, Rinn and colleagues focused on a lincRNA upstream of the p21 gene that was induced by p53, was necessary for apoptosis, and could act as a global repressor of gene expression. In regard to the regulatory mechanism, this lincRNA was found to be bound to a highly conserved region in a specific heterogeneous nuclear ribonucleoprotein.

In a highlight talk, Christopher Burge (Massachusetts Institute of Technology, Cambridge, USA) set a target for the alternative splicing field to further develop computational tools that can combine data from high-resolution, high-throughput technologies to build predictive models for the transcriptome, and he described new technological advances that will facilitate these predictive models. Earlier, at the AS-SIG, YB presented work by the authors of this report on the tissue-regulated splicing code, which was mentioned by Burge during his talk. He focused on a new technology currently being developed in his lab that uses Illumina high-throughput sequencing technology to quantitatively assess the DNA-binding affinity of transcription factors. Burge discussed current efforts to extend this approach for RNA-binding proteins and illustrated its usefulness by describing insights gained into the binding affinity of the yeast transcription factor GCN4.

Making sense of high-throughput data

As experiments using the new high-throughput technologies generate data at an increasingly rapid rate, computational biologists are developing algorithms that can accurately extract information from these data. Specifically, the analysis of sequencing data still poses many challenges. Xiaoyu Chen (University of Washington, Seattle, USA) described a dynamic Bayesian network model aimed at identifying transcription-factor-binding sites in yeast DNA from genomic footprinting experiments involving DNase I digestion of chromatin followed by high-throughput sequencing. Jared Simpson (Wellcome Trust Sanger Institute, Hinxton, UK) described work to construct a tool that can assemble a genome from short reads and can handle genomes up to several gigabases in size, such as that of human. Unlike current commonly used algorithms that utilize de Bruijn graphs, the new method efficiently constructs a string graph from a set of reads, an approach particularly well suited for the assembly of mixed-length read data. Simpson described hurdles still to be overcome in constructing a genome assembler, such as the need to accommodate for possible sequencing errors in the short reads.

Many high-throughput technologies produce data that are difficult to interpret directly with the human eye. Consequently, another area that received much attention was the visualization of biological datasets. Gary Bader (University of Toronto, Ontario, Canada) presented new plug-ins for the Cytoscape software commonly used for visualizing molecular interaction networks that will help perform enrichment analysis of sets of nodes in large networks. Lincoln Stein (Ontario Institute for Cancer Research, Toronto, Canada) presented new features of Reactome, a manually curated database for biological pathways, with a web interface that lets users perform queries and searches. He also described a Reactome standalone plug-in for Cytoscape that enables users to integrate and visualize Reactome data with data from other sources.

There's plenty of room at the bottom

The study of molecular biology at increasingly fine resolution was another popular theme. Jonathan Widom (Northwestern University, Evanston, USA) described the confirmation of previous studies of nucleosome-binding preferences and locations in yeast, but with new data at single-nucleotide resolution. Eran Segal (Weizmann Institute, Rehovot, Israel) described an experimental set-up to investigate, at much higher resolution than before, the regulatory effect of poly(A) tracts near transcription factor binding sites in yeast promoter regions. Revisiting a model suggested by Iyer and Struhl in 1995, in which poly(A) tracts serve to deplete promoters of repressive nucleosomes, Segal and his colleagues can now quantify the effect on gene expression of parameters such as the poly(A) tract length, the proportion of adenine nucleotides, and the distance from the transcription factor binding sites. They showed that there is a trade-off between the length of the poly(A) tract and the proportion of adenine nucleotides in it, and that the influence of the tract falls steadily with distance from the transcription factor binding site, until a distance of 200 bases, where a local peak is reached. The location of the peak determined experimentally in this way does not, however, match well with current computational models. An interesting question still to be answered is whether poly(A) tract boundaries serve to fine-tune the expression levels of adjacent genes.

Jean Beggs (University of Edinburgh, UK) described the use of RiboSys reporters to enable her group to monitor cell number, DNA yield, RNA yield, reverse transcription efficiency, quantitative PCR efficiency and RNA copy number per cell (using single-molecule FISH) in yeast at 30-second intervals. Using this system, she and her colleagues identified a 30-second pause in the progress of the RNA polymerase II (Pol II) complex when splicing is occurring in yeast RNAs. They have also identified what appears to be a novel surveillance mechanism, in which phosphorylation of the Pol II complex at Ser5 is used to first pause it and then at Ser2 to resume its progress, and involving several proteins such as Ioc2 and Ioc4 (members of the Isw1 chromatin-remodeling complex). As Beggs pointed out, one could only observe this mechanism with a high-resolution system such as theirs, and only when the yeast cell population was well synchronized.

Computational approaches for studying disease

Opening the ISMB main conference, Steven Brenner (University of California, Berkeley, USA), told an inspiring story about a lab member whose successful treatment for lung cancer was a result of a better molecular understanding of a specific cancer mutation she carried. The interest in using computational biology to better understand the etiology and pathology of cancer and other diseases was exemplified by the fact that more than half of the timeslots at this year's conference featured at least one talk on studying human disease with computational methods. In a prize-winning talk, Peter Van Loo (University of Leuven, Belgium) presented an algorithm called ASCAT (Allele-Specific Copy number Analysis of Tumors), which takes signal intensities from whole-genome single-nucleotide polymorphism array data to estimate the proportion of tumor cells and tumor ploidy in a heterogeneous sample. On a genome-wide scale, ASCAT can then identify cancer-associated copy-number changes for individual loci.

In relation to pharmacogenomics, Rachel Karchin (Johns Hopkins University, Baltimore, USA) described computational approaches for distinguishing driver mutations, which can identify genes that may be promising drug targets, from passenger mutations in the genetic landscape of a tumor. She discussed a tool called CHASM (Cancer-specific High-throughput Annotation of Somatic Mutations) developed in her lab, which uses a machine-learning approach to classify missense mutations into these two categories on the basis of properties of the mutated amino acids that may affect the translated protein. David Altshuler (Harvard Medical School, Cambridge, USA) supplied a word of caution during his keynote talk about the value of disease prediction. Giving the example of the detection of prostate cancer using prostate-specific antigen, Altshuler emphasized that being able to predict outcomes does not automatically translate to being able to help patients. In this case, the corresponding treatment did not increase the patients' life expectancy, but did saddle many individuals with debilitating side effects.

The themes reviewed here reflect how the field of computational biology is constantly shifting, adapting to the new high-resolution and high-throughput technologies that are being developed, as well as to the new scientific questions these technologies allow us to ask. Data processing and visualization can therefore be expected as central themes at ISMB conferences in the future, but we should also expect to see work exploring the unified view of the transcriptome and its complexity, with connections to focused disease studies.

Acknowledgements
We thank Ben Blencowe and John Calarco for providing helpful feedback.

Published: 26 August 2010

doi:10.1186/gb-2010-11-8-307
Cite this article as: Barash Y, Wang X: An illuminated view of molecular biology. *Genome Biology* 2010, **11**:307.