Report on the CYRENE Project: A cis-Lexicon containing the regulatory architecture of 586 regulatory genes experimentally validated using the “Davidson Criteria”

Ryan Tarpine, James Hart, Timothy Johnstone, Derek Aguiar, Sorin Istrail

Center for Computational Molecular Biology, Brown University

All correspondence including getting the cisGRN Browser to Istrail Lab, Center for Computational Molecular Biology and Department of Computer Science, Brown University, sorin_istrail@brown.edu

The CYRENE cis-Lexicon presently contains the regulatory architecture of 393 transcription-factor-encoding genes and 194 other regulatory genes in species: human, mouse, fruit fly, sea urchin, nematode, rat, chicken, and zebrafish, with a higher priority on the first five species. The regulatory architectures of each of these CYRENE genes are validated using the "Davidson Criteria:" sites must be shown to physically bind proteins and functionally confirmed by in-vivo disruption. The cis-Lexicon annotations include confirmed transcription factor binding sites, the cis Regulatory Module (CRM) boundaries, the spatial and temporal functionality of the CRM, and the molecular function and classification of the encoded protein. Included is an update on the CLOSE System (cis-Lexicon Ontology Search Engine) -- a set of algorithmic strategies for automated literature extraction of cis-regulation articles -- that is used to speed up the identification of new CYRENE genes in the literature and to estimate the "completeness" of the CYRENE transcription factor universe. Here also we discuss the newly released CYRENE cisGRN-Browser, a full genome browser dedicated to cis-regulatory genomics. This work has been done jointly with Eric Davidson of Division of Biology at California Institute of Technology.

cis-Lexicon Connectivity Map (D. Melanogaster)

cis-Lexicon

Distribution of cis-Lexicon transcription factor binding sites by TF superfamily.

Distribution of cis-Lexicon transcription factor binding sites by species.

Virtual Sea Urchin

The Virtual Sea Urchin (VSU) uses spatial models and a graphics engine to simulate the 4-dimensional sea urchin embryo, allowing the researcher to probe the GRN at various levels of granularity -- from the multicellular embryo to the gene-regulatory network of an individual cell-type. The VSU currently provides models for the S. purpuratus embryo at 6h (shown), 10h, 15h, 20h, and 24h, which were created by extrapolating cross sectional color-coded tracings from photomicrographs to three dimensions (Eric H. Davidson. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. Academic Press, May 2006).

Virtual Sea Urchin's view of the Strongylocentrotus purpuratus embryo at 0, 1, 2, 3, and at 6 hours. VSU distinguishes cell type by color.

Future Direction: Cross-Platform Integration

The computational and data model for the VSU was recently completely rebuilt in Java using JGGL bindings to accommodate animation and integration with the cis-Browser. The development of an embryo can now be modeled using flat text files. The computational modeling of embryonic development will eventually feature realistic cell models and dynamics simulators. We also plan to combine the cis-regulatory sequence analysis capabilities of Cyrene and the network building, visualization, and simulation capabilities of BioTapestry with the temporal and spatial analysis of the 4D Virtual Sea Urchin to get a complete characterization of the S. purpuratus GRN.