“Comprehensive Characterization of Functional RNA Elements Encoded in the Human Genome”

Brenton R. Graveley

Department of Genetics and Genome Sciences

University of Connecticut Health Center

400 Farmington Avenue

Farmington, CT 06030-3301

Genomes contain all of the information necessary to specify the development and function of an organism. In addition to genes, genomes also contain a myriad of functional elements that control regulatory processes involved in gene expression. One class of these elements only operate when they are transcribed into RNA and serve as the binding sites for RNA binding proteins (RBPs) which act to control post-transcriptional processes including splicing, cleavage and polyadenylation, RNA editing, RNA localization, translation, and RNA stability. Despite the importance of these functional RNA elements encoded in the genome, they have been largely understudied in comparison to genes and DNA elements. We are working to develop a foundational, functional map of protein-RNA interactions of RNA binding proteins (RBPs) encoded in the human genome, and the RNA elements they bind to across the transcriptome. These RNA elements, when expressed, form the basis of co- and post-transcriptional regulation of human genes. Our strategy consists of developing and integrating a physical map of hundreds of RBPs in two different human cell lines (HepG2 and K562) with transcriptome-wide measurements of the effects of depleting these RBPs, their localization patterns and their binding preferences independent of co-factor associations. To date, we have generated 1,076 replicated datasets for 361 RBPs. Specifically, we have performed eCLIP for 97 and 84 RBPs in K562 and HepG2 cells, respectively, for a total of 126 RBPs of which 55 RBPs were characterized in both cell types. We have also used shRNA- and/or CRISPR-mediated depletion followed by RNA-seq of 225 RBPs in K562 and 227 RBPs in HepG2 cells, covering a total of 251 RBPs with 200 RBPs characterized in both cell types. We also performed RBNS on 78 RBPs, 20 of which have eCLIP-seq data in at least one cell type. We have further determined the subcellular distribution properties of 270 RBPs via systematic immunofluorescence experiments in HepG2 cells. Finally, we identified the DNA elements associated with 56 RBPs in HepG2 and 40 RBPs in K562 by ChIP-seq, covering 61 unique RBPs of which 35 RBPs were characterized in both cell types. We will present an overview of this dataset and examples of insights that have been obtained. All data is publicly available at <http://www.encodeproject.org> as soon as it is generated and verified and can be used without restriction.