



University
of Basel

Department
Biozentrum



Swiss Institute of
Bioinformatics

BIOZENTRUM

The Center for
Molecular Life Sciences

Basel Computational Biology Seminar:
22830-01 Current Research in Bioinformatics I

Roderic Guigó

*Center for Genomic Regulation, Universitat Pompeu Fabra
Barcelona, Spain*

The GENCODE CLS project: massively expanding the lncRNA catalog through capture long-read RNA sequencing

Accurate and complete gene annotations are indispensable for understanding how genome sequences encode biological functions. For more than twenty years, the GENCODE consortium has developed reference annotations for the human and mouse genomes, becoming a foundation for biomedical and genomics communities worldwide. Nevertheless, collections of important yet poorly-understood gene classes like long non-coding RNAs (lncRNAs) remain incomplete and scattered across multiple, uncoordinated catalogs. To address this, GENCODE has undertaken the most comprehensive lncRNA annotation effort to date. This is founded on the manually supervised computational annotation of full-length targeted long-read sequencing, on matched embryonic and adult tissues, of orthologous regions in human and mouse. Altogether 17,931 human genes (140,268 transcripts) and 22,784 mouse genes (136,169 transcripts) have been added to the GENCODE catalog representing a 2-fold and 6-fold growth in transcripts, respectively - the greatest increase in the number of annotated human genes since the sequencing of the human genome. Our targeted design assigned human-mouse orthologs at a rate beyond previous studies, tripling the number of human disease-associated lncRNAs that have mouse orthologs. Novel lncRNA genes consistently exhibit biological signals of functionality, and they greatly enhance the functional interpretability of the human genome. While poorly expressed in bulk RNA-Seq samples, many of them are highly expressed in specific cell populations, maybe even contributing to cell-type determination. The expanded GENCODE lncRNA annotations mark a critical step toward deciphering the human and mouse genomes.

Date: Monday, October 20, 2025
Time: 16:15 h – 17:30h
Location: Biozentrum, 02.073
Contact: Aleksei Mironov (aleksei.mironov@unibas.ch)