



Reference Epigenomes at the ENCODE portal

Jason A. Hilton¹, Cricket A. Sloan¹, Benjamin C. Hitz¹, Esther T. Chan¹, Jean M. Davidson¹, Idan Gabdank¹, J. Seth Strattan¹, Aditi Narayanan¹, Kathrina Onate¹, Marcus Ho¹, J. Michael Cherry¹

1) Department of Genetics, School of Medicine, Stanford University, Stanford, CA 94305

Abstract

The Encyclopedia of DNA elements (ENCODE) project, currently in its 10th year of production scale, is a collaborative effort toward cataloging genomic annotations. The research institutes within the consortium have conducted more than 5,000 experiments using a variety of techniques to study the structure, regulation, and transcription profiles of human and mouse genomes. The data from these experiments first passes through the ENCODE Data Coordination Center (DCC) for basic validation and metadata standardization before they are openly available at the ENCODE site (<https://www.encodeproject.org/>). From these data, we have gathered experiments conducted on a similar biosample type into datasets to be analyzed as reference epigenomes. Currently, ENCODE has assembled more than 20 reference epigenomes from human cell lines and tissues, and additional reference epigenomes from experiments conducted on mouse tissue samples along a development series. The ENCODE project has also imported data from all of the 111 reference epigenomes that were completed through the Roadmap Epigenomics Project, and we are currently analyzing them using the ENCODE analysis pipelines. Upon assemblage of a reference epigenome on the ENCODE site, an automated check compares the individual experiments within each reference epigenome to determine if the reference epigenome is from a single donor (or a single strain, in the case of mouse experiments), if any biosample treatments differ than the others in the reference epigenome, or if any IHEC-required assays are absent from the reference epigenome. Notifications are then displayed on the reference epigenome page to fully inform the user on the presented data. Additionally, ENCODE is registering each of the reference epigenomes, including the Roadmap datasets, with EpiRR. The collection, careful curation, and organization of these datasets and the integration of experimental metadata that comprise them will maximize accessibility to epigenomic data and analysis.