A Brief Tour Of ENCODE

Mike Pazin
National Human Genome Research Institute, NIH
ASHG Advanced Workshop On Integrative Analysis Using ENCODE And Roadmap Epigenomics Data
October 10, 2015
Welcome

• Objectives
  – We want to tell the community about the ENCODE resource
  – We want to hear community experiences and suggestions
ENCODE:
Encyclopedia Of DNA Elements

• Identify all candidate functional elements in the genome

• Make resource freely available to community for use in studies of:
  – genetic basis of disease
  – gene regulation
ENCODE Data Types

Science 306:636, 2004
Non-coding DNA Is Important For Disease

- About 90% of GWAS findings lie outside of protein-coding regions
- About 80% of heritability for common diseases found in regulatory regions
- Non-coding DNA variants are known to cause human diseases and alter human traits
  - Fragile X Syndrome
  - ALS (amyotrophic lateral sclerosis)

Stamatoyannopoulos, Science 337:1190, 2012
PMID: 17477822
PMID: 25679767
Richer Maps Provide More Information
Richer Maps Provide More Information
Richer Maps Provide More Information
Richer Maps Provide More Information
Standard ENCODE Use Cases: Hypothesis Generation

- Prediction of causal variants/regulatory elements
- Prediction of target genes
- Prediction of target cell types
- Prediction of upstream regulatory factors
- Prediction of mechanism for phenotype changes

- Genetic v. epigenetic
- Germline v. somatic
Publications Using ENCODE Data

Hundreds of Consortium publications

Over 1000 community publications using ENCODE data:

~440 Human Disease
~600 Basic Biology
~200 Methods/Software Development

https://www.encodeproject.org
Accessing ENCODE

• 1000s of shared datasets
  – No embargo
  – High quality
  – Uniformly processed

• Sharing software

• Data interoperability

• Informed consent for unrestricted-access sharing of genomic data
International Human Epigenome Consortium (IHEC)

- Data Portal: [http://epigenomesportal.ca/ihec/](http://epigenomesportal.ca/ihec/)
- Goal: Coordinate production of 1000 human epigenome maps for cellular states relevant to health and disease [http://ihec-epigenomes.org](http://ihec-epigenomes.org)
- Can view by consortium, by assay, by cell type
- Data from 7 consortia
Summary- Accessing ENCODE Resources

• ENCODE portal [https://www.encodeproject.org](https://www.encodeproject.org)
  – Display/download ENCODE and Roadmap Epigenomics data
  – Data Standards
  – Software tools
  – Publications
  – Encyclopedia prototype

• ENCODE Analysis Tools
  – Regulatory Elements Database [http://dnase.genome.duke.edu](http://dnase.genome.duke.edu)

• ENCODE Tutorials
  – [https://www.encodeproject.org/tutorials/](https://www.encodeproject.org/tutorials/)

• ENCODE mailing list :

• IHEC resources