



National Human Genome
Research Institute

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



Elise Feingold



Dan Gilchrist



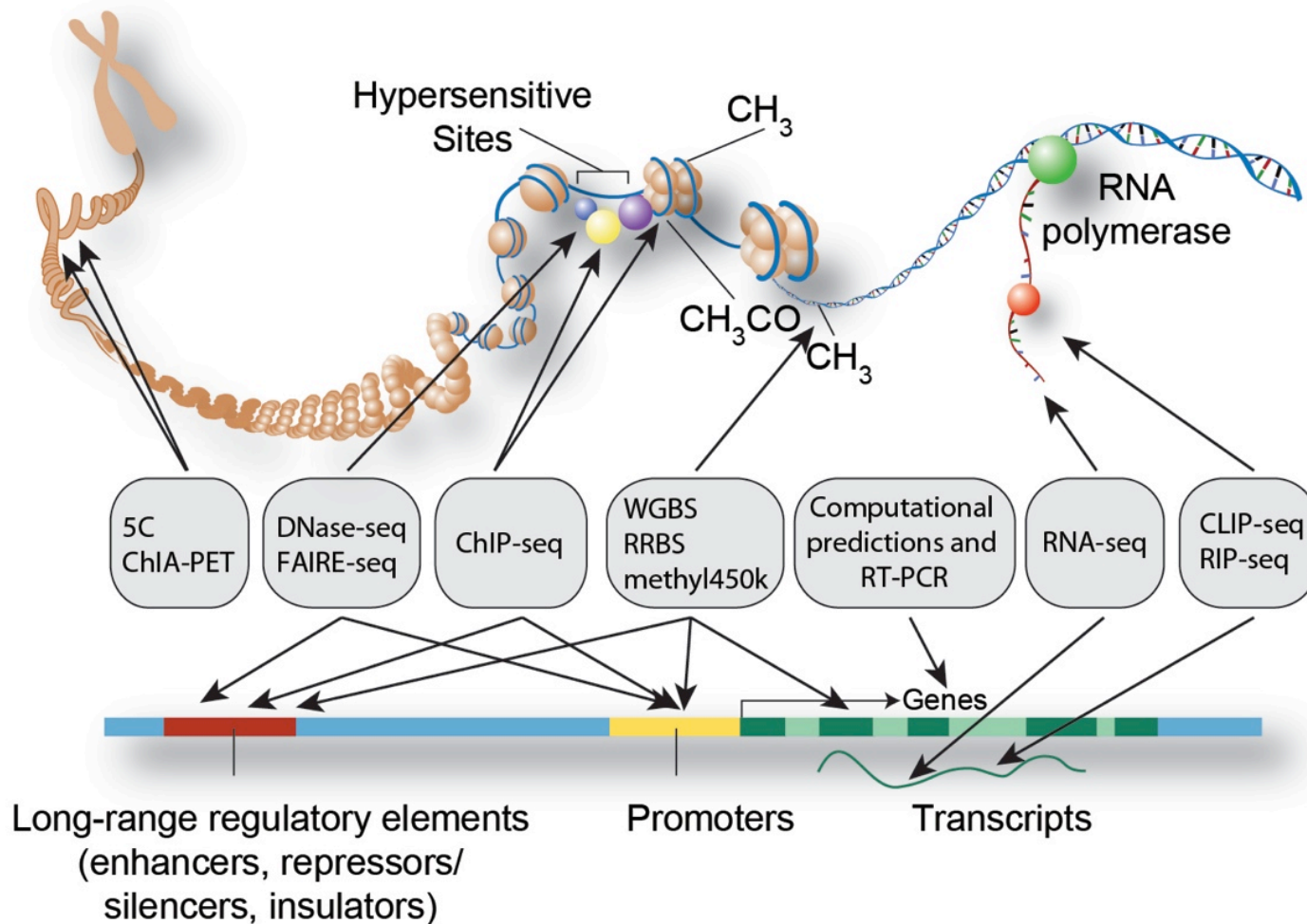
ENCODE:

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





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

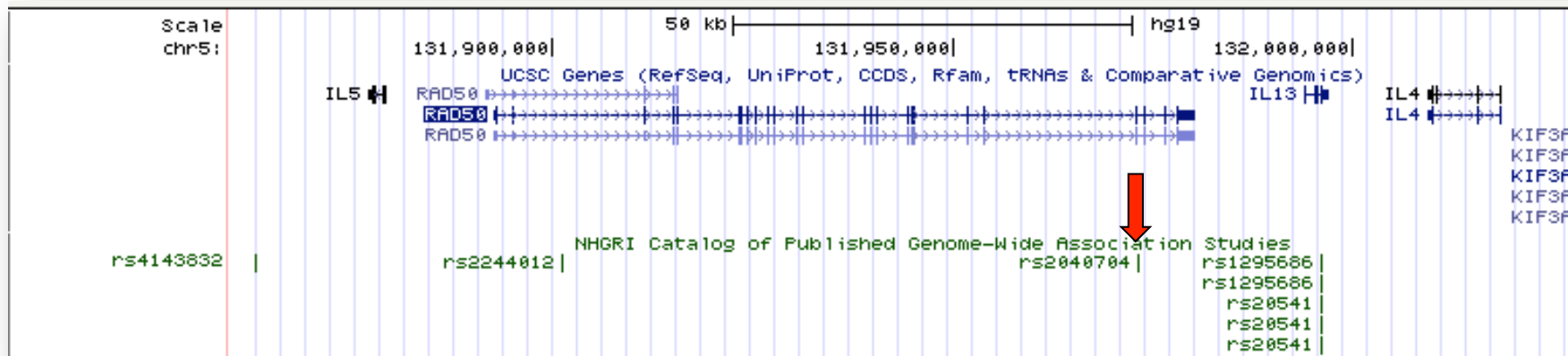
Gusev, Am J Hum Gen 95:535, 2014

PMID: 17477822

PMID: 25679767

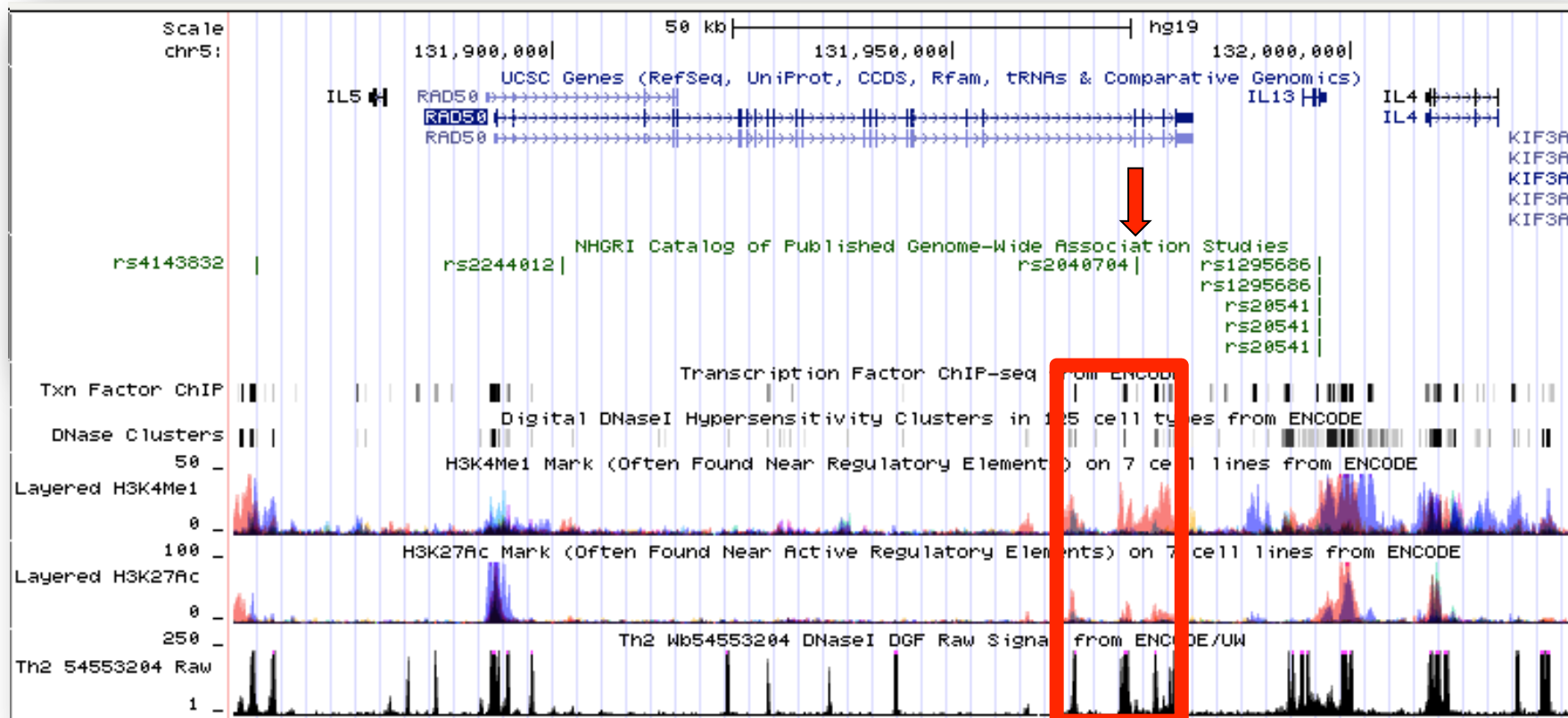


Richer Maps Provide More Information



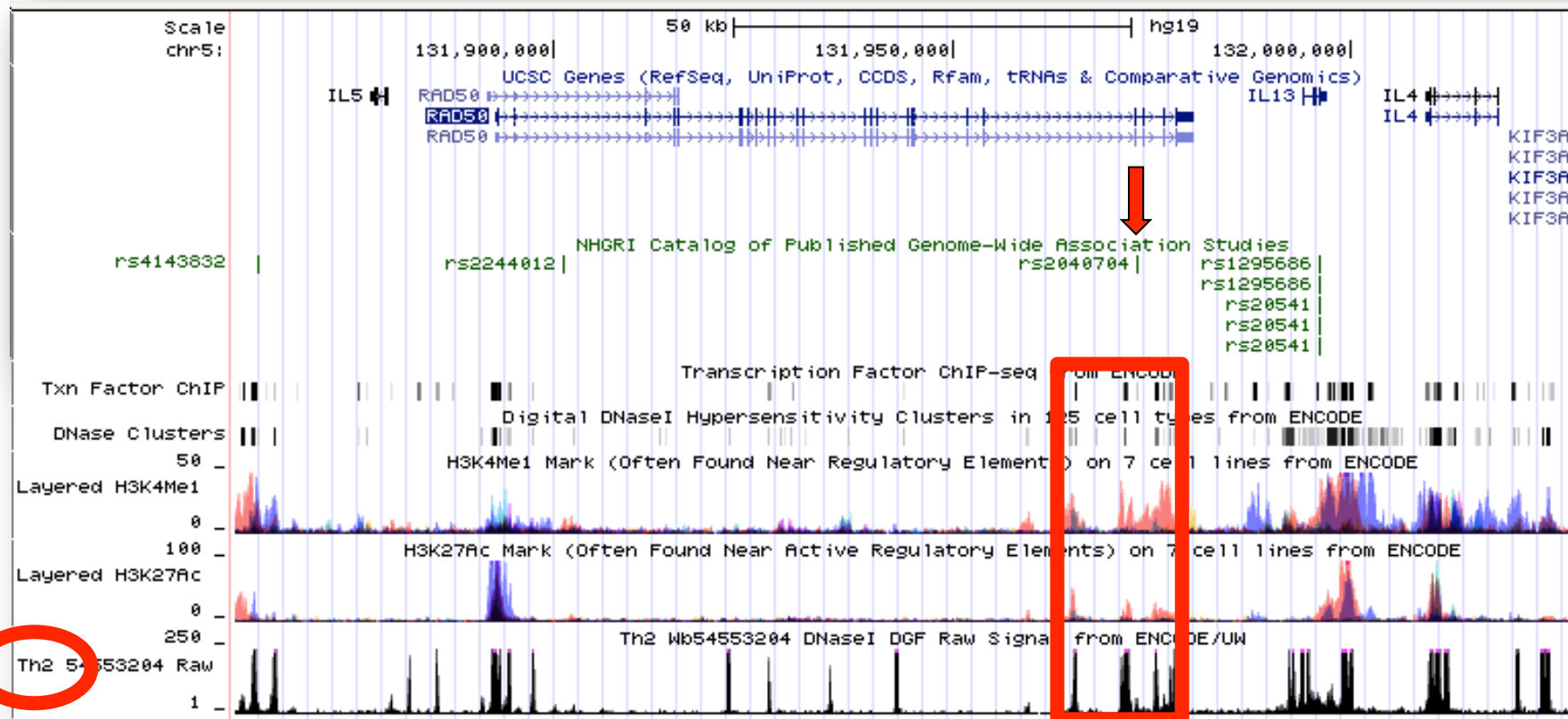


Richer Maps Provide More Information





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

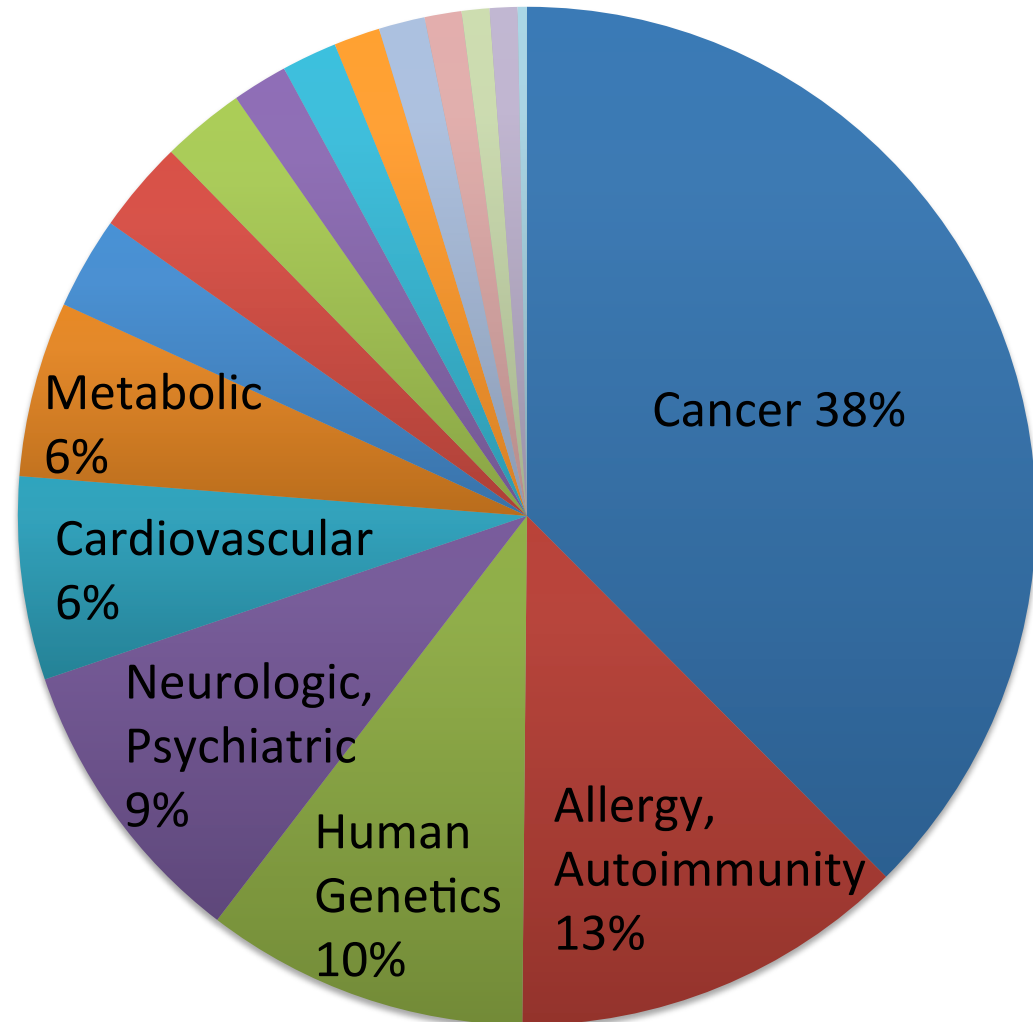


Publications Using ENCODE Data

Hundreds of Consortium publications

Over 1000 community publications using ENCODE data:

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





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/>
- Goal: Coordinate production of 1000 human epigenome maps for cellular states relevant to health and disease <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>
 - Display/download ENCODE and Roadmap Epigenomics data
 - Data Standards
 - Software tools
 - Publications
 - Encyclopedia prototype
- ENCODE Analysis Tools
 - RegulomeDB <http://regulomedb.org/>
 - HaploReg <http://www.broadinstitute.org/mammals/haploreg/>
 - Regulatory Elements Database <http://dnase.genome.duke.edu>
 - RegulomeDB GWAS Database <http://www.regulomedb.org/GWAS/>
- ENCODE Tutorials
 - <http://www.genome.gov/27553900>
 - <https://www.encodeproject.org/tutorials/>
<http://www.ncbi.nlm.nih.gov/pubmed/25762420>
- ENCODE mailing list :
 - <https://mailman.stanford.edu/mailman/listinfo/encode-announce>
- IHEC resources
 - IHEC Home Page <http://ihec-epigenomes.org>
 - IHEC Data Portal <http://epigenomesportal.ca/ihec/>