# ENCODE Overview

**Mike Pazin**

Program Director, Functional Genomics, Division of Genome Sciences

19 March 2019

National Human Genome Research Institute

The **Forefront** of **Genomics**®

# Welcome

- Objectives
  - We want to tell you about the ENCODE resource
  - We want to hear ideas, experiences, suggestions, and questions

# Non-coding DNA Is Important For Disease And Gene Regulation

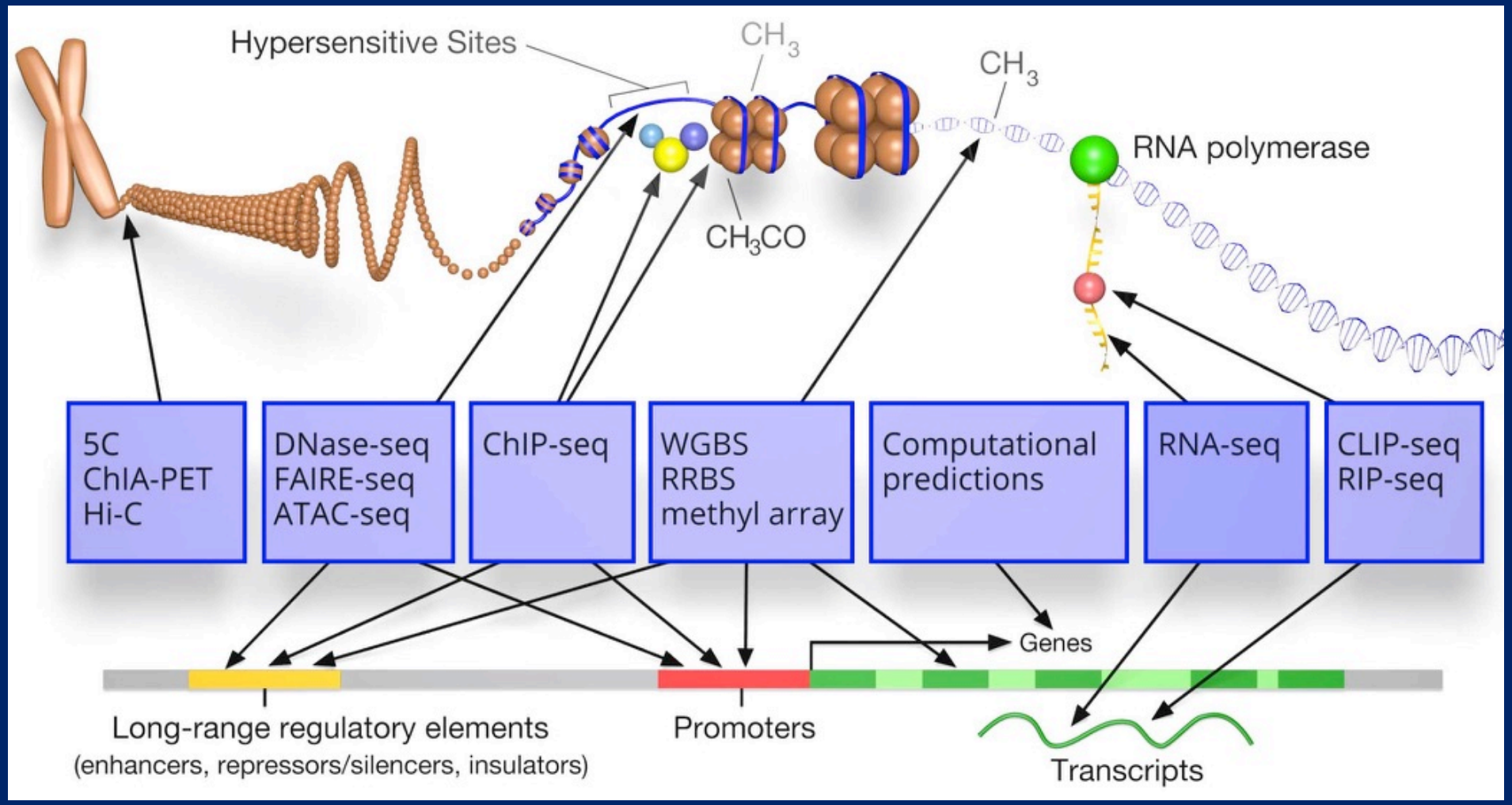
- Vast majority of common disease associations and heritability appear to lie outside of protein-coding regions
- Non-coding DNA variants are known to cause human diseases and alter human traits (FXS, ALS, polydactyly)
- BUT: Reading the genome is difficult

PMID: 22955828, PMID: 25439723, PMID: 23128226
PMID: 17477822, PMID: 25679767, PMID: 27569544

# ENCODE Goals

- Catalog all candidate functional elements in the genome
- Make resource freely available to community

# ENCODE Assays




Biochemical Features

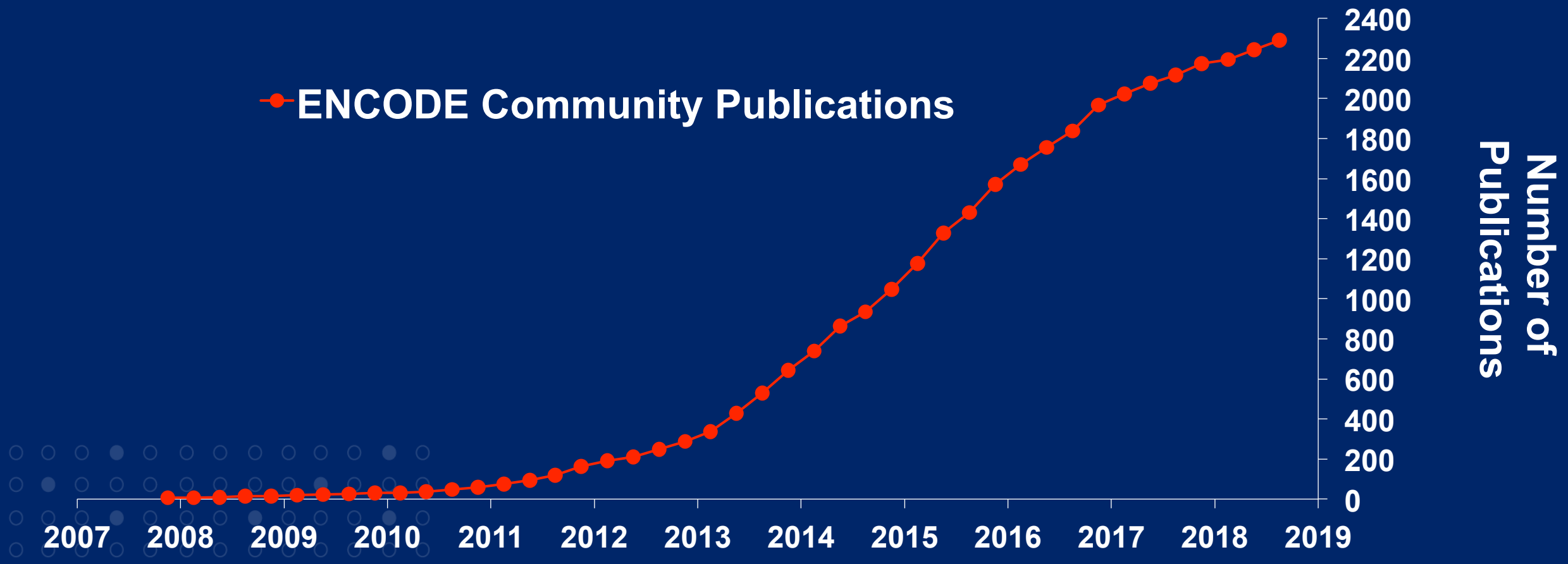
Assays

Genomic elements

# ENCODE Components

- Functional Element Mapping Centers
- Functional Element Characterization Centers (NEW)
- Computational Analysis Research Projects
- ENCODE Data Coordination Center (DCC)
- ENCODE Data Analysis Center (DAC)

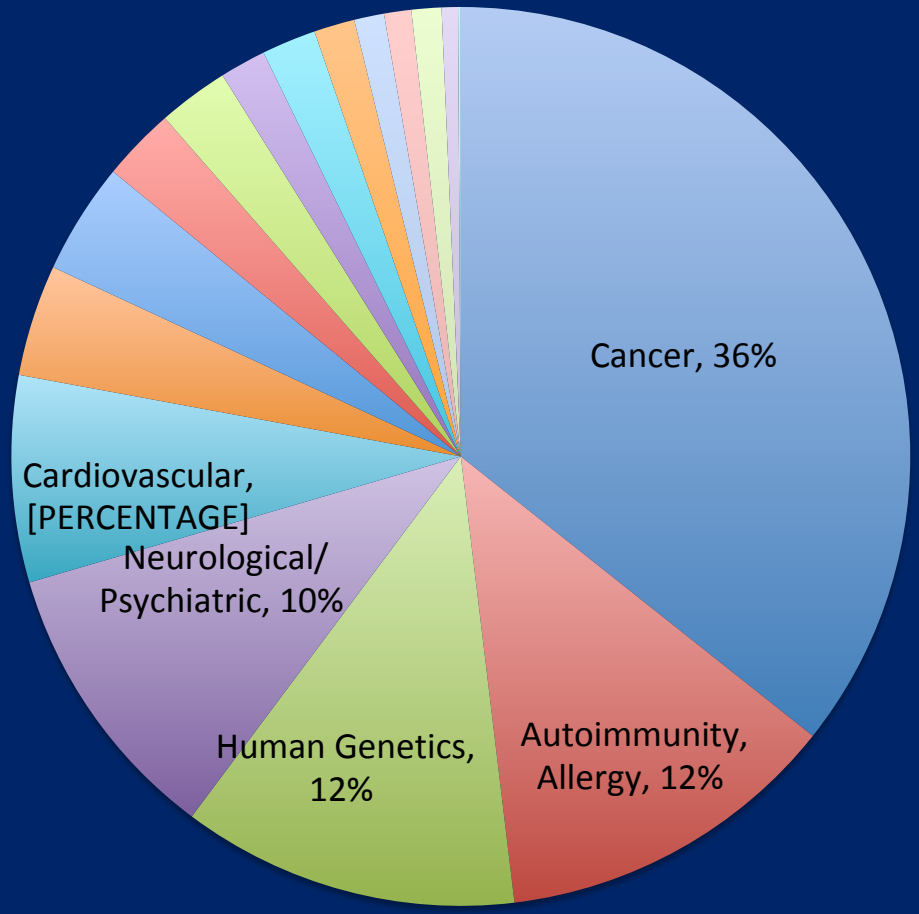
# Publications Using ENCODE Data




https://www.encodeproject.org/publications/

NIH
NHGRI

# ENCODE Community Publications, By Disease Category



# ENCODE Use Cases

- Protocols and Approaches
- Methods development/validation
- Disease association/gene regulation studies
  - Affected cell type
  - Upstream regulators
  - Target gene
  - Causal variant

# ENCODE Outreach

## ENCODE Tutorials

These tutorials were prepared for biologists using human and/or mouse gene
biology. They explain what data are available, what they mean, how they can
in genetic research on human disease.

- **ENCODE: Getting Started** New
- **2016 Society of Toxicology Annual Meeting Workshop**
- **ENCODE/Roadmap Epigenomics Tutorial, October 2015, ASHG**
- **ENCODE 2015: Research Applications and Users Meeting**
- **ENCODE Workshop, April 2015, Keystone Symposia**

## Workshop materials

- **NEW:** Uniform Processing Pipelines/DNAnexus Tutorial **from the ASHG 2016 Workshop in Vancouver, BC O**
- Presentations, tutorials, and video **from the ENCODE Users Meeting 2016 in Palo Alto, CA June 8 - 10, 2016**
  - Video and workshop materials from hands-on tutorial sessions on accessing, processing, analyzing, and ut resources, along with presentations from leading experts in disease, biology, and computational fields expla ENCODE resources in their work.
  - Pictures by Forrest Tanaka can be viewed using #EUM16 on twitter
- Keystone Symposium 2016, Chromatin and Epigenomics, **Whistler, British Columbia, March 21, 2016.**
  - J. Seth Strattan, Find and Visualize ENCODE Data
- Asia Pacific Bioinformatics Conference, **San Francisco, January 10, 2016.**
  - J. Michael Cherry, Introduction to ENCODE and the ENCODE DCC
  - Aditi Narayanan, The ENCODE Portal: Searching for Metadata and Data
  - J. Seth Strattan, ENCODE Data Availability and Standardized Processing

https://www.genome.gov/27553900/encode-tutorials/ https://www.encodeproject.org/tutorials/

NIH NHGRI

# ENCODE Users Meeting

- Save the date:
- July 8-11 2019
- Seattle, WA

# International Human Epigenome Consortium (IHEC)

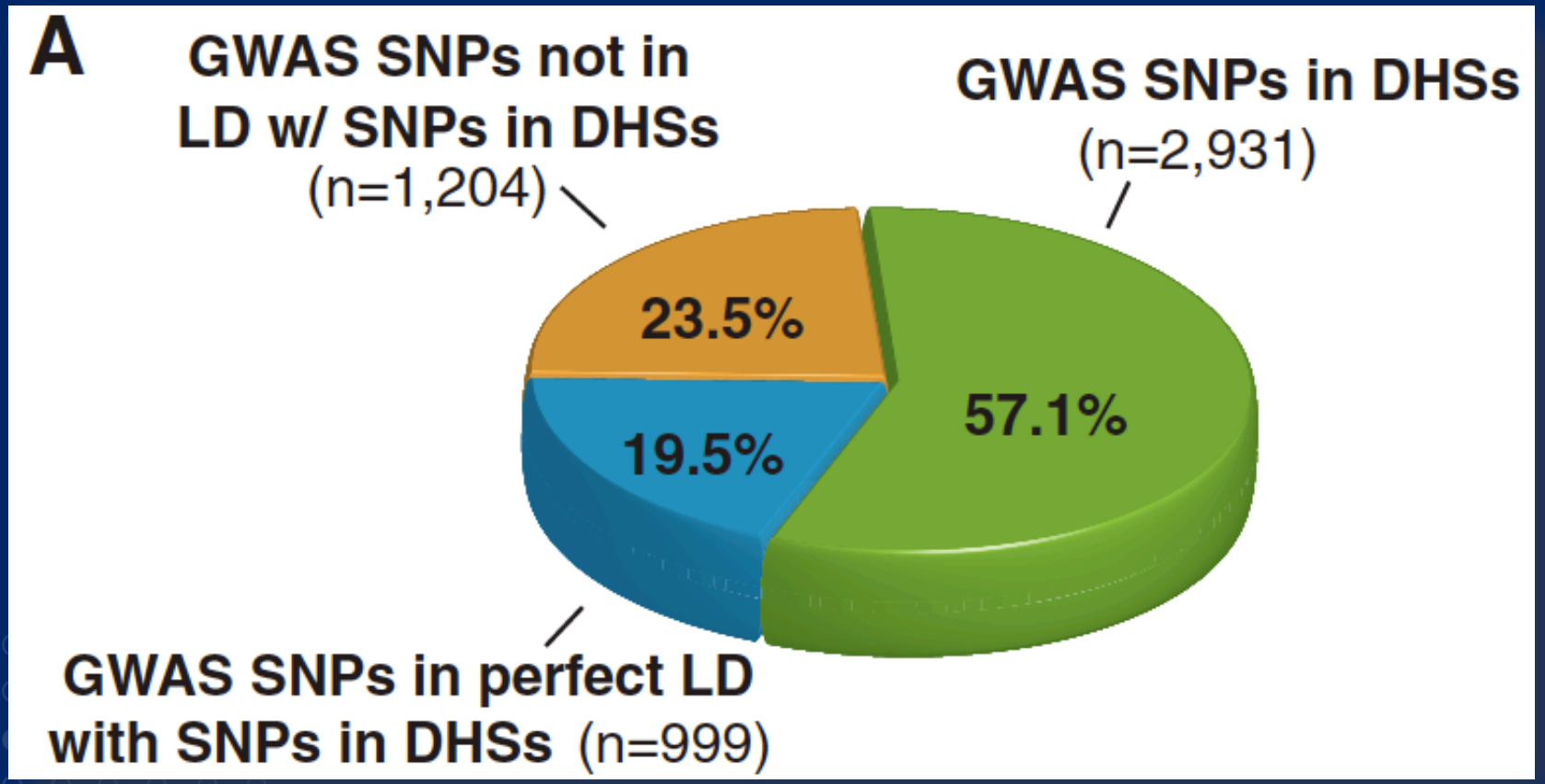
- Data Portal: http://epigenomesportal.ca/ihec/
- Goal: Production of human epigenome maps relevant to health and disease http://ihec-epigenomes.org
- Can view by consortium, by assay, by cell type

# ENCODE Consortium

NIH
NHGRI

The **Forefront** of **Genomics**®

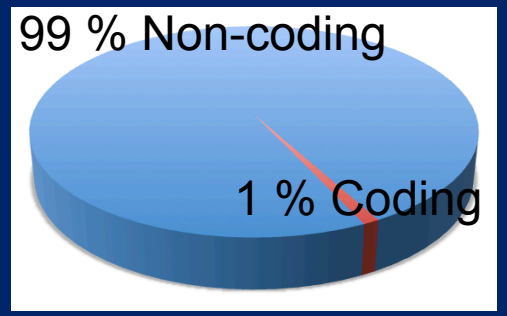
# Many GWAS Associations Lie In Regions Annotated By ENCODE And Roadmap Epigenomics Data




Stamatoyannopoulos, Science 337:1190, 2012

# Reading The Genome Is Difficult

- No simple regulatory code
- Sequence conservation can identify some candidate functional elements (but not when or where they act)
- Regulatory regions aren't always in the same order as gene targets



# Gene Regulation and Disease: hypothesis generation and refinement

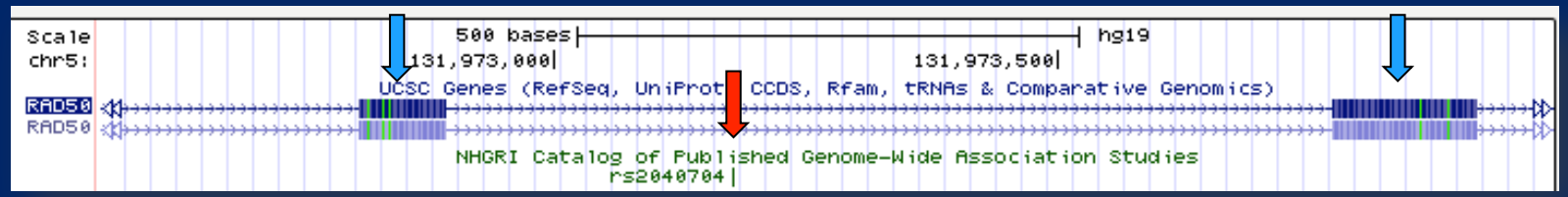
- Causal variants
- Target Gene
- Affected Cell type
- Upstream regulators

# 1,500 Letters Of Our 3 Billion Letter Genome

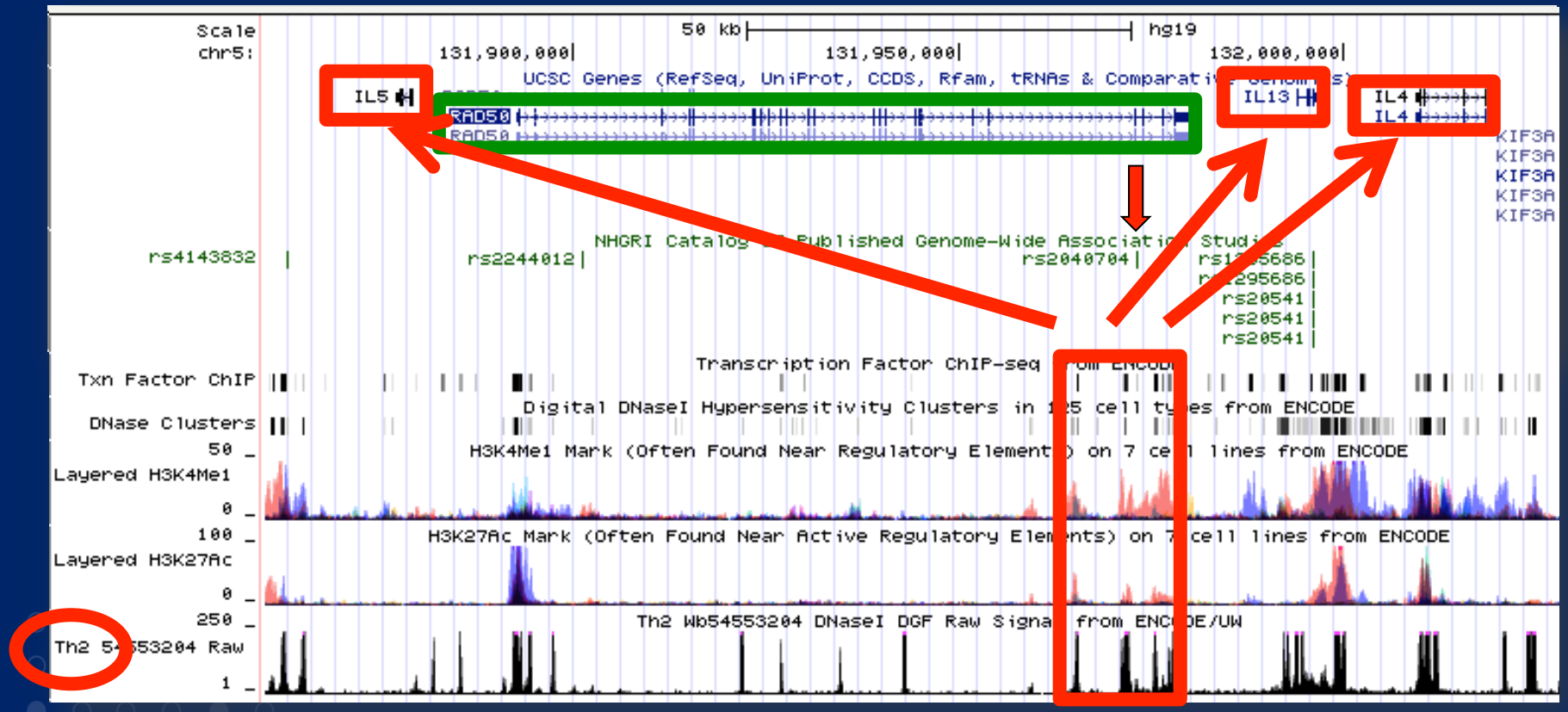
agccaagcagcaaagttttgctgctgtttatttttgtagctcttactatattctacttttaccattgaaaatattgagga
agttatttatatttctatttttttatatattatatattttatgtattttaatattactattacacataattatttttttatatatatgaag
taccaatgacttccttttccagagcaataatgaaatttcacagtatgaaaatggaagaaatcaataaaattatac
gtgacctgtggcgaagtacctatcgtggacaaggtgagtaccatggtgtatcacaaatgctctttccaaagccc
tctccgcagctcttccccttatgacctctcatcatgccagcattacctccctggacccctttctaagcatgtctttga
gattttctaagaattcttatcttggcaacatcttgtagcaagaaaatgtaaagttttctgttccagagcctaacagg
acttacatatttgactgcagtaggcattatatttagctgatgacataataggttctgtcatagtgtagatagggata
agccaaaatgcaataagaaaaaccatccagaggaaactctttttttttctttttctttttttttttttccagatggagtct
cgcacttctctgtcacccgggctggagcgcagtggtgcaatcttggctcactgcaacctccacctcctgggttca
ggtgattctcccacctcagcctcccgagtagtagctggaattacaggtgcgcgctcccacacctggctaatttttt
gtattcttagtagagatggggtttcaccatgttggccaggctggtctcaaactcctgccctcaggtgatctgccca
ccttggcctcccagtgttgggtttacaggcgtgagccaccgcgcctggcctggaggaaactcttaacaggga
aactaagaaagagttgaggctgaggaactggggcatctgggttgcttctggccagaccaccaggctcttgaa
tcctcccagccagagaaagagtttccacaccagccattgttttcctctggtaatgtcagcctcatctgttgttccta
ggcttacttgatatgtttgtaaatgacaaaaggctacagagcataggttcctctaaaatattcttcttcctgtgtcag
atattgaatacatagaaatacggtctgatgccgatgaaaatgtatcagcttctgataaaaggcggaattataac
taccgagtggtgatgctgaagggagacacagccttggatatgcgaggacgatgcagtgctggacaaaagg
caggtatctcaaaagcctggggagccaactcacccaagtaactgaaagagagaaacaaacatcagtgca
gtggaagcacccaaggctacacctgaatggtgggaagctctttgctgctatataaaatgaatcaggctcagct
actattattacactctcctgaagctaaccaacatttcctgcaacattatgtagactt

# Maps And Annotation Help Us To Understand The Sequence

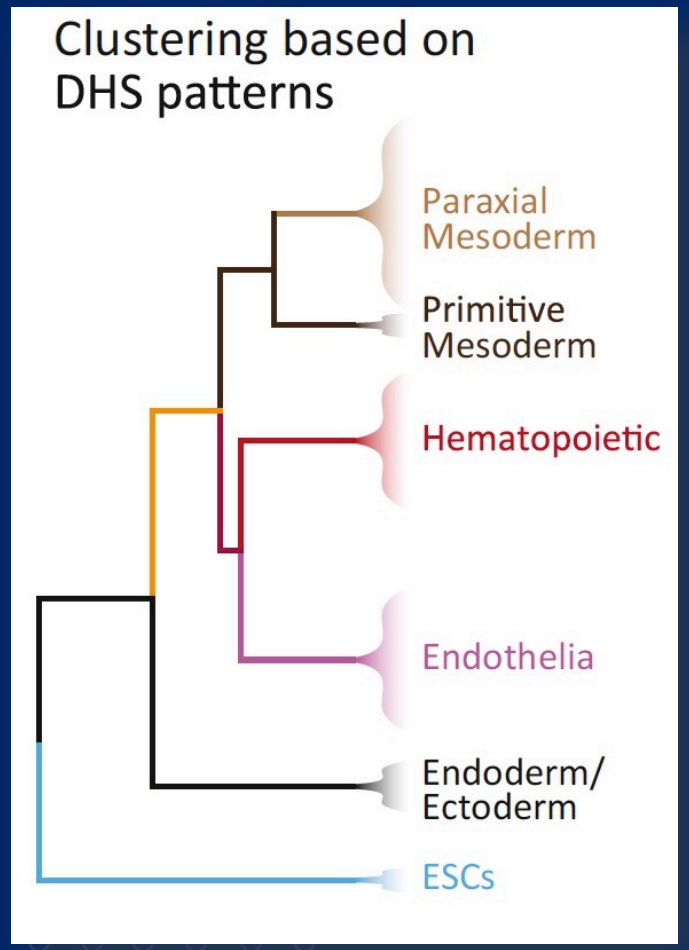
agccaagcagcaaagttttgctgctgtttattttgtagctcttactatattctacttttaccattgaaaatattgagga
agttatttatatttctatttttttatatattatatattttatgtattttaatattactattacacataattatttttttatatatatgaag
taccaatgacttccttttccagagcaataatgaaatttcacagtatgaaaatggaagaaatcaataaaattatac
gtgacctgtggcgaagtacctatcgtggacaaggtgagtaccatggtgtatcacaaatgctctttccaaagccc
tctccgcagctcttccccttatgacctctcatcatgccagcattacctccctggacccctttctaagcatgtctttga
gattttctaagaattcttatcttggcaacatcttgtagcaagaaaatgtaaagttttctgttccagagcctaacagg
acttacatatttgactgcagtaggcattatatttagctgatgacataataggttctgtcatagtgtagatagggata
agccaaaatgcaataagaaaaaccatccagaggaaactctttttttttctttttctttttttttttttccagatggagtct
cgcacttctctgtcacccgggctggagcgcagtggtgcaatcttggctcactgcaacctccacctcctgggttca
ggtgattctcccacctcagcctcccgagtagtagctggaattacaggtgcgcgctcccacacctggctaattttt
gtattcttagtagagatggggtttcaccatgttggccaggctggtctcaaactcctgccctcaggtgatctgccca
ccttggcctcccagtgttgggtttacaggcgtgagccaccgcgcctggcctggaggaaactcttaacaggga
aactaagaaagagttgaggctgaggaactggggcatctgggttgcttctggccagaccaccaggctcttgaa
tcctcccagccagagaaagagtttccacaccagccattgttttcctctggtaatgtcagcctcatctgttgttccta
ggcttacttgatatgtttgtaaatgacaaaaggctacagagcataggttcctctaaaatattcttcttcctgtgtcag
atattgaatacatagaaatacggtctgatgccgatgaaaatgtatcagcttctgataaaaggcggaattataac
taccgagtggtgatgctgaagggagacacagccttggatatgcgaggacgatgcagtgctggacaaaagg
caggtatctcaaaagcctggggagccaactcacccaagtaactgaaagagagaaacaaacatcagtgca
gtggaagcacccaaggctacacctgaatggtgggaagctctttgctgctatataaaatgaatcaggctcagct
actattattacactctcctgaagctaaccaacatttcctgcaacattatgtagactt

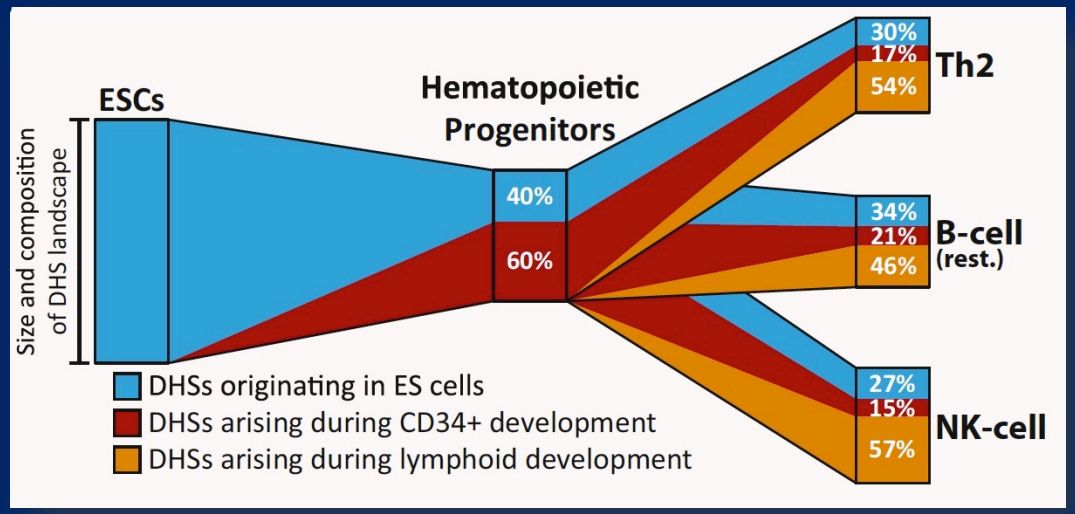


# Richer Maps Provide More Information



# ENCODE Data Are Cell-Type Specific




Stamatoyannopoulos, Cell 154:888, 2013