# Genomics of Model Organisms and Human Biology: Insights from the modENCODE Project

**Natcher Conference Center**

**June 20, 2012**

# How can we “read” the human genome sequence?

- No “instruction manual”/language rules
- Evolutionary conservation helps to identify functionally important regions

  ~5% conserved/ ~1.5% protein coding
- Moderately good at identifying protein-coding regions, but fine structures difficult to predict from sequence
- Regulatory regions can be very far away from genes
- Need unbiased experimental investigation

# ENCODE: Encyclopedia of DNA Elements

**Compile a *comprehensive encyclopedia* of all sequence features in the human genome and in the genomes of selected model organisms**

Approach:

- Apply lessons learned from the success of the Human Genome Project
- Start with well-defined pilot project
- Develop and test high-throughput technologies

# ENCODE Components

**ENCODE**

- Pilot Project Phase (9/03 – 9/07)
  Studied defined 1% of the human genome sequence using existing technologies
- Production Phase (9/07 – 9/12)
  New/continued pilot projects and expansion to whole genome studies in human

**modENCODE** (5/07 – 5/12)

Production projects to comprehensively identify functional elements in the genomes of *C. elegans* and *D. melanogaster*

**Mouse ENCODE** (9/09 – 9/11 with ARRA funds; 9/11 - 9/12 with NHGRI funds)

Limited production projects to identify functional elements in the mouse genome to inform annotation of human genome

**Technology Development** (9/03 -9/15)

High-throughput methods to identify functional elements
Funded solicitations in 2003, 2004, 2007, 2012

***Coming soon: Next phase of ENCODE*** *(2012 – 2016)*

- Data Production
- Data Analysis

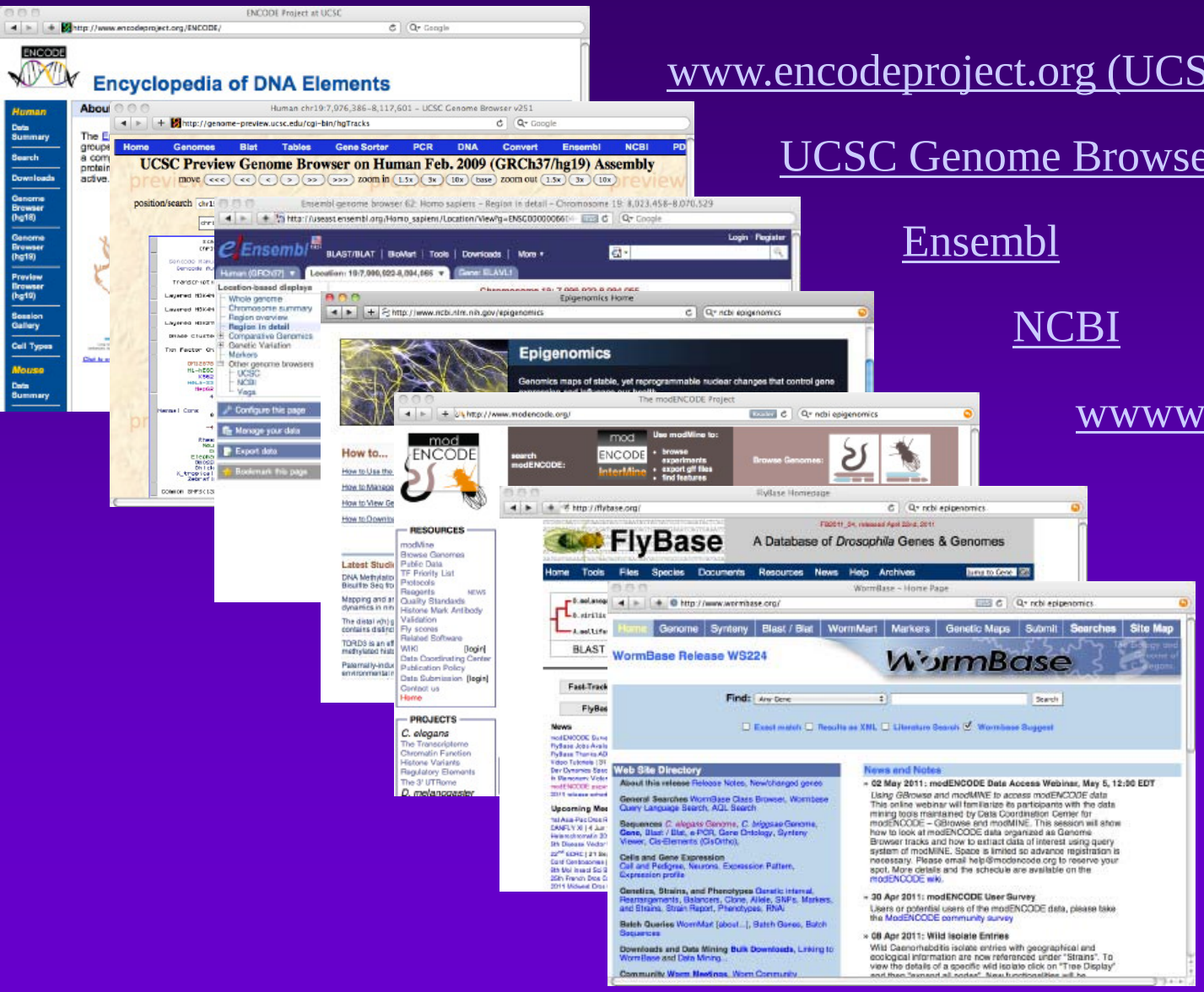
# Research Consortia

- Monthly teleconference calls
- Working groups to address specific issues
  - Data Management
  - Resources
  - Data Release
- Data Analysis Working Groups
- Annual meetings
- Inter-consortia collaborations
- Consortia publications

# Community Resources

- Use by research community to enhance understanding of:
  - regulation of gene expression on a spatial, temporal and quantitative level
  - genetic basis of disease
- Rapid pre-publication data release
- Consortia publications
- Analysis requires development of:
  - Common data reporting formats
  - Data standards
  - Analytical tools

# Data Access




www.encodeproject.org (UCSC)

UCSC Genome Browser

Ensembl

NCBI

wwww.modENCODE.org

FlyBase

WormBase

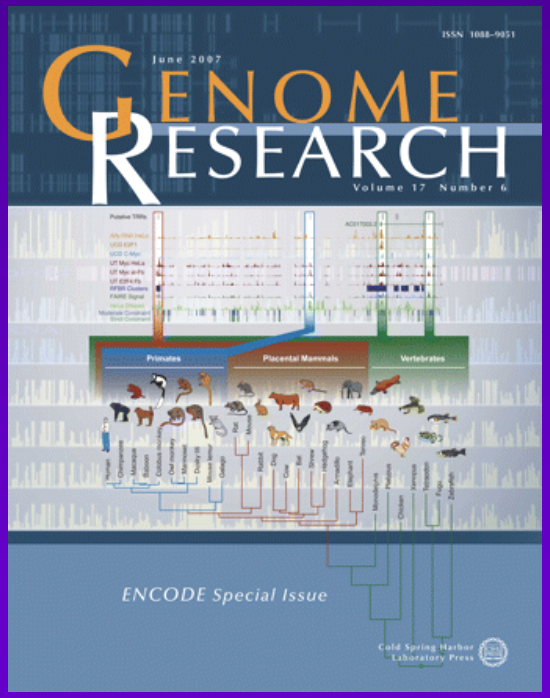
Vol 447|14 June 2007|doi:10.1038/nature05874

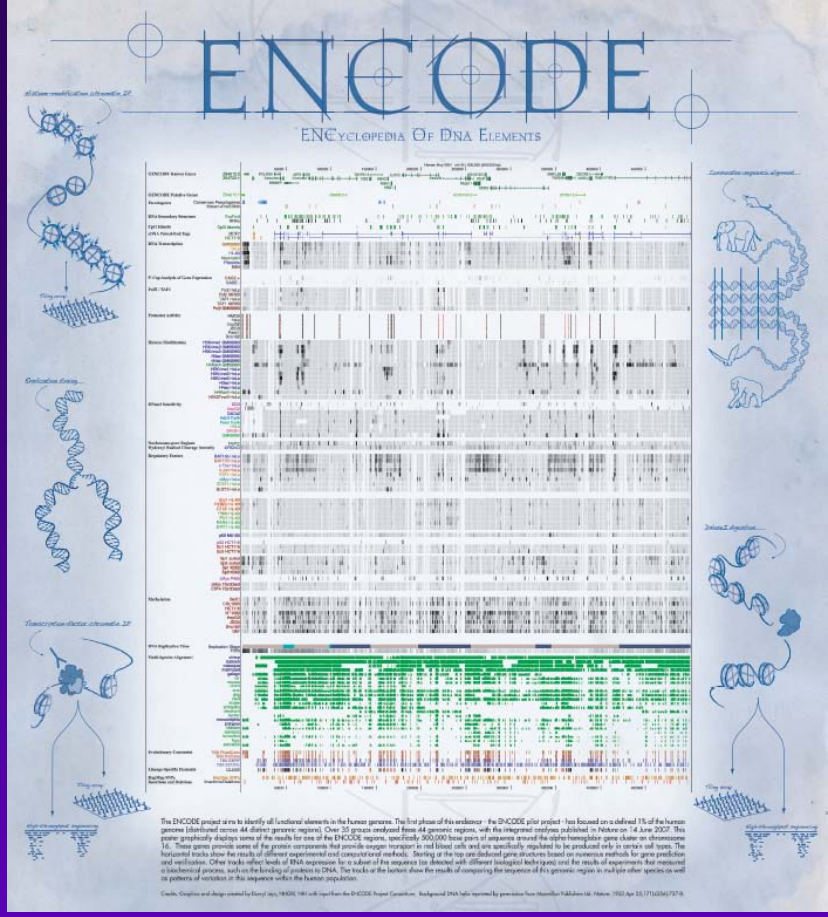


ARTICLES

# Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project

The ENCODE Project Consortium*





# modENCODE "Marker" Paper

Vol 459|18 June 2009 nature

FEATURE

## Unlocking the secrets of the genome

Despite the successes of genomics, little is known about how genetic information produces complex organisms. A look at the crucial functional elements of fly and worm genomes could change that.

Susan E. Celniker, Laura A. L. Dillon, Mark B. Gerstein, Kristin C. Gunsalus, Steven Henikoff, Gary H. Karpen, Manolis Kellis, Eric C. Lai, Jason D. Lieb, David M. MacAlpine, Gos Micklem, Fabio Piano, Michael Snyder, Lincoln Stein, Kevin P. White and Robert H. Waterston, for the modENCODE Consortium

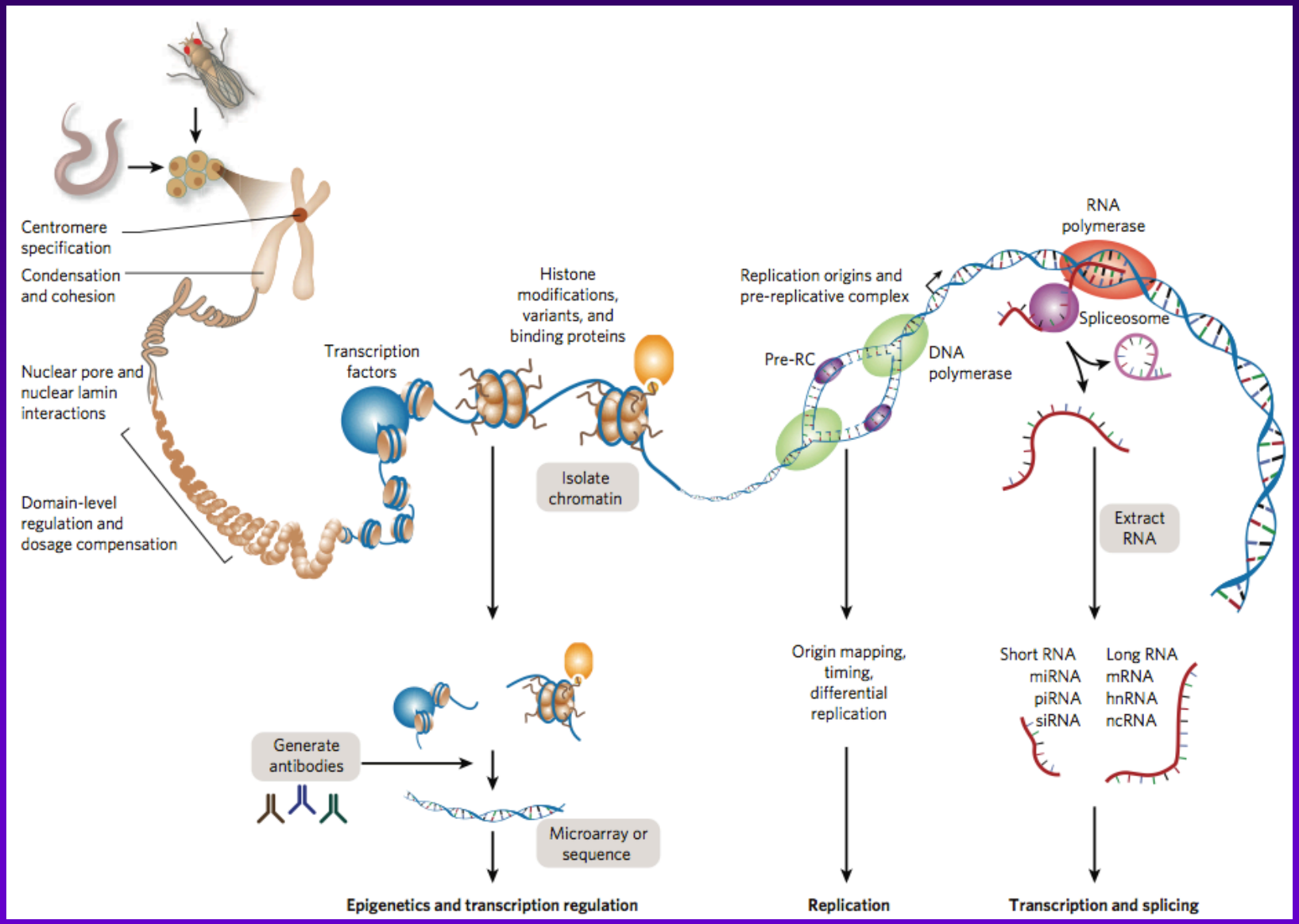
The primary objective of the Human Genome Project was to produce high-quality sequences not just for the human genome but also for those of the chief model organisms: *Escherichia coli*, yeast (*Saccharomyces cerevisiae*), worm (*Caenorhabditis elegans*), fly (*Drosophila melanogaster*) and mouse (*Mus musculus*). Free access to the resultant data has

DNA and expressed sequence tags, have been invaluable, but unfortunately these data sets remain incomplete[7]. Non-coding RNA genes present an even greater challenge[8–10], and many remain to be discovered, particularly those that have not been strongly conserved during evolution. Flies and worms have roughly the same number of known transcription factors as humans[11], but comprehensive molecular studies of gene regulatory networks have yet to be tackled in any of these species.

In an attempt to remedy this situation, the National Human Genome Research Institute (NHGRI) launched the ENCODE (Encyclopedia of DNA Elements) project in 2003, with the goal of defining the functional elements in the human genome. The pilot phase of the project focused on 1% of the human genome and a

These two model organisms, with their ease of husbandry and genetic manipulation, are pillars of modern biological research, and a systematic catalogue of their functional genomic elements promises to pave the way to a more complete understanding of the human genome. Studies of these animals have provided key insights into many basic metazoan processes, including developmental patterning, cellular signalling, DNA replication and inheritance, programmed cell death and RNA interference (RNAi). The genomes are small enough to be investigated comprehensively with current technologies and findings can be validated *in vivo*. The research communities that study these two organisms will rapidly make use of the modENCODE results, deploying powerful experimental approaches that are often not possible or practical in mam-

# Mapping Functional Elements




*Nature 459-927,2009*

# modENCODE Publications

**Integrative Analysis of the *Caenorhabditis elegans* Genome by the modENCODE Project**

**Identification of Functional Elements and Regulatory Circuits by *Drosophila* modENCODE**

19 companion papers in Nature, Genome Research, Genome Biology and Database

# Social Media

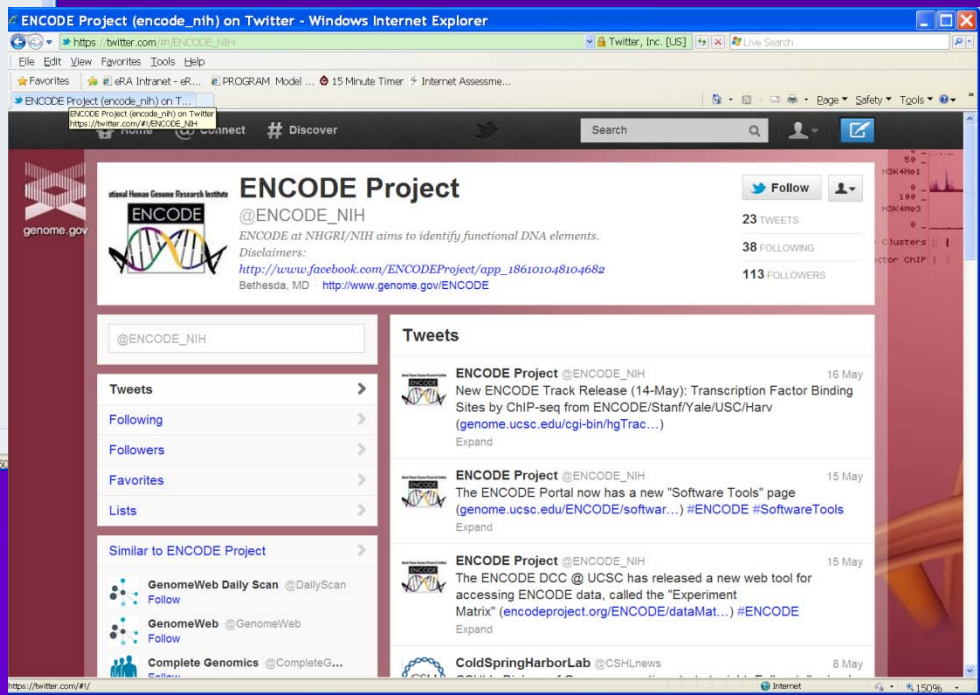
## Facebook

**ENCODE (ENCyclopedia Of DNA Elements)**

## Twitter

**@ENCODE_NIH**

**#modSymp2012**



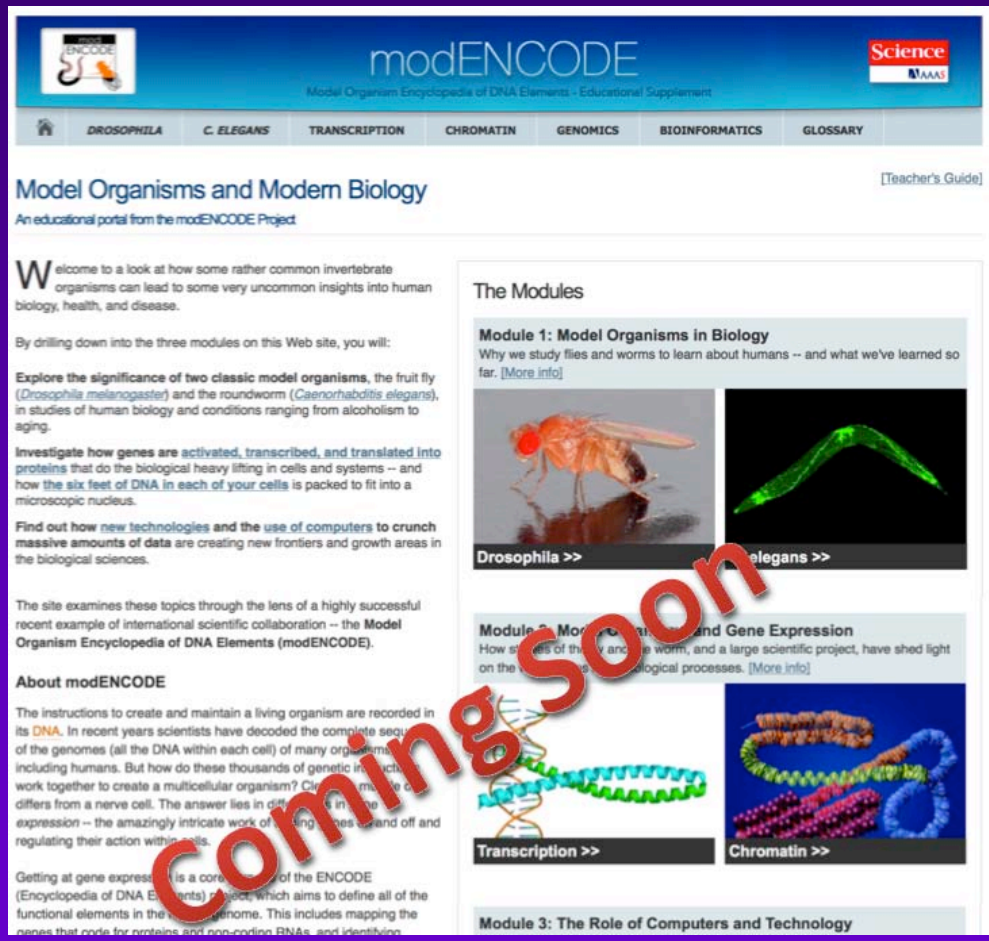
# Education Outreach

Volunteer Scientists from modENCODE and Genomics Education Partnership teamed up with Science/AAAS to create education website

Target audience is general public – high school students in particular

Six segments provide rich background on work in flies and worms

- Introduction to chromatin structure and eukaryotic transcription
- Description of modern high-throughput genomic technologies
- Bioinformatics approaches for data analysis




- To receive notification when beta test site goes live, email Stewart Wills at soleditor@aaas.org.
- See poster and handout in lobby for more information.

# Genomics of model organisms and human biology: Insights from the modENCODE Project

June 20-21, 2012

- Showcase modENCODE findings
  - Data access
  - Data analysis
- Panel discussion on utility of modENCODE data
  - basic biological processes
  - human biology and disease
- Speakers from within modENCODE and other members of research community
- Held in conjunction with Model Organism to Human Biology: Cancer Genomics meeting

# The modENCODE Consortium

**Susan Celniker** (Justen Andrews, Steven Brenner, Michael Brent, Peter Cherbas, Tom Gingeras, Brenton Graveley, Roger Hoskins, Thom Kaufman, Norbert Perrimon)

**Stephen Henikoff** (Kamran Ahmed)

**Gary Karpen** (Sarah Elgin, Mitzi Kuroda, Peter Park, Vincent Pirrotta)

**Manolis Kellis** (Mark Gerstein)

**Eric Lai**

**Jason Lieb** (Julie Ahringer, Abby Dernberg, Xiaole Liu, Eran Segal, Susan Strome)

**David MacAlpine** (Terry Orr-Weaver)

**Fabio Piano** (Kris Gunsalus, John Kim)

**Michael Snyder** (Mark Gerstein, Anthony Hyman, David Miller, Valerie Reinke, Robert Waterston)

**Lincoln Stein** (Gos Micklem, Suzanna Lewis)

**Robert Waterston** (Mark Gerstein, Phillip Green, Michael MacCoss, David Miller, Valerie Reinke, Frank Slack)

**Kevin White** (Hugo Bellen, Manolis Kellis, James Posakony, Bing Ren, Steven Russell, Robert White)

**… and many senior scientists, postdocs, students, technicians, computer scientists, statisticians and administrators in these groups**

# Many Thanks!

# NHGRI Staff

## Program Directors

Elise Feingold

Peter Good

Michael Pazin

## Division Director

Mark Guyer

## Program Analysts

Leslie Adams

Caroline Kelly

### Former Program Analysts

Laura Liefer Dillon

Rebecca Lowdon

Jessica Melone

Judith Wexler

Julia Zhang

ENCODE

mod
ENCODE

ENCODE

mod
ENCODE