Disclosures

Presenters have no relevant
conflicts to disclose

Introduction to ENCODE

2:30 - 2:45   **Introduction & overview of the Consortium**
Eurie Hong, Ph.D.
Project Manager, ENCODE Data Coordination Center
Stanford University

2:45 - 3:00   **Using the ENCODE Portal (hands-on)**
J. Seth Strattan, Ph.D.
Data Wrangler, ENCODE Data Coordination Center
Stanford University

3:00 - 3:15   **Genome annotations**
Mark Gerstein, Ph.D.
Albert L. Williams Professor of Biomedical Informatics
Yale University

3:15 - 3:30   **How to interpret and read a track (hands-on)**
Richard Sandstrom
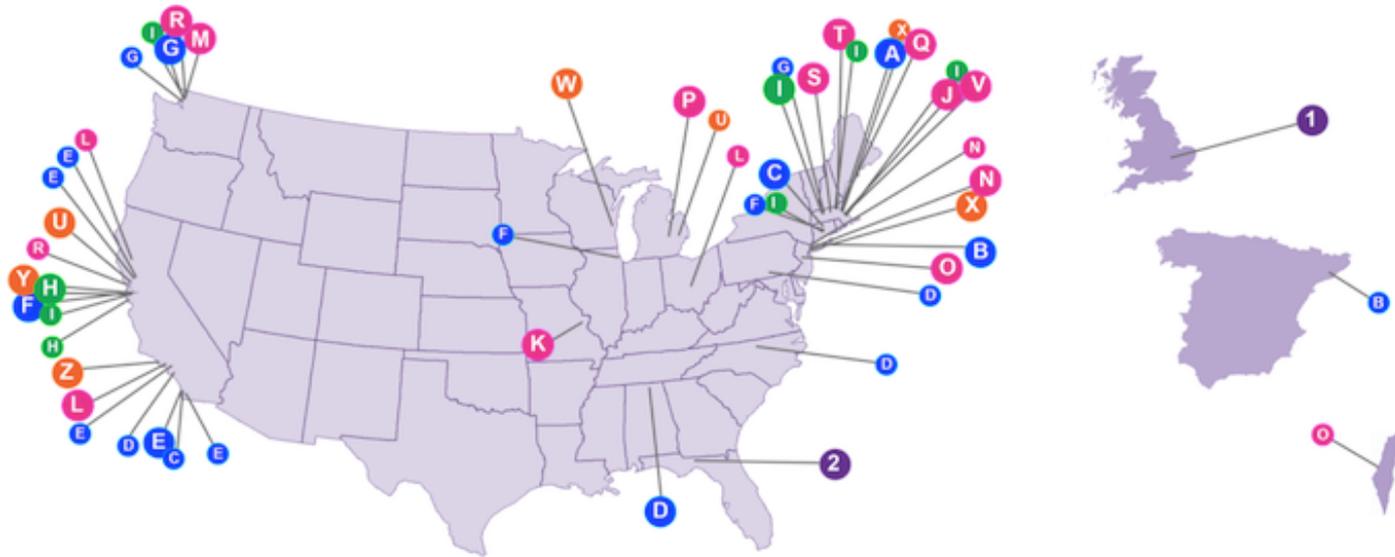Research Scientist, Stamatoyannopoulos lab
University of Washington

3:30 - 4pm   **Variant annotation using ENCODE data (hands-on)**
Luke Ward, Ph.D.
MIT

Who is the ENCODE Consortium?

Production Groups
A Broad Institute
B Cold Spring Harbor;
Centre for Genomic Regulation (CRG);
C University of Connecticut Health Center;
UCSD
D HudsonAlpha; Pennsylvania State;
UC Irvine; Duke; Caltech
E UCSD; Salk Institute ; Joint Genome Institute;
Lawrence Berkeley National Laboratory; UCSD
F Stanford; University of Chicago; Yale
G University of Washington;
Fred Hutchinson Cancer Research Center;
University of Massachusetts Medical School

Data Coordination Center
H Stanford; UCSC

Data Analysis Center
I University of Massachusetts Medical School;
Yale; MIT; Stanford; Harvard; University of Washington

Technology Development Groups
J MIT
K Washington University, St. Louis
L USC; Ohio State University; UC, Davis
M University of Washington
N Sloan-Kettering; Weill Cornell Medical College
O Princeton; Weizmann
P University of Michigan
Q Broad Institute
R University of Washington; UCSF
S Advanced RNA Technologies, LLC
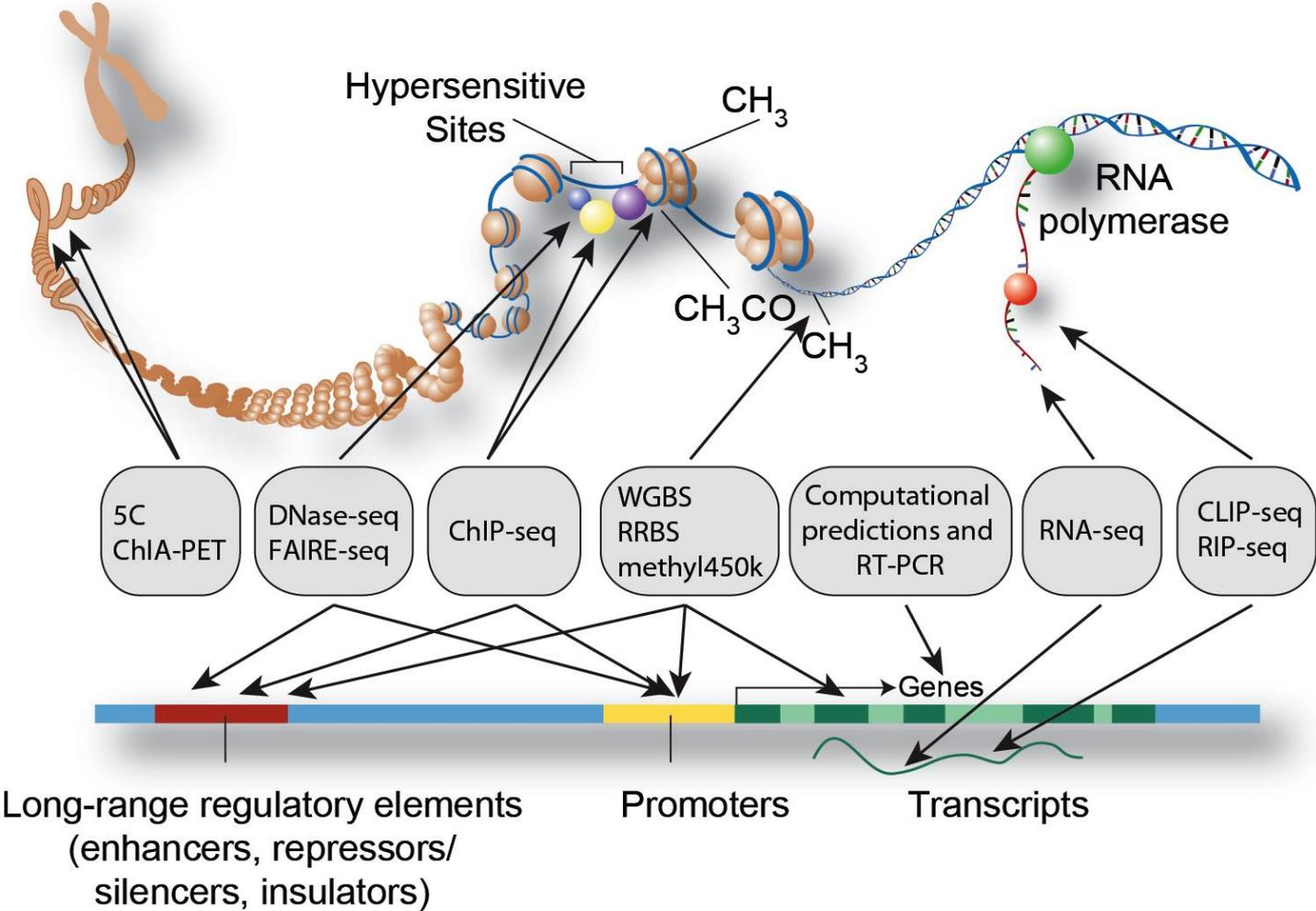**T** Harvard

Computational Analysis Groups
U Berkeley; Wayne State University
V MIT
W University of Wisconsin
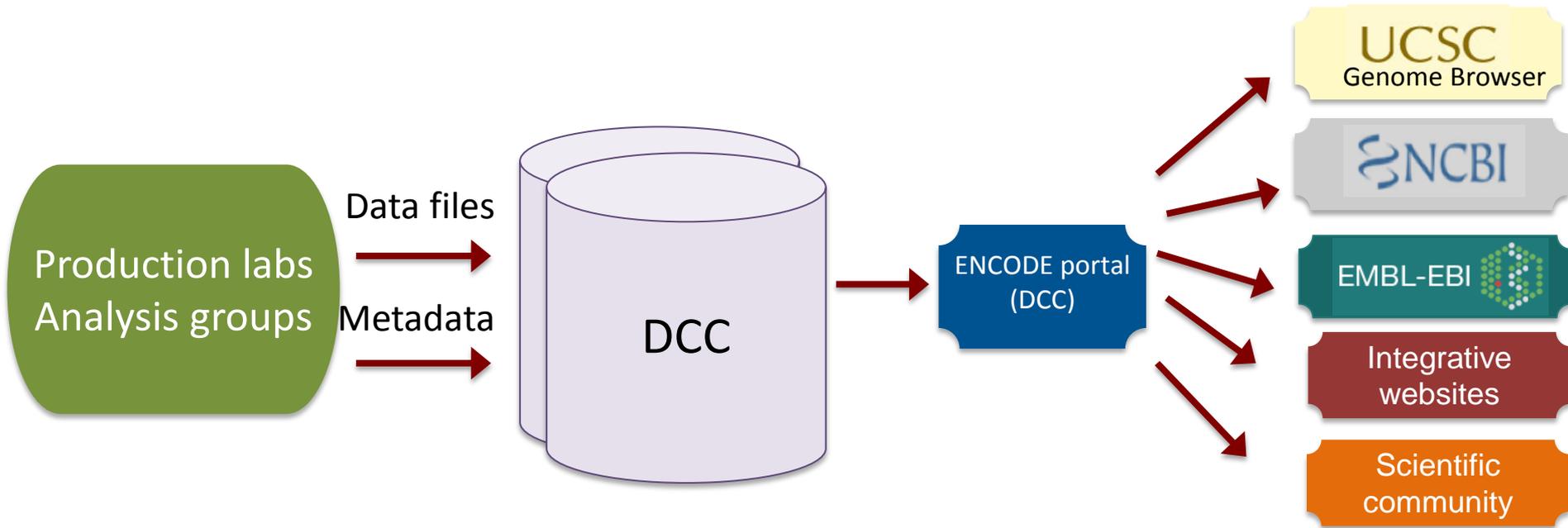**X** Sloan-Kettering; Broad Institute
Y Stanford
Z UCLA

Affiliated Groups
1 Wellcome Trust Sanger Institute
2 Florida State University

Image credit: NHGRI

What data does the consortium generate?

ENCODE data flow

Role:	**Data generation**	**Data organization**	**Data access**
Tasks:	Perform assays	Define submission process	Web-based searches
	Perform analyses	Data processing & validation	Data downloads
	Validate data	Data file storage	
	Submit data files	Metadata curation	
	Submit metadata		

Principles of data organization

Transparency of methods
- How was the experiment performed?
- Are ENCODE data standards met?

Reproducibility of results
- What files were used?
- What software & parameters were used to analyze the data?

Interoperability with other genomic projects
- Are standard file formats used?
- Can the metadata allow easy integration with other data?

Metadata: capture the experimental design
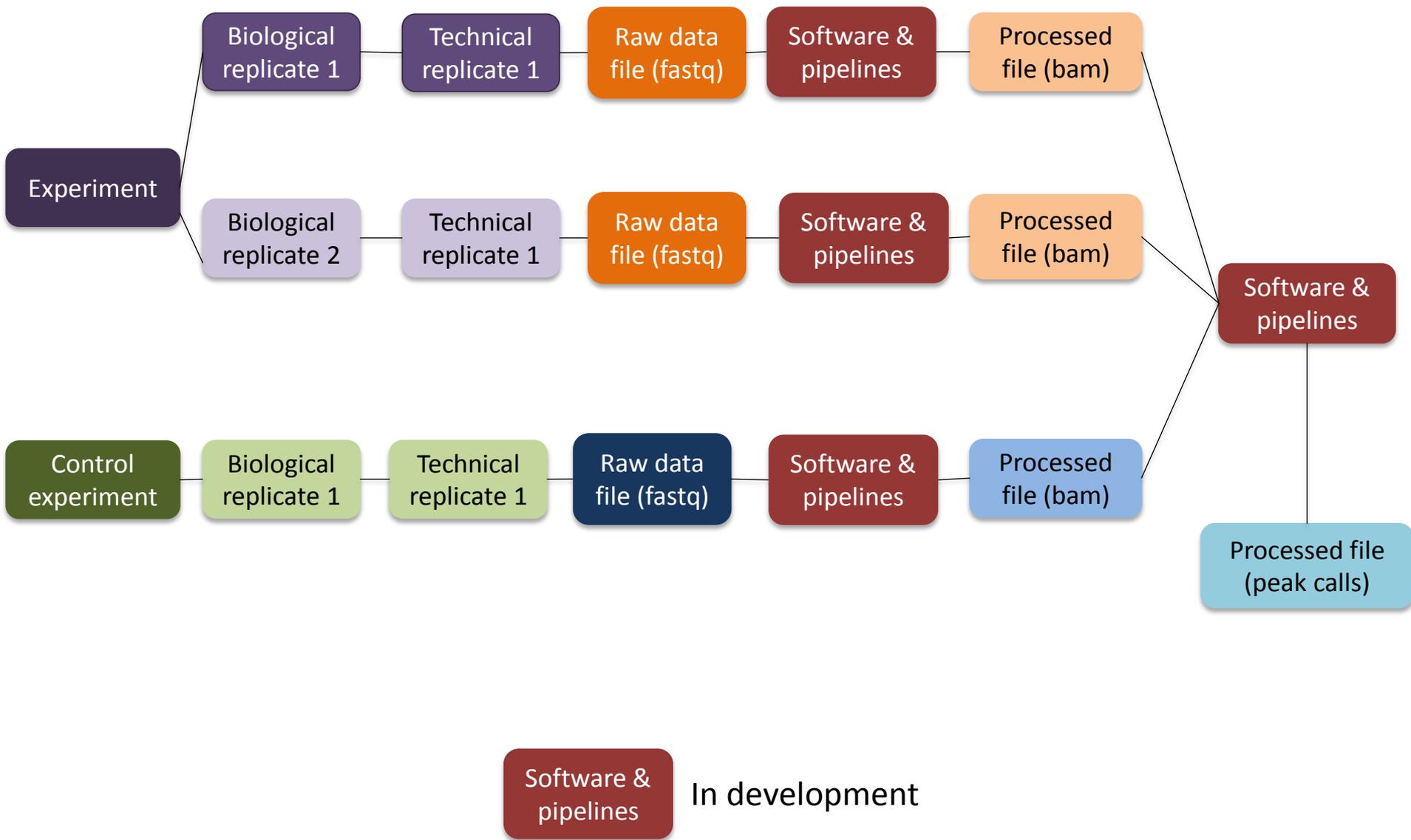
Metadata: capture details of methods

Experiment with replicates (*ENCSR000DRY*)

Biosamples
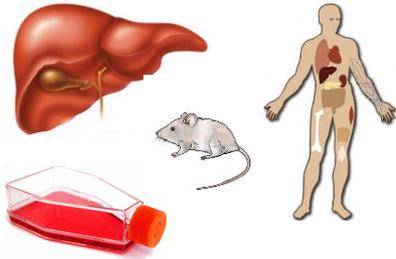
- Type (e.g. tissue, cell line)
- Ontology term name
- Source, product id, lot id
- Treatments
- Knockdown
- Fusion construct information
- Donor or strain information
- Dates (e.g. growth, harvest, procurement)
- Passage number
- Starting amount
- Lab assigned IDs

Antibodies
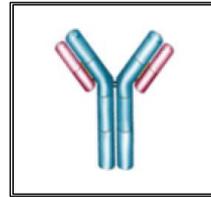
- Source, product id, lot id
- Isotype
- Antigen
- Host
- Purification method
- Validation status
- NHGRI approval status
- Target
- Species
- Dbxrefs

Libraries
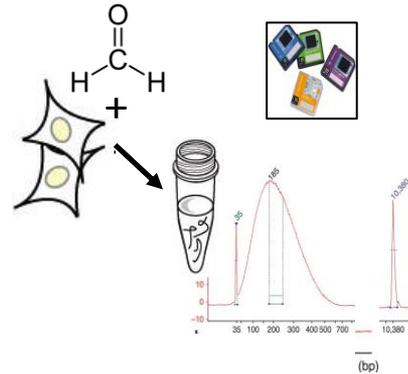
- Library preparation protocol
- Strand specificity
- Size selection method
- Validation document
- Lysis method
- Sonication method
- Extraction method
- Nucleic acid type
- Nucleic acid size range

Files

Alignment

Peak calls
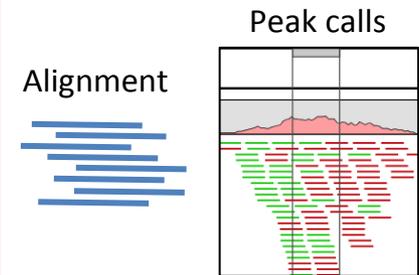
- Reference genome version
- Alignment software
- Software parameters
- Software version
- Quality metrics (e.g. NRF, FRiP)

(selected subset of all metadata)

ENCBS095DKV (biosample)
ENCDO826IFN (donors)

ENCAB964IAU

ENCLB239KAN

ENCFF254TDA

Introduction to the ENCODE portal
https://www.encodeproject.org/