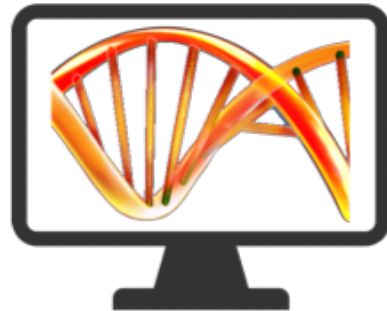




# Navigating the ENCODE Encyclopedia: Exploring Candidate Regulatory Elements, Linked Genes, and Genetic Variation with SCREEN



Jill Moore, Michael Purcaro, and Zhiping Weng  
University of Massachusetts Medical School

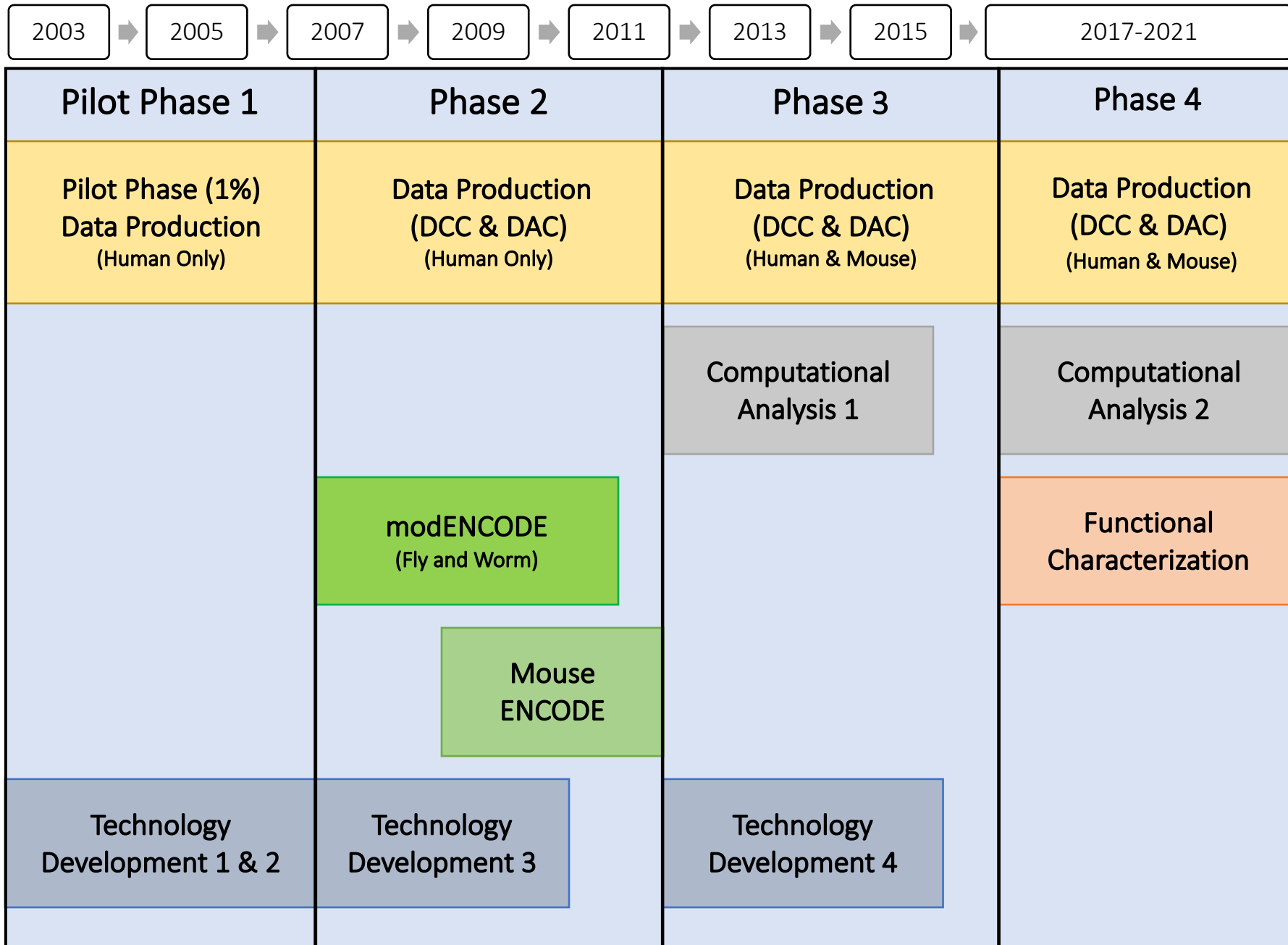
# The Encyclopedia Of DNA Elements Consortium



## Goals:

- Catalog all functional elements in the genome
- Develop freely available resource for research community
- Study human as well as other species
- Project components: data generation, analysis, and repository

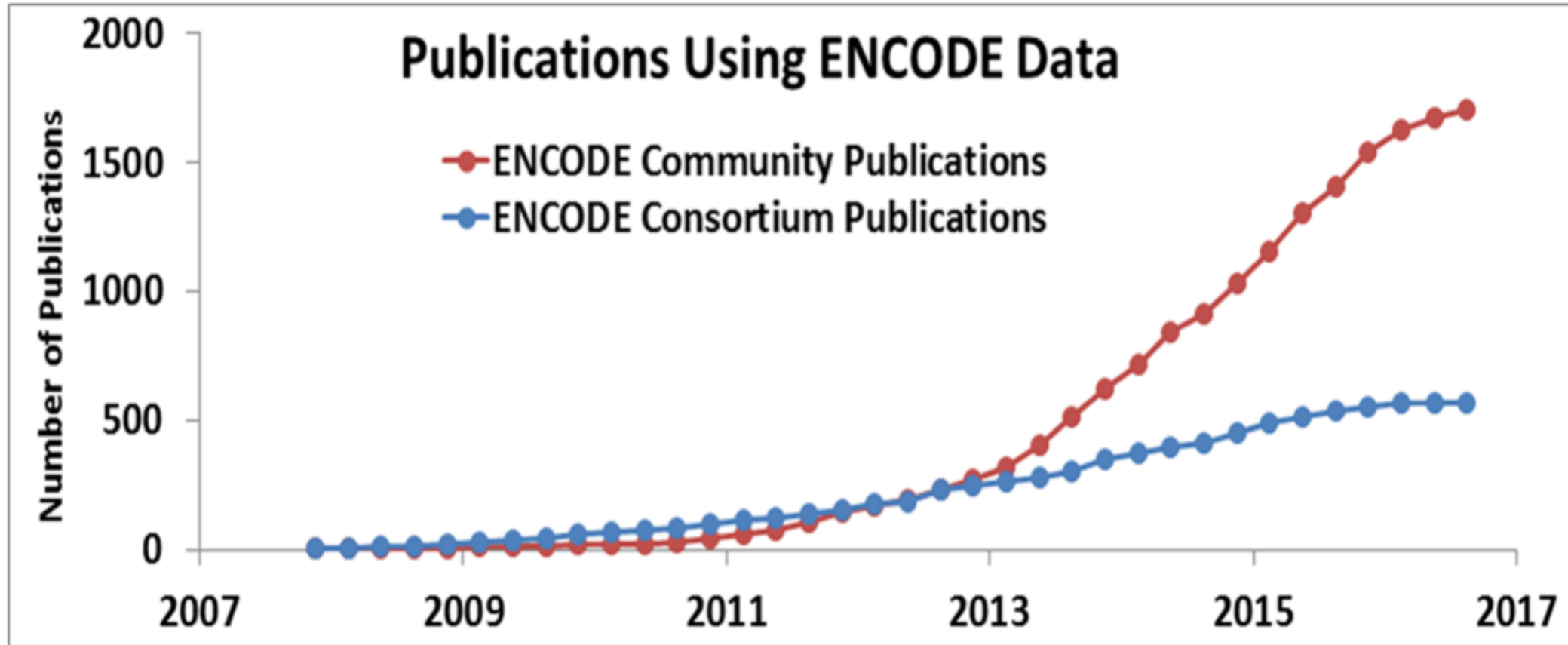
# ENCODE Timeline



# Community Resource

- Rapid pre-publication data release
- Open access of software tools and analysis pipelines
- Ensure high data quality
  - Data standards
  - Quality control metrics
  - Analytical tools
- ENCODE Encyclopedia: analysis results

# Publications Using ENCODE Data



# Publications Using ENCODE Data

~ 550 Consortium publications

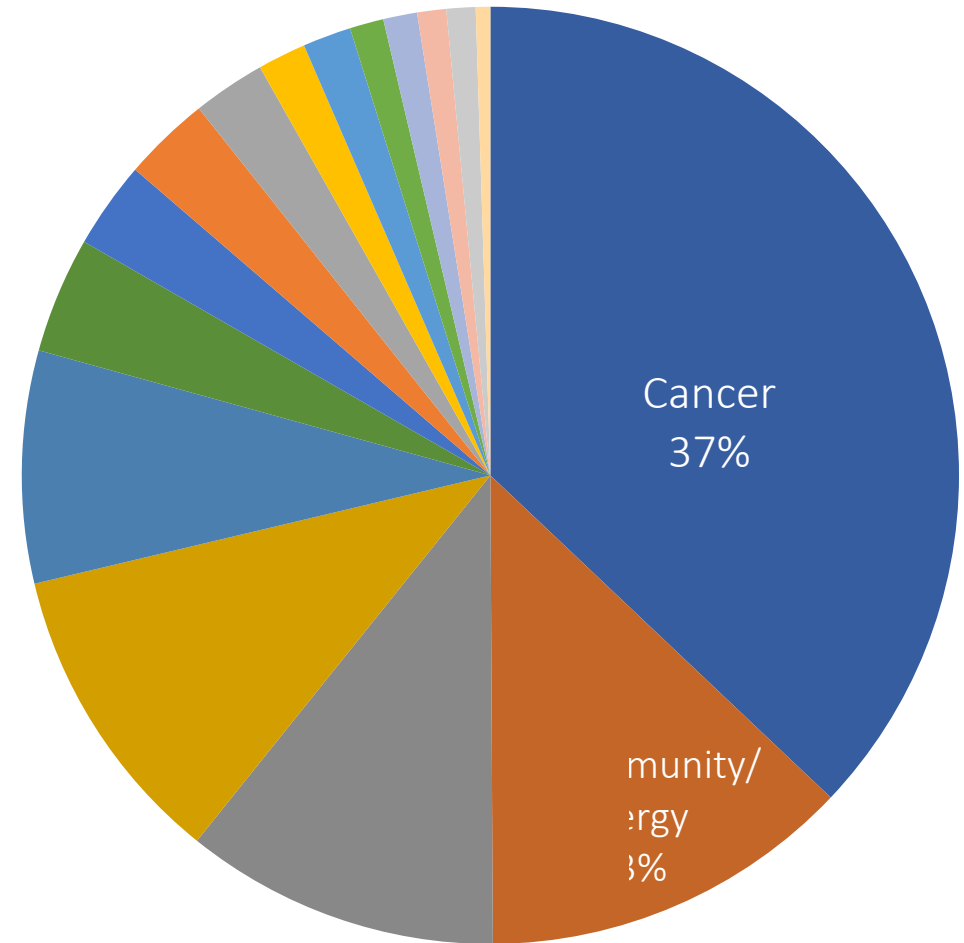
~1750 community publications using ENCODE data:

~725 Human Disease

~770 Basic Biology

~260 Methods/Software  
Development

<https://www.encodeproject.org>



# Advantages of a Consortium

1. Publicly available datasets, and lots of them!
2. Integrative analysis of many assays performed on the same cells or tissues
3. Uniform data processing and quality control, rich metadata, and standards

# But, ENCODE Teleconferences

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
	<b>1</b> RNA Call Binding Call Cancer	<b>2</b> Outreach Call Human Resources Call	<b>3</b> Data Release Call	<b>4</b>	<b>5</b> AWG: GWAS	<b>6</b>
<b>7</b>	<b>8</b> RNA Call Peak Calling	<b>9</b> Human Resources Operational Call	<b>10</b> Data Release Call	<b>11</b> DCC Antibody Call DCC Call	<b>12</b> Production PI Call AWG Call	<b>13</b>
<b>14</b>	<b>15</b> RNA Call Binding Call Cancer	<b>16</b> Outreach Call Human Resources Call	<b>17</b> Data Release Call	<b>18</b> DAC Call	<b>19</b> <b>Consortium Call</b> DNase Call AWG: GWAS	<b>20</b>
<b>21</b>	<b>22</b> RNA Call Peak Calling	<b>23</b> Human Resources Operational Call	<b>24</b> Data Release Call	<b>25</b> DCC Antibody Call EDCAC Call	<b>26</b> Functional Characterization Call AWG Call	<b>27</b>
<b>28</b>	<b>29</b> RNA Call	<b>30</b>	<b>31</b> Data Release Call			

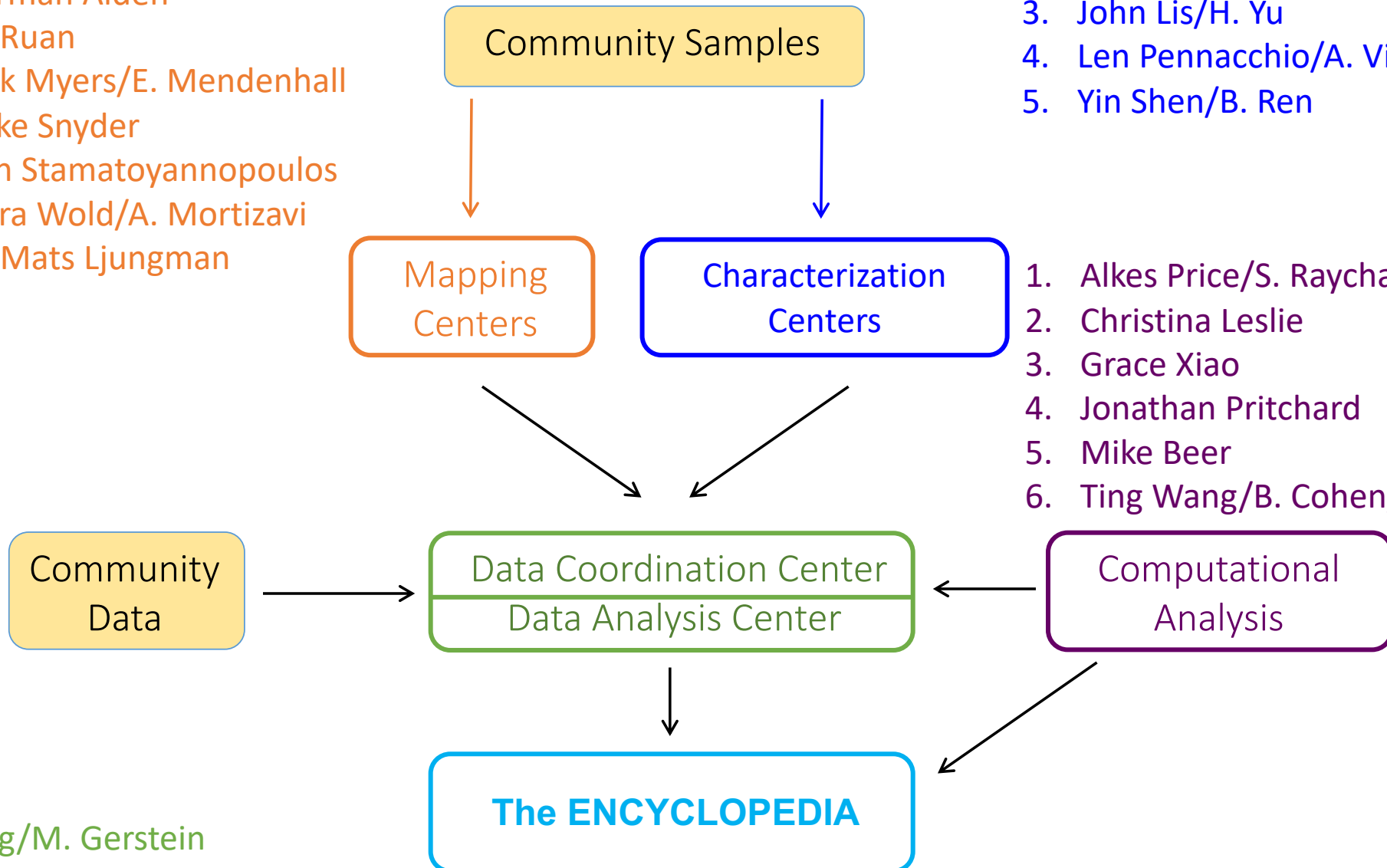


# ENCODE 4

1. **Histone Marks:** Brad Bernstein/C. Nusbaum
2. **Hi-C:** Erez Lieberman Aiden
3. **ChIA-PET:** Yijun Ruan
4. **TF CHIP-seq:** Rick Myers/E. Mendenhall
5. **TF CHIP-seq:** Mike Snyder
6. **DNase-seq:** John Stamatoyannopoulos
7. **RNA-seq:** Barbara Wold/A. Mortizavi
8. **RNA dynamics:** Mats Ljungman

1. Nadav Ahitav/J. Shendure
2. Will Greenleaf/M. Bassik
3. John Lis/H. Yu
4. Len Pennacchio/A. Visel
5. Yin Shen/B. Ren

1. Alkes Price/S. Raychaudhuri
2. Christina Leslie
3. Grace Xiao
4. Jonathan Pritchard
5. Mike Beer
6. Ting Wang/B. Cohen/C. Feschotte



DCC: Mike Cherry  
DAC: Zhiping Weng/M. Gerstein

# Summary of ENCODE 3 Data Production

# The ENCODE Portal: [encodeproject.org](http://encodeproject.org)

ENCODE

Data

Encyclopedia

Materials & Methods

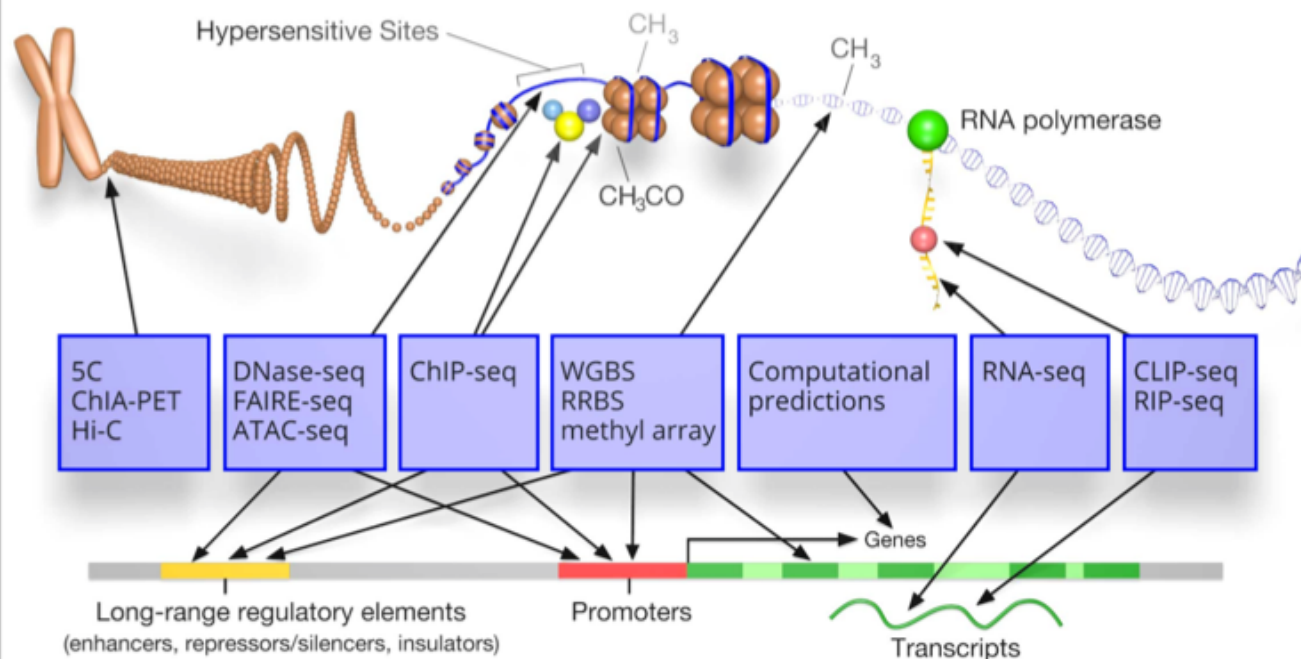
Help

Search...



The Epigenomic Roadmap data are also on the ENCODE Portal!

## ENCODE: Encyclopedia of DNA Elements



The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

[Get Started](#)



Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

# Total ENCODE Data

BIOSAMPLE

9154 results

Clear Filters

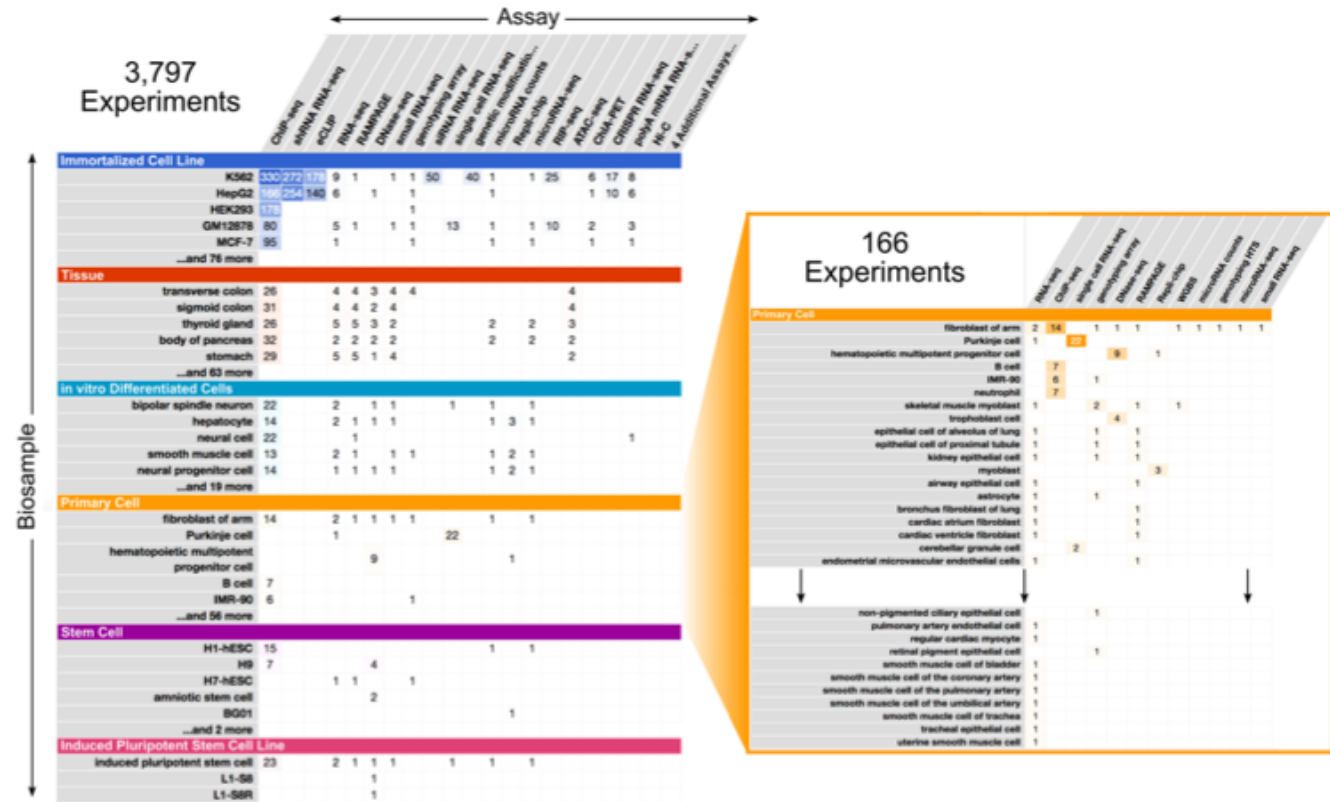
ASSAY

...and 20 more

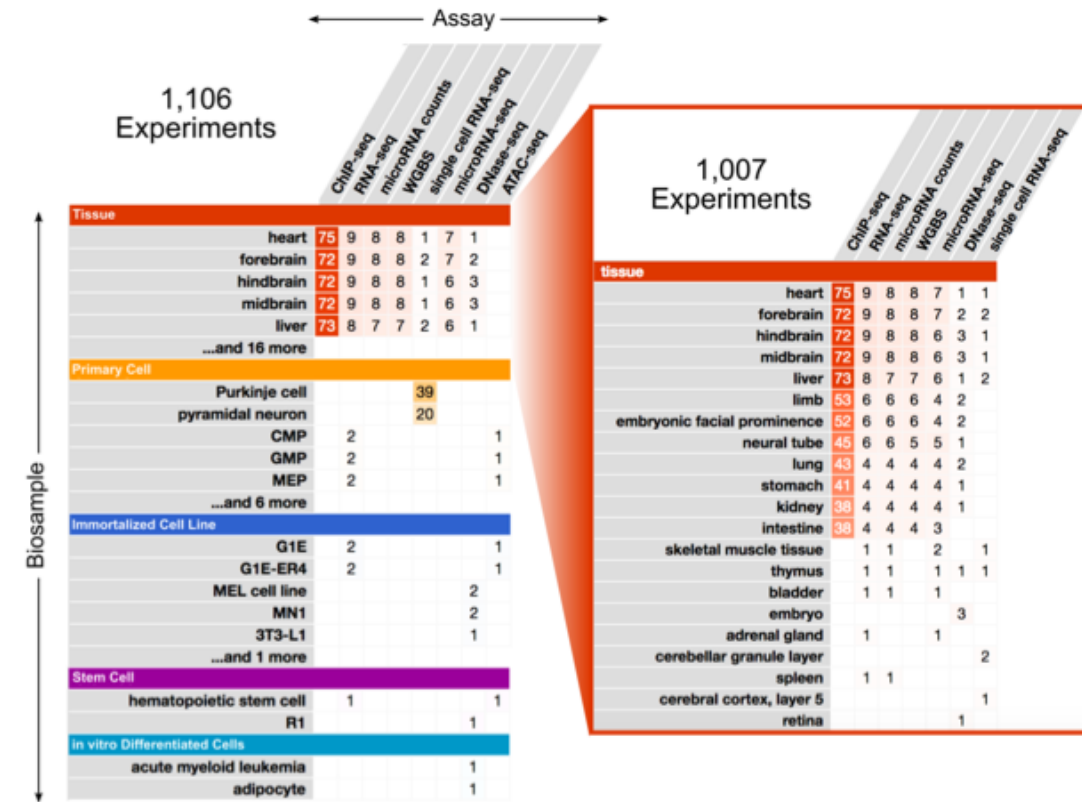
	ChIP-seq	shRNA RNA-seq	DNase-seq	total RNA-seq	eCLIP	DNAse array	polyA mRNA RNA-seq	RNA microarray	small RNA-seq	RNA Bind-n-Seq	RAMPAGE	ATAC-seq	genotyping array	single cell RNA-seq	microRNA counts	Replic-seq	RRBS	WGBS	CAGE
<b>immortalized cell line</b>																			
K562	609272	7	12	180	3	19	12	8	1	2	1	1	6	1	1	9			
HepG2	325254	3	6	148	3	11	7	3		2	1		6	2	1	6			
GM12878	237	2	6		3	11	8	6	1	2	13	1	1	6	2	2	6		
HEK293	239				2	1				2					2				
MCF-7	176	8	1		2	5	7	7		2		1	1	6	5	3			
+ See 150 more...																			
<b>tissue</b>																			
liver	128	5	11		1	11	1	2	7	2	7	7		1	7				
heart	91	8	11		6	1	2	7	1	8	9				8				
forebrain	72	4	9					7	2	8	9				8				
stomach	73	5	9	3	1	4	5	6		4	4			1	4				
midbrain	72	5	9					7	1	8	8			8					
+ See 104 more...																			
<b>primary cell</b>																			
endothelial cell of umbilical vein	36	2		1	5	2	1			1	1	6		5					
Purkinje cell		1								6									
keratinocyte	24	2		5	6	3				6				5					
IMR-90	18	1	1	2	3	9			2	6	1	3							
fibroblast of lung	27	3	1	3	3	3	1		2			1	1						
+ See 120 more...																			
<b>in vitro differentiated cells</b>																			
neural cell	26			1	3	1													
myocyte	26	1	1	2						1	1								
bipolar neuron	22	1	2	1	1					1	1	1							
myotube	14	2	1	2	3		1	3						2					
hepatocyte	15	1	2	1	1	1				1	1			1					
+ See 31 more...																			
<b>stem cell</b>																			
H1-hESC	105	2		2	7	1	3		1	1	1		1	2	4				
G1E-ER4	23	1	2	7					1										
ES-E14	16	1	1																
G1E	11	1	2	2					1										
H9	7	5		1															
+ See 22 more...																			
<b>in vitro sample</b>																			
<b>induced pluripotent stem cell line</b>																			
GM23338	23	1	2	1	1	1			1	1	1								
induced pluripotent stem cell		1			1														
CWRU1		1			1														
iPS-NIH11		1			1														
iPS-NIH7		1			1														
+ See 3 more...																			

# Data Generated During ENCODE Phase 3

## Human



## Mouse



# ENTEx: ENCODE + GTEx

- Human tissue samples from multiple donors

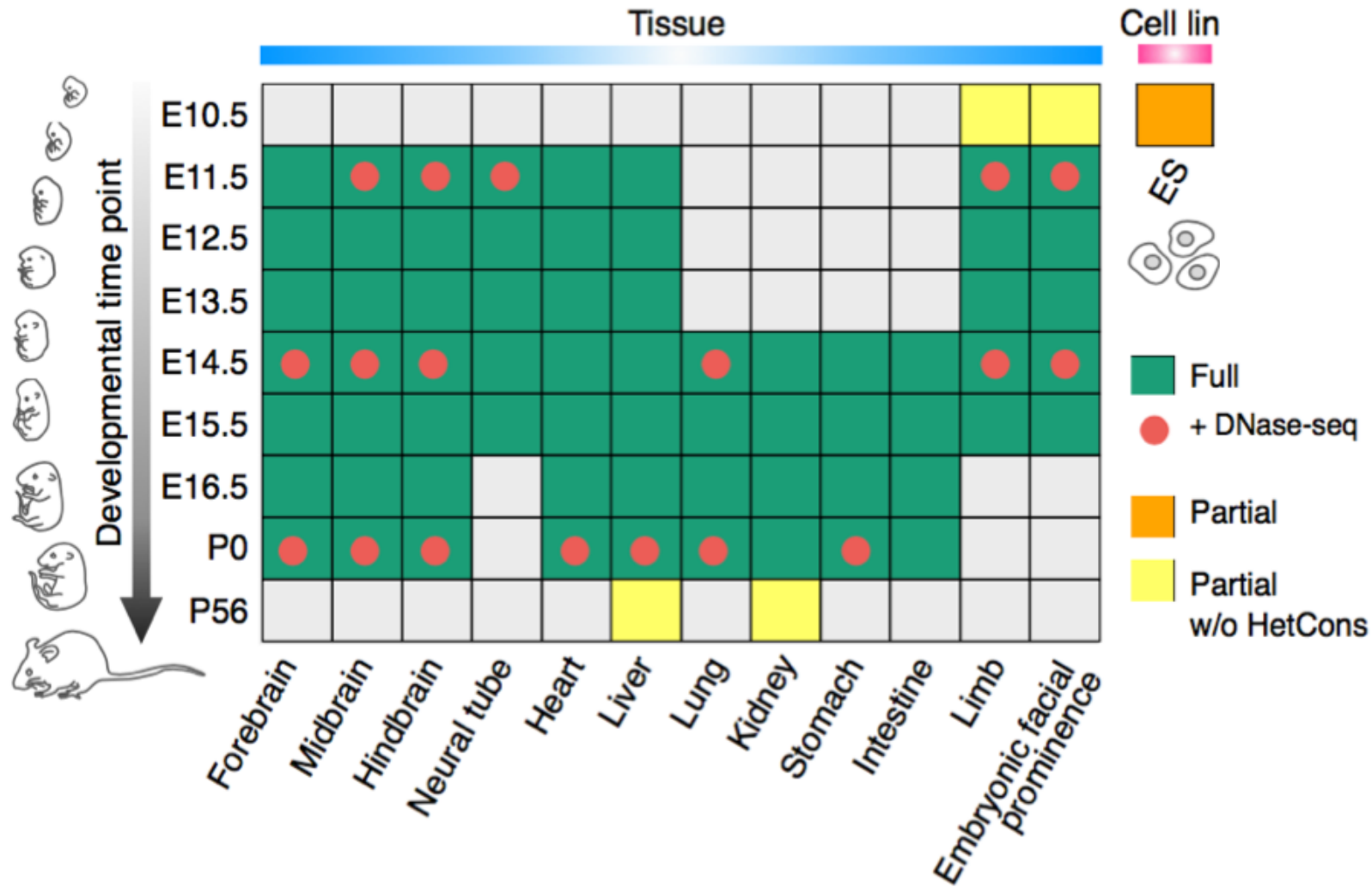
1062 results

Clear Filters

ChIP-seq  
 DNase array  
 total RNA-seq  
 RAMPAGE  
 small RNA-seq  
 ATAC-seq  
 microRNA counts  
 microRNA-seq  
 DNase-seq  
 genotyping HTS  
 Hi-C  
 genotyping array  
 eCLIP

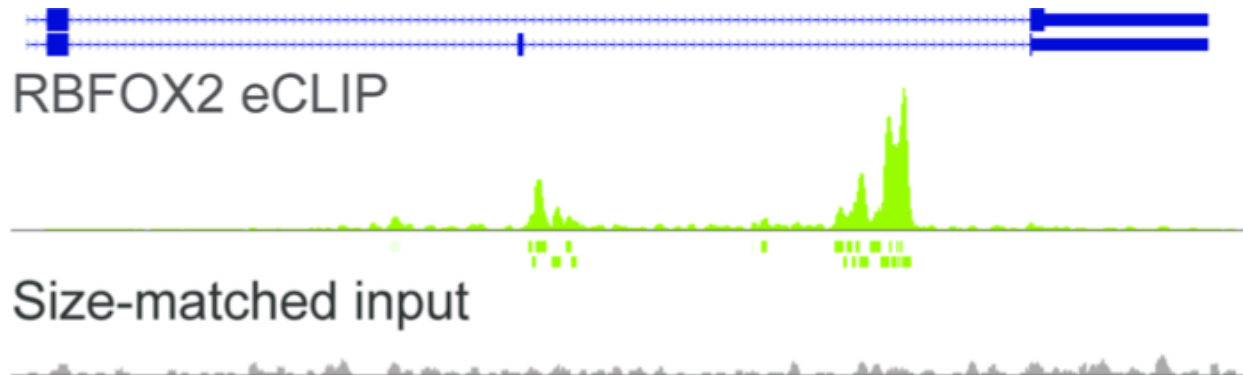
tissue	ChIP-seq	DNase array	total RNA-seq	RAMPAGE	small RNA-seq	ATAC-seq	microRNA counts	microRNA-seq	DNase-seq	genotyping HTS	Hi-C	genotyping array	eCLIP
transverse colon	29	4	4	4	4	4		3	8	4	4		
gastrocnemius medialis	33	4	4	3		4	2	2	1		3		
sigmoid colon	36	4	4	4	4	4							
adrenal gland	26	5	4	4	2	3	2	2	3				3
body of pancreas	32	4	2	2	2	4	2	2	2				
thyroid gland	28	4	4	4	1	3	2	2	3				
stomach	32	3	4	4	4	2							
upper lobe of left lung	29	4	4	4	4								
gastroesophageal sphincter	26	4	4	4	4	2							
spleen	26	5	4	4	4	1							
esophagus muscularis mucosa	25	4	4	4	3	1							
breast epithelium	29	3	3	2		3							
esophagus squamous epithelium	22	4	4	4	4	1							
suprapubic skin	24	4	4	4									
tibial nerve	21	4	4	3	1	1							
Peyer's patch	17	3	4	4	2	1							
omental fat pad	15	4	4	4	1	1							
subcutaneous adipose tissue	16	4	4	3	1	1							
lower leg skin	14	4	4	2			2	2					
heart left ventricle	13	2	2	2		2	2	2					
prostate gland	11	2	2	2	2	1	2	2	1				
vagina	16	2	1	1	1		2	2					
ovary	11	2	2	2	1	1	1	1	1				
testis	7	2	2	2	2	1	2	2	1				
right lobe of liver	10	1	2	2	2	1	1	1					
thoracic aorta	16		2	1		1							
uterus	8	2	2	1	1		2	2	2				
ascending aorta	16	2	1										
right atrium auricular region	8	2	2	2		1	2	2					
tibial artery	7	4				1							
coronary artery	5	2				1							

# Mouse Embryonic Development



# RNA Binding Proteins

- eCLIP for 90 RBPs in K562 and HepG2 cell lines
- Over 70 RNA Bind-n-Seq experiments
- RIP-seq for over 30 RBPs across 3 cell types



**BIOSAMPLE**

**1351 results**

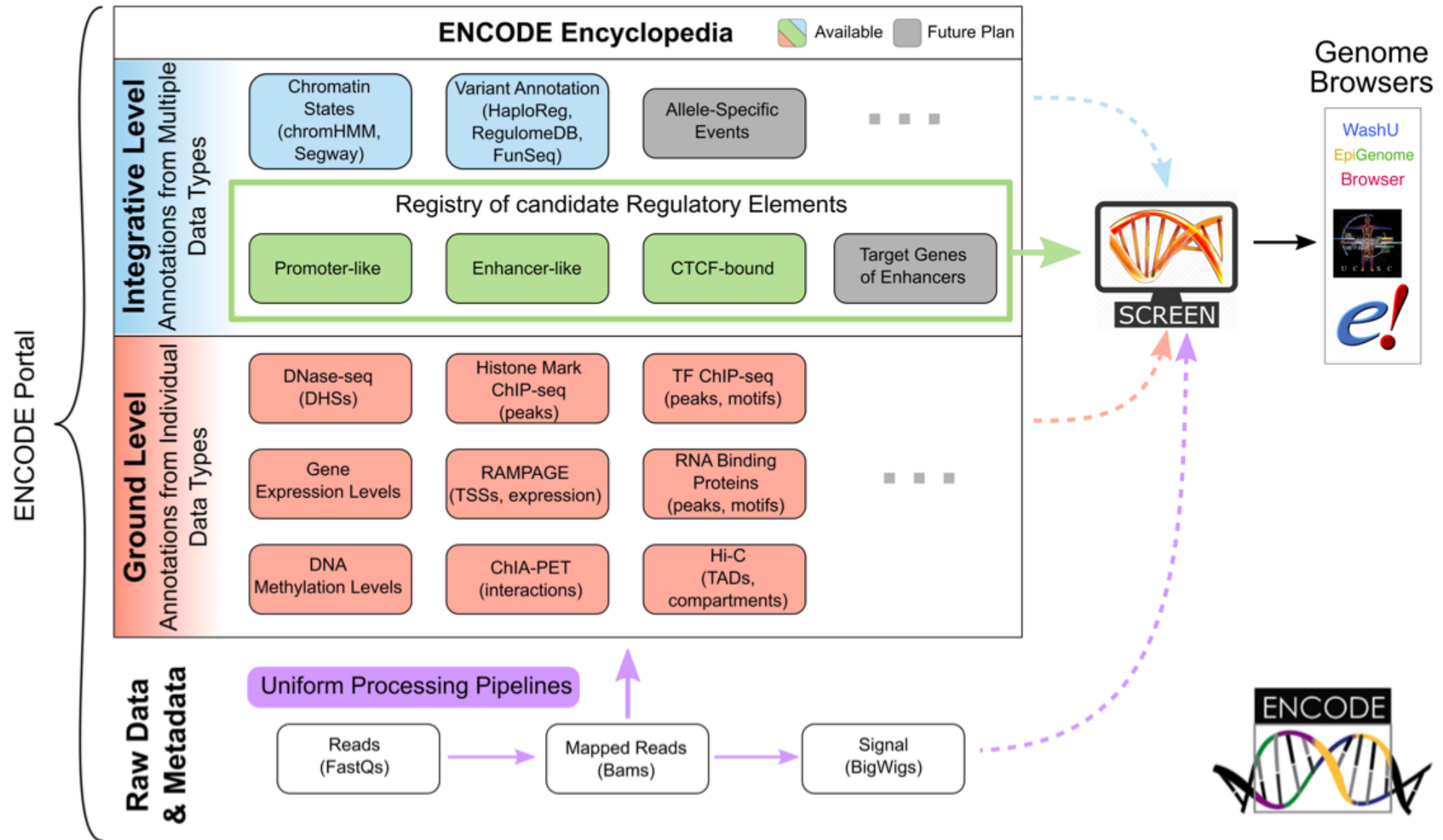
Clear Filters

**ASSAY**

	shRNA RNA-seq	ChIP-seq	eCLIP	RNA Bind-n-Seq	RIP-seq	siRNA RNA-seq	CRISPR RNA-seq	RIP-chip	iCLIP	ChIA-PET	Switchgear
<b>Immortalized cell line</b>											
K562	244	77	90	26	6	11	6	7	1		
HepG2	230	35	74			8	2				
S2-DRSC					57						
GM12878		26		11		7					
A549		31									
<a href="#">+ See 61 more...</a>											
<b>tissue</b>											
transverse colon		7									
adrenal gland		4	2								
gastroesophageal sphincter		6									
body of pancreas		5									
esophagus squamous epithelium		5									
<a href="#">+ See 30 more...</a>											
<b>In vitro sample</b>											
<b>primary cell</b>											
fibroblast of lung		5									
endothelial cell of umbilical vein		4									
keratinocyte		4									
IMR-90		3									
foreskin fibroblast		3									
<a href="#">+ See 31 more...</a>											
<b>stem cell</b>											
H1-hESC		17					1				
E14TG2a.4		1									
<b>In vitro differentiated cells</b>											
bipolar neuron		3									
hepatocyte		2									
myotube		2									
neural cell		2									
<b>Induced pluripotent stem cell line</b>											
GM23338		2									
<a href="#">See all biosamples</a>											



# The ENCODE Encyclopedia



# Disclosure for:

ASHG Interactive Workshop: Navigating the ENCODE Encyclopedia: Exploring Candidate Regulatory Elements, Linked Genes, and Genetic Variation with SCREEN

No Relevant Conflicts to Disclose:

Jill Moore

Zhiping Weng

Michael Purcaro