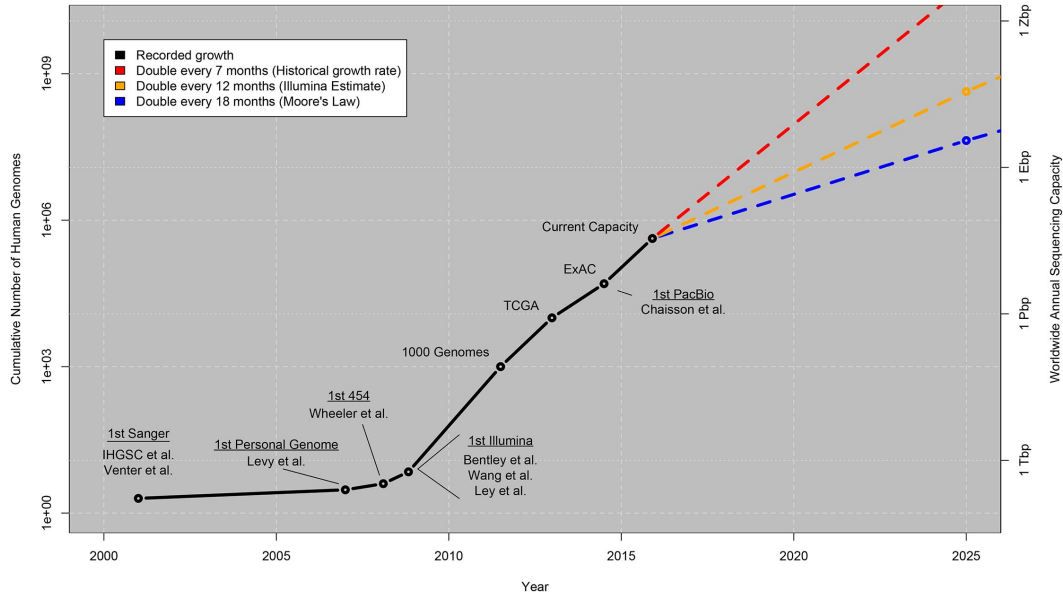


Public resources from Large Scale Omic Projects

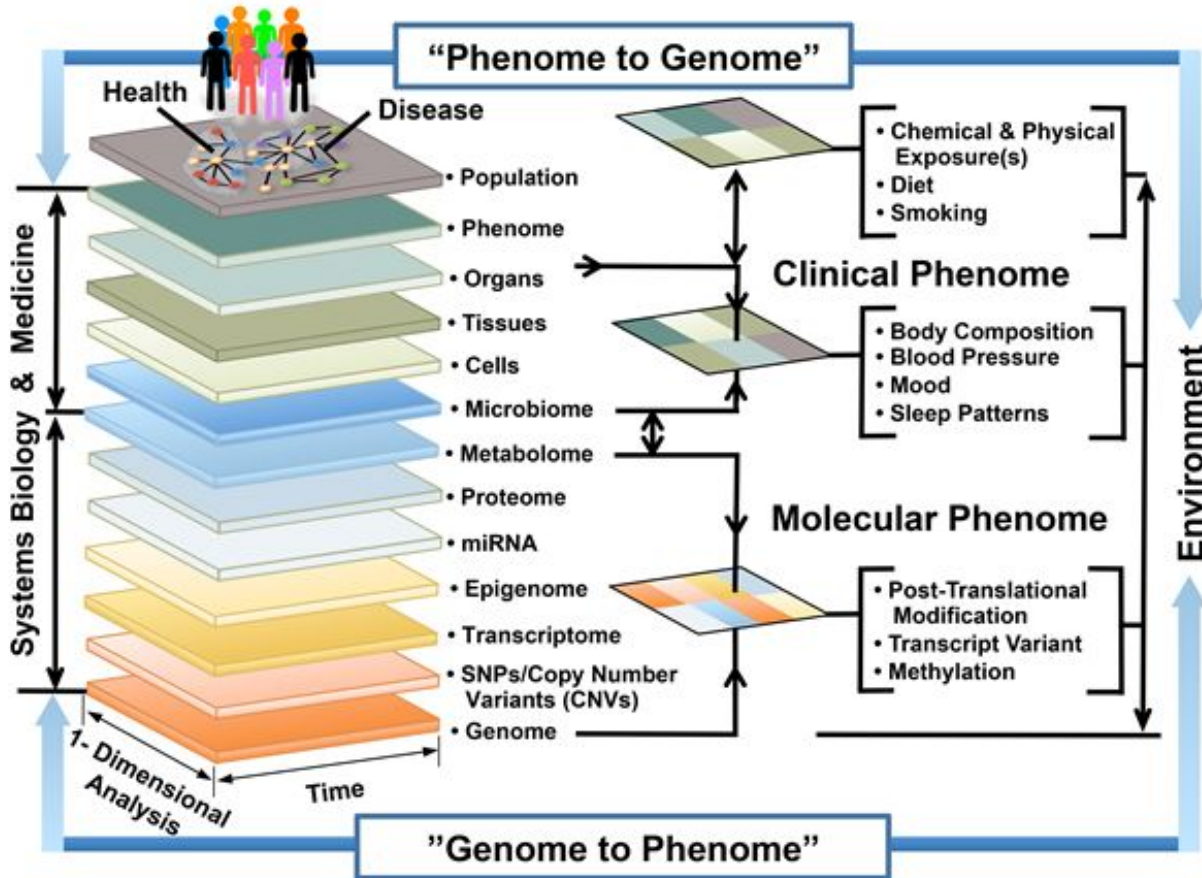
Growth of DNA Sequencing



- Large amount of biological data are currently available for researchers
- How we can use this large amount of information in our research?
- We can use it easily?
- We can integrate these information with other biological knowledges?

From: *PLoS biology*, 13(7), 2015, e1002195.

<https://doi.org/10.1371/journal.pbio.1002195>



- A biological phenotype is the result of the combination of multiple molecular layers
- The contribution of these layers can be measured by high-throughput technologies
- Most of results of these experiments are largely available to researchers through web-based resources and data repositories

Main data types

- **Raw:** Sequencing reads (Fastq), microarray signals (CEL), proteomic spectra (mzML), ...
- **Processed:** Expression tables, ChIP-Seq peaks (BED), lists of genomic variants (VCF), ...

gene	ctrl_1	ctrl_2	exp_1	exp_2
geneA	10	11	56	45
geneB	0	0	128	54
geneC	42	41	59	41
geneD	103	122	1	23
geneE	10	23	14	56
geneF	0	1	2	0
...
...
...

id	treatment	sex
ctrl_1	control	male
ctrl_2	control	female
exp_1	treatment	male
exp_2	treatment	female

Sample names:
ctrl_1, **ctrl_2**, **exp_1**, **exp_2**

- **Integrated:** Genome browsers, interactive web platforms, JAVA graphical interfaces, ...

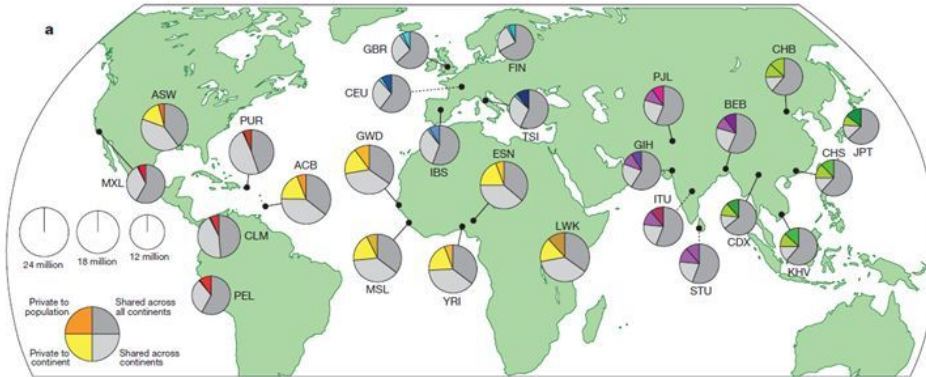
1000 Genomes Project

A global reference for human genetic variation

The 1000 Genomes Project Consortium*

OCTOBER 2015 | VOL 526 | NATURE

Phase III: 2,504 humans : 84.8 million SNPs



- **Level:** Genome
- **Project:** [1000 Genome Project](#)
- **Aim:** Creation of a catalog of human genetic variations
- **Samples:** primary lymphoblastoid cell lines
- **Technologies:** WGS, WES
- **Data:** Raw, processed, integrated

Data access: <http://www.internationalgenome.org/data>

Available data

1000 Genomes Project

1000 Genomes Release	Variants	Individuals	Populations	VCF	Alignments	Supporting Data
Phase 3	84.4 million	2504	26	VCF	Alignments	Supporting Data
Phase 1	37.9 million	1092	14	VCF	Alignments	Supporting Data
Pilot	14.8 million	179	4	VCF	Alignments	Supporting Data

Sequence Alignment Map (SAM) Format

```
@HD VN:1.3 SO:coordinate
@SQ SN:ref LN:45
r001 163 ref 7 30 8M2I4M1D3M = 37 39 TTAGATAAAGGATACTG *
r002 0 ref 9 30 3S6M1P1I4M * 0 0 AAAAGATAAGGATA *
r003 0 ref 9 30 5H6M * 0 0 AGCTAA * NM:i:1
r004 0 ref 16 30 6M14N5M * 0 0 ATAGCTTCAGC *
r003 16 ref 29 30 6H5M * 0 0 TAGGC * NM:i:0
r001 83 ref 37 30 9M = 7 -39 CAGCGCCAT *
```

```
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT N
A00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:
HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:
HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:
HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:
HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTCT G,GTACT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP
0/1:35:4 0/2:17:2 1/1:40:3
```

Variant Call Format (VCF)

Data visualization and integration (e.g. [Ensembl Genome Browser](#))

Variant: rs1333049

rs1333049 SNP

Most severe consequence

[downstream gene variant](#) | [See all predicted consequences](#)

Alleles

[G/C](#) | Ancestral: C | MAF: 0.42 (C) | Highest population MAF: 0.50

Location

[Chromosome 9:22125504](#) (forward strand) | VCF: 9 22125504 rs1333049 G C

Evidence status



HGVS name

[NC_000009.12:g.22125504G>C](#)

Synonyms

This variant has 4 synonyms - [Show](#)

Genotyping chips

This variant has assays on 6 chips - [Show](#)

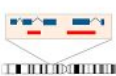
Original source


Variants (including SNPs and indels) imported from dbSNP (release 150) | [View in dbSNP](#)


About this variant


This variant overlaps [13 transcripts](#), has [4039 sample genotypes](#), is associated with [9 phenotypes](#) and is mentioned in [250 citations](#)


Explore this variant


 **Genomic context**


 **Genes and regulation** 13


 **Flanking sequence**
ATTCATT
CGGSGTG
TCATGCT


 **Population genetics** 72

 **Phenotype data** 9

 **Sample genotypes** 4039

 **Linkage disequilibrium**

 **Phylogenetic context**

 **Citations** 250

Using the website

Programmatic access

Population	Allele: frequency (count)
ALL	G: 0.582 (2914) C: 0.418 (2094)
AFR	G: 0.787 (1040) C: 0.213 (282)
ACB	G: 0.760 (146) C: 0.240 (46)
ASW	G: 0.730 (89) C: 0.270 (33)
ESN	G: 0.783 (155) C: 0.217 (43)
GWD	G: 0.876 (198) C: 0.124 (28)
LWK	G: 0.737 (146) C: 0.263 (52)
MSL	G: 0.800 (136) C: 0.200 (34)
YRI	G: 0.787 (170) C: 0.213 (46)
AMR	G: 0.545 (378) C: 0.455 (316)
CLM	G: 0.553 (104) C: 0.447 (84)
MXL	G: 0.508 (65) C: 0.492 (63)
PEL	G: 0.635 (108) C: 0.365 (62)
PUR	G: 0.486 (101) C: 0.514 (107)
EAS	G: 0.463 (467) C: 0.537 (541)
CDX	G: 0.425 (79) C: 0.575 (107)
CHB	G: 0.524 (108) C: 0.476 (98)
CHS	G: 0.486 (102) C: 0.514 (108)
JPT	G: 0.481 (100) C: 0.519 (108)
KHV	G: 0.394 (78) C: 0.606 (120)
EUR	G: 0.528 (531) C: 0.472 (475)
CEU	G: 0.556 (110) C: 0.444 (88)
FIN	G: 0.601 (119) C: 0.399 (79)
GBR	G: 0.505 (92) C: 0.495 (90)
IBS	G: 0.500 (107) C: 0.500 (107)
TSI	G: 0.481 (103) C: 0.519 (111)
SAS	G: 0.509 (498) C: 0.491 (480)

Applications of the 1000 Genomes Project resources

Xiangqun Zheng-Bradley, Paul Flicek

Briefings in Functional Genomics, Volume 16, Issue 3, 1 May 2017, Pages 163–170,
<https://doi.org/10.1093/bfgp/elw027>

Published: 19 July 2016

Abstract

The 1000 Genomes Project created a valuable, worldwide reference for human genetic variation. Common uses of the 1000 Genomes dataset include genotype imputation supporting Genome-wide Association Studies, mapping expression Quantitative Trait Loci, filtering non-pathogenic variants from exome, whole genome and cancer genome sequencing projects, and genetic analysis of population structure and molecular evolution. In this article, we will highlight some of the multiple ways that the 1000 Genomes data can be and has been utilized for genetic studies.

Brief. in Functional Genomics, 16(3), 2017, 163-70,
<https://doi.org/10.1093/bfgp/elw027>

What else?

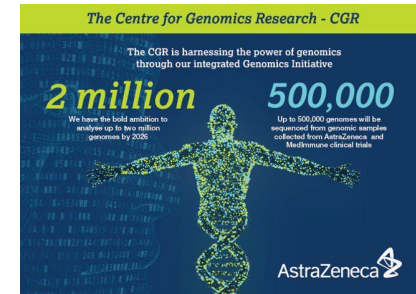
- [Exome Aggregation Consortium](#) (ExAC)
- [European Genome-phenome Archive](#) (EGA)
- [The database of Genotypes and Phenotypes](#) (dbGaP)

- [10K Genomes project](#) (Non human)
- [1000 Plant Genomes Project](#) (Non human)

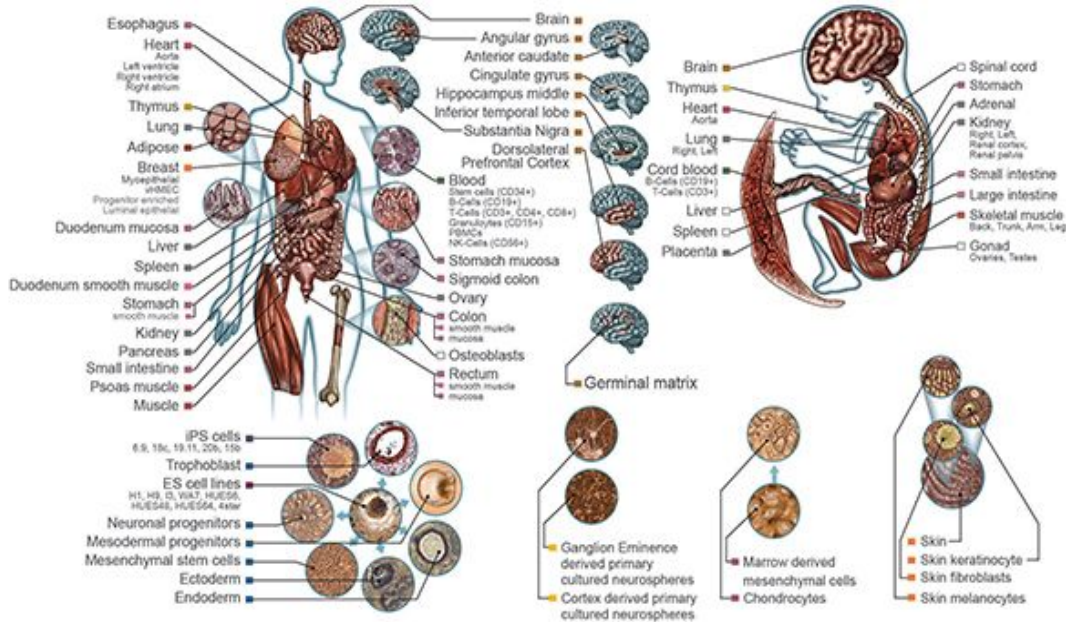
What's next?



[100,000 Genomes Project](#)



[AstraZeneca 2 million genomes](#)



- **Level:** Epigenome
- **Project:** [Roadmap Epigenomics](#)
- **Aim:** Characterization of healthy human tissue epigenomes
- **Samples:** tissues / cell types
- **Technologies:** DNA methylation, Histone modification ChIP-Seq, DNase-Seq, and RNA-Seq
- **Data:** Raw, processed, integrated

Data access: http://egg2.wustl.edu/roadmap/web_portal/



Home /
Grid

Metadata

Processed
Data ▾

Imputed
Data ▾

Chromatin
State
Learning ▾

Epigenomes
Clustering ▾

DNaseI-accessible
Regulatory
Regions ▾

Predicting
Regulators
and Motifs ▾

Disease
Variant
Interpretation ▾

Processed data for single dataset:

- Cell line, dataset Metadata
- Quality control results
- Read Alignments
- Signal peaks
- Genome-wide signal coverage
- Expression quantification

Integrated data from multiple datasets

- Chromatin states
- Comparison between epigenomes
- Predicted promoters and enhancers
- Tissue-specific regulators
- Prediction of disease variant functional effect

Data visualization and integration

Processed data for single dataset / tissue



Integrated data from multiple datasets / tissues



Integrative analysis of 111 reference human epigenomes

Roadmap Epigenomics Consortium†, Anshul Kundaje^{1,2,3*}, Wouter Meuleman^{1,2*}, Jason Ernst^{1,2,4*}, Misha Bilenky^{5*}, Angela Yen^{1,2}, Alireza Heravi-Moussavi⁵, Pouya Kheradpour^{1,2}, Zhizhuo Zhang^{1,2}, Jianrong Wang^{1,2}, Michael J. Ziller^{2,6}, Viren Amin⁷, John W. Whitaker⁸, Matthew D. Schultz⁹, Lucas D. Ward^{1,2}, Abhishek Sarkar^{1,2}, Gerald Quon^{1,2}, Richard S. Sandstrom¹⁰, Matthew L. Eaton^{1,2}, Yi-Chieh Wu^{1,2}, Andreas R. Pfenning^{1,2}, Xinchun Wang^{1,2,11}, Melina Claussnitzer^{1,2}, Yaping Liu^{1,2}, Cristian Coarfa⁷, R. Alan Harris⁷, Noam Shores², Charles B. Epstein², Elizabeta Gjoneska^{2,12}, Danny Leung^{8,13}, Wei Xie^{8,13}, R. David Hawkins^{8,13}, Ryan Lister⁹, Chibo Hong¹⁴, Philippe Gascard¹⁵, Andrew J. Mungall⁵, Richard Moore⁵, Eric Chuah⁵, Angela Tam⁵, Theresa K. Canfield¹⁰, R. Scott Hansen¹⁶, Rajinder Kaul¹⁶, Peter J. Sabo¹⁰, Mukul S. Bansal^{1,2,17}, Annaick Carles¹⁸, Jesse R. Dixon^{8,13}, Kai-How Farh², Soheil Feizi^{1,2}, Rosa Karlic¹⁹, Ah-Ram Kim^{1,2}, Ashwinikumar Kulkarni²⁰, Daofeng Li²¹, Rebecca Lowdon²¹, GiNell Elliott²¹, Tim R. Mercer²², Shane J. Neph¹⁰, Vitor Onuchic⁷, Paz Polak^{2,23}, Nisha Rajagopal^{8,13}, Pradipta Ray²⁰, Richard C. Sallari^{1,2}, Kyle T. Siebenthal¹⁰, Nicholas A. Sinnich-Armstrong^{1,2}, Michael Stevens^{21,42}, Robert E. Thurman¹⁰, Jie Wu^{24,25}, Bo Zhang²¹, Xin Zhou²¹, Arthur E. Beaudet²⁶, Laurie A. Boyer¹¹, Philip L. De Jager^{2,23,27}, Peggy J. Farnham²⁸, Susan J. Fisher²⁹, David Haussler³⁰, Steven J. M. Jones^{5,31,32}, Wei Li³³, Marco A. Marra^{5,32}, Michael T. McManus³⁴, Shamil Sunyaev^{2,23,27}, James A. Thomson^{35,41}, Thea D. Tlsty¹⁵, Li-Huei Tsai^{2,12}, Wei Wang⁸, Robert A. Waterland²⁰, Michael Q. Zhang^{20,37}, Lisa H. Chadwick³⁸, Bradley E. Bernstein^{2,39,40§}, Joseph F. Costello^{14§}, Joseph R. Ecker^{8§}, Martin Hirst^{3,18§}, Alexander Meissner^{2,10§}, Aleksandar Milosavljevic^{7§}, Bing Ren^{8,13§}, John A. Stamatoyannopoulos^{10§}, Ting Wang^{21§} & Manolis Kellis^{1,2§}

The reference human genome sequence set the stage for studies of genetic variation and its association with human disease, but epigenomic studies lack a similar reference. To address this need, the NIH Roadmap Epigenomics Consortium generated the largest collection so far of human epigenomes for primary cells and tissues. Here we describe the integrative analysis of 111 reference human epigenomes generated as part of the programme, profiled for histone modification patterns, DNA accessibility, DNA methylation and RNA expression. We establish global maps of regulatory elements, define regulatory modules of coordinated activity, and their likely activators and repressors. We show that disease- and trait-associated genetic variants are enriched in tissue-specific epigenomic marks, revealing biologically relevant cell types for diverse human traits, and providing a resource for interpreting the molecular basis of human disease. Our results demonstrate the central role of epigenomic information for understanding gene regulation, cellular differentiation and human disease.

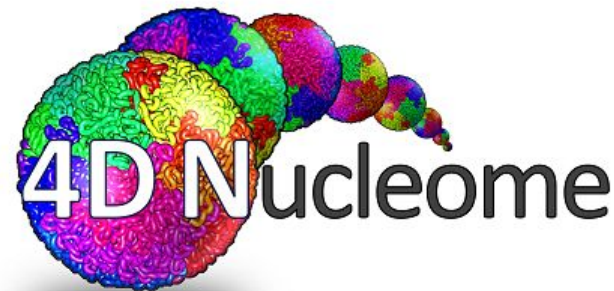
Nature, 518, 2015, 317-330,

<http://doi.org/10.1038/nature14248>

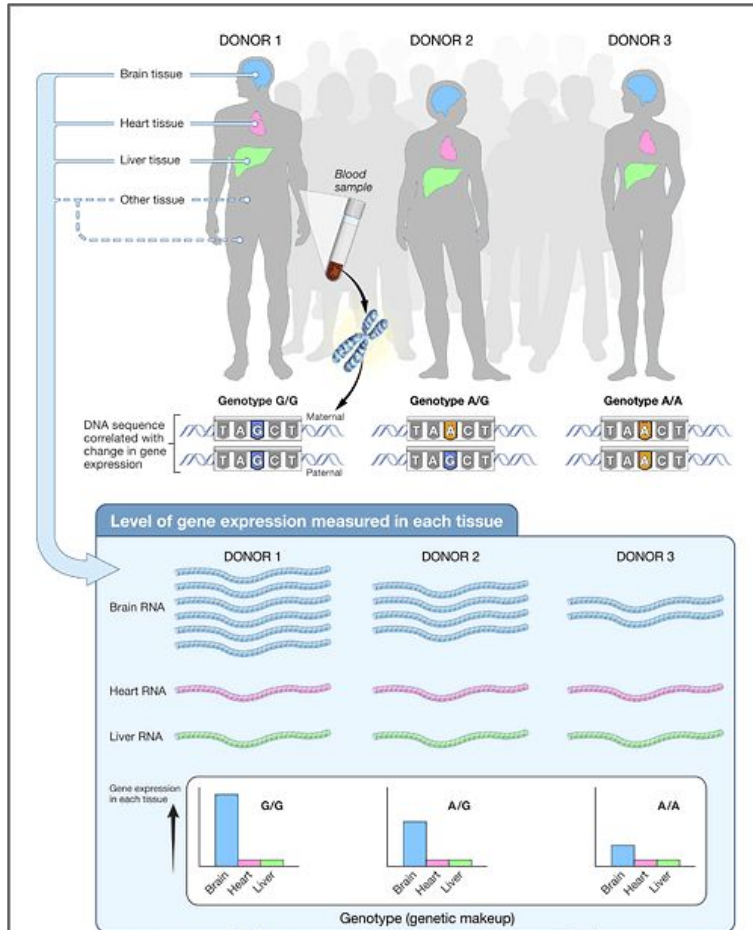
What else?

- [Blueprint](#) > ([Cell special issue](#))
- [CEEHRC](#)
- [modENCODE](#) (Non-human model organisms)

What's next?



[NIH 4D Nucleome](#)



- **Level:** Transcriptome
- **Project:** [GTEx](#)
- **Aim:** Large scale analysis of expression Quantitative Traits Loci (eQTL)
- **Samples:** Post-mortem tissues
- **Technologies:** WES, WGS, SNP microarray, RNA-Seq
- **Data:** Processed and integrated

ARTICLE

OPEN

doi:10.1038/nature24277

Genetic effects on gene expression across human tissues

GTEx Consortium*

Characterization of the molecular function of the human genome and its variation across individuals is essential for identifying the cellular mechanisms that underlie human genetic traits and diseases. The Genotype-Tissue Expression (GTEx) project aims to characterize variation in gene expression levels across individuals and diverse tissues of the human body, many of which are not easily accessible. Here we describe genetic effects on gene expression levels across 44 human tissues. We find that local genetic variation affects gene expression levels for the majority of genes, and we further identify inter-chromosomal genetic effects for 93 genes and 112 loci. On the basis of the identified genetic effects, we characterize patterns of tissue specificity, compare local and distal effects, and evaluate the functional properties of the genetic effects. We also demonstrate that multi-tissue, multi-individual data can be used to identify genes and pathways affected by human disease-associated variation, enabling a mechanistic interpretation of gene regulation and the genetic basis of disease.

Nature, 550, 2017, 204-213.

<http://doi.org/10.1038/nature24277>

What else?

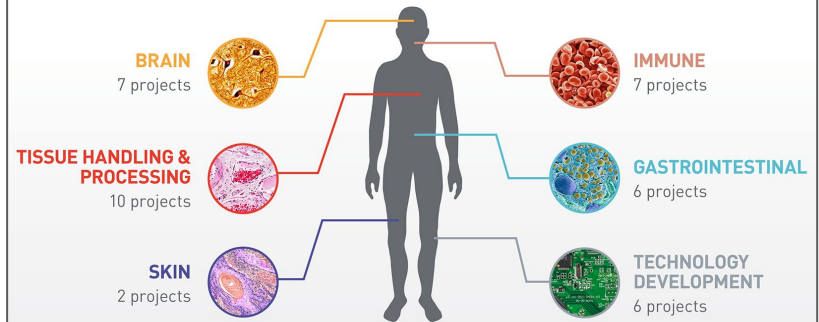
- [FANTOM5](#) ([Nature Collection paper](#))
- [Human Developmental Biology Resource](#) (HDBR)

What's next?



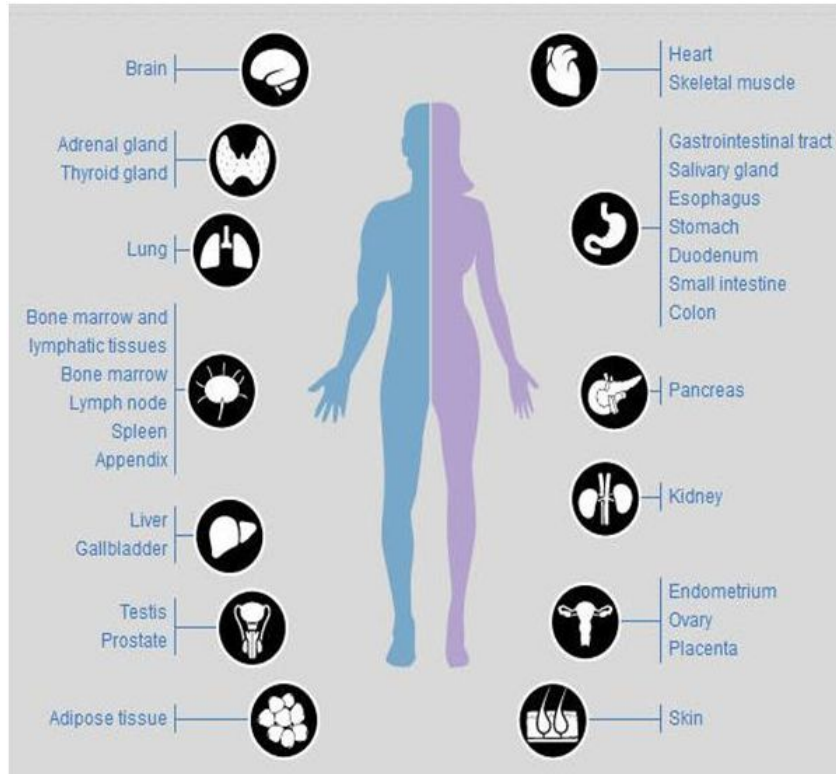
MAPPING THE BASIC UNITS OF LIFE

CZI proudly supports **38 new projects** in these six areas for the Human Cell Atlas.



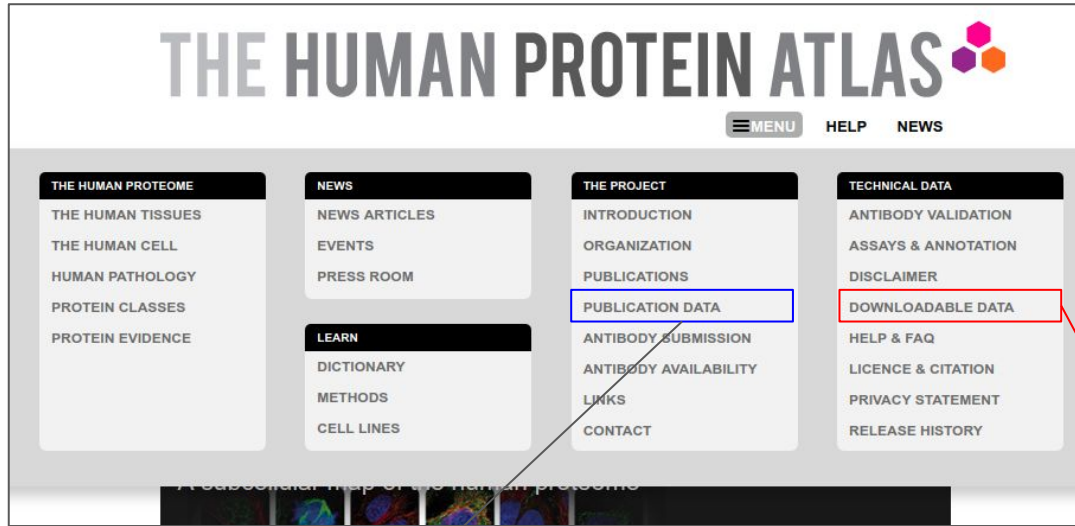
[Human Cell Atlas](#)

THE HUMAN PROTEIN ATLAS



- **Level:** Proteome
- **Project:** [The Human Protein Atlas](#)
- **Aim:** Analysis of protein tissue expression and localization
- **Samples:** Tissues, cell lines
- **Technologies:** Protein array, IHC, Immunofluorescence, Western blot, RNA-Seq,
- **Data:** Processed and integrated

Data access



Publication Data

- Antibody response
- Protein array expression data - Tissue
- Protein array expression data - Cells
- Immunofluorescence expression data
- RNA isoform expression data

Downloadable Data

- Normal tissue data (level and reliability)
- Pathology data (level, reliability, correlation with survival)
- Subcellular localization data (reliability, level, between cell variations, cell cycle dependency)
- RNA-Seq data (gene/isoform level in TPM)

Data visualization and integration


THE HUMAN PROTEIN ATLAS

MENU HELP NEWS

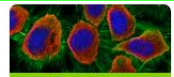
SEARCH¹

e.g. RBM3, insulin, CD36

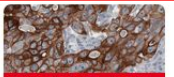
Search Fields >



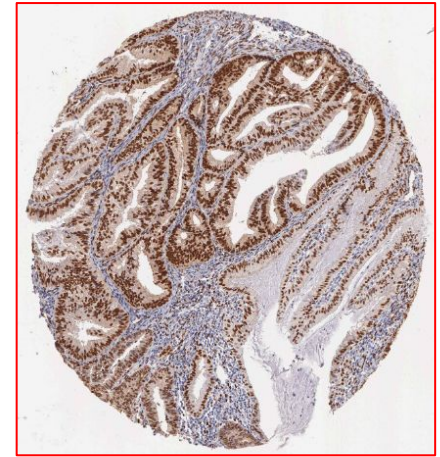
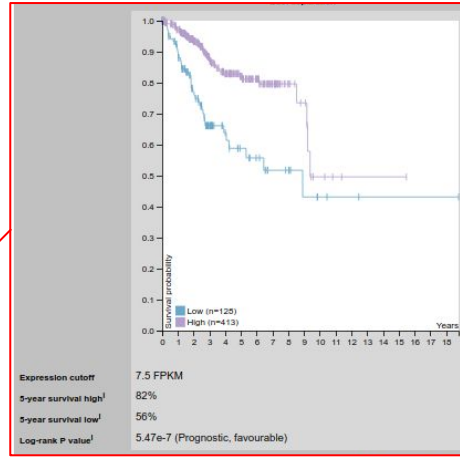
TISSUE ATLAS



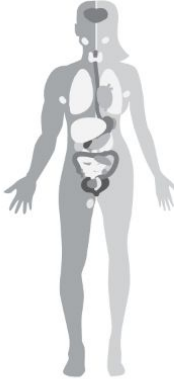
CELL ATLAS



PATHOLOGY ATLAS



RNA expression (TPM)¹ Protein expression (score)¹



Organ System	RNA expression (TPM)	Protein expression (score)	Thumbnail
Brain	Low	Low	Cerebral cortex
Endocrine tissues	Low	Low	
Bone marrow & immune system	Low	Low	Lymph node
Muscle tissues	Low	Low	
Lung	Low	Low	Liver
Liver & gallbladder	Low	Low	
Pancreas	Low	Low	Lung
Gastrointestinal tract	Low	Low	Kidney
Kidney & urinary bladder	Low	Low	
Male tissues	Low	Low	Kidney
Female tissues	Low	Low	Testis
Adipose & soft tissue	Low	Low	
Skin	Low	Low	Endometrium

GENERAL INFORMATION

Gene name¹: ESR1

Gene description¹: Estrogen receptor 1

Protein class¹: Cancer-related genes, Disease related genes, FDA approved drug targets, Nuclear receptors, Predicted intracellular proteins, Transcription factors

Predicted localization¹: Intracellular

Number of transcripts¹: 11

[SHOW MORE](#)

HUMAN PROTEIN ATLAS INFORMATION

Summary¹: Mainly localized to vesicles. In addition localized to the nucleus.

RNA cell category¹: Group enriched (T-47d, MCF7)

Protein evidence¹: Evidence at protein level

Main location¹: Localized to the Vesicles (approved) [View proteome in REACTOME](#)

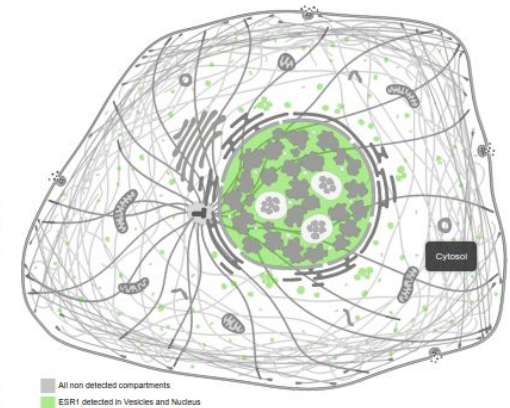
Additional location¹: In addition localized to the Nucleus (approved) [View proteome in REACTOME](#)

DATA RELIABILITY

Reliability score¹: ● Approved

Antibodies¹: HPA000449

[SHOW MORE](#)



RESEARCH ARTICLE SUMMARY

PROTEOMICS

Tissue-based map of the human proteome

Mathias Uhlén,* Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjödéd, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szgyarto, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per-Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M. Schwenk, Marica Hamsten, Kalle von Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar von Heijne, Jens Nielsen, Fredrik Pontén

Science. 2015, 347(6220), 1260419.
<http://doi.org/10.1038/nature24277>

What else?

- [Human Proteome Map](#)
- [Human Metabolome Project](#)
- [Clinical Proteomic Tumor Analysis Consortium](#) (CPTAC)

What's next?



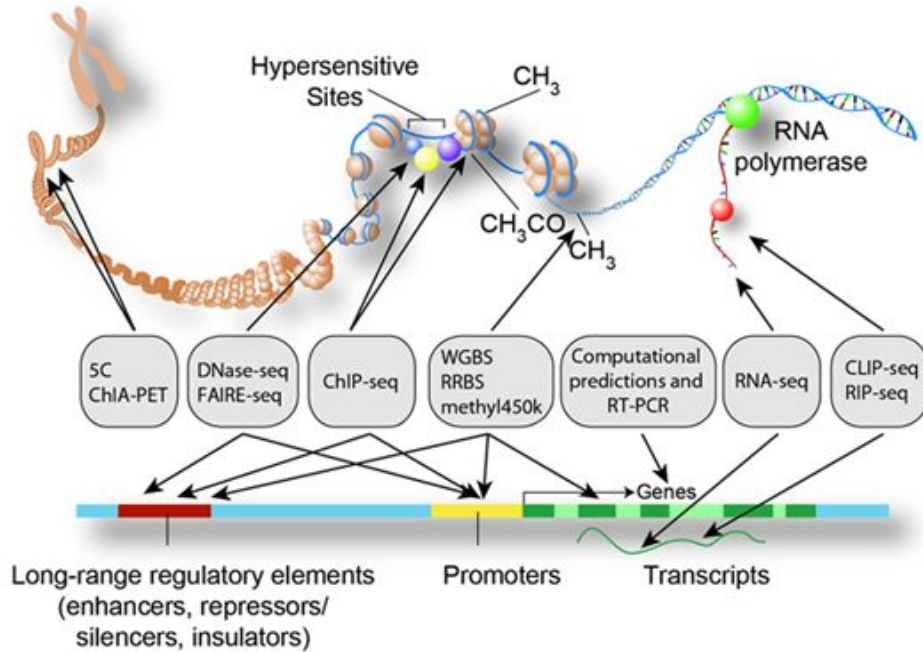
CHP

Chromosome-based Human Proteome Project
(C-HPP)

B/D HPP

Biology/Disease Human Proteome Project
(B/D-HPP)

ENCODE



[Nature Collection](#)

- **Level:** Multi omics - cellular level
- **Project:** [ENCODE](#)
- **Aim:** Functional annotation of human genome
- **Samples:** Cell lines, primary cells, tissue (non human), whole organism
- **Technologies:** ChIP-Seq, RNA-Seq, CLIP/RIP-Seq, WGBS, RRBS, DNase-seq, ChIA-PET, 5C, FAIRE-Seq, Repli-Seq, WGS, Hi-C
- **Data:** Raw, processed, integrated

Data access: <https://www.encodeproject.org/matrix/?type=Experiment>

Experiment Matrix

Click or enter search terms to filter the experiments included in the matrix.

Enter search term(s)

Organism

<i>Homo sapiens</i>	10847
<i>Mus musculus</i>	1863
<i>Drosophila melanogaster</i>	1435
<i>Caenorhabditis elegans</i>	974
<i>Drosophila pseudoobscura</i>	12

+ See more...

Biosample type

cell line	5746
tissue	4373
whole organisms	2051
primary cell	1674
in vitro differentiated cells	676

+ See more...

Organ

blood	2608
bodily fluid	2608
liver	1177
embryo	982
brain	935

+ See more...

Project

ENCODE	9657
--------	------

Assay	Assay category	Target of assay	Date released	Available data
ChIP-seq 8908	DNA binding 8908	transcription factor 4024	July, 2013 2765	fastq 13700
DNase-seq 863	Transcription 3331	histone 3102	March, 2014 887	bam 12733
polyA RNA-seq 817	DNA accessibility 1117	histone modification 3102	July, 2016 614	bigWig 11688
shRNA RNA-seq 533	DNA methylation 865	control 2568	May, 2016 569	bed narrowPeak 7298
total RNA-seq 413	RNA binding 630	broad histone mark 1727	October, 2016 485	bigBed narrowPeak 7287

15321 results

BIOSAMPLE	ASSAY																				
	ChIP-seq	DNase-seq	polyA RNA-seq	shRNA RNA-seq	total RNA-seq	eCLIP	RNA microarray	DNase array	WGBS	RRBS	small RNA-seq	microRNA-seq	ATA-seq	RAMPAGE	RNA Bind-n-Seq	genotyping array	single cell RNA-seq	Replic-seq	microRNA counts	...and 26 more	
cell line																					
K562	698	4	10	276	12	190	12	3	1	1	8	1	1	1	2	9	6	1			
HepG2	374	3	11	257	6	161	7	3	2	2	3				2	6	6	1			
A549	384	14	27				2	2	1	1	9				2	3	2				
GM12878	249	3	14	4			8	3	2	2	6	2	1		2	6	13	6	1		
HEK293	257						1	2	2						2						
tissue																					
liver	162	9	22	3			1	11	1	1	7	7	2	3	2	7					
heart	100	22	16	3	10	10	1	1	1	9	7	2			1	8					
stomach	98	21	15	5			3	10	1	4	4	6	5	1		4					
lung	80	16	12	1	10	2	8	3	1	4	4	1				4					
kidney	69	17	13				2	2	5	4	4	4				4					
whole organisms																					
whole organism	1580			73	50											15					
multi-cellular organism	291																				
carcass				12	4											4					

+ See 196 more...

+ See 160 more...

Clear Filters

Assay category

DNA binding 698

Assay

ChIP-seq 698

shRNA RNA-seq	276
eCLIP	190
CRISPRi RNA-seq	77
siRNA RNA-seq	50

+ See more...

Project

ENCODE 698

RFA

ENCODE3	478
ENCODE2	220

Experiment status

released	669
revoked	26
archived	3

Genome assembly (visualization)

hg19	519
GRCh38	491

Organism

Homo sapiens 697

Target of assay

transcription factor 543

Showing 25 of 698 results

View All

ChIP-seq of K562

Homo sapiens K562

Target: MTA1
Lab: Michael Snyder, Stanford
Project: ENCODE

ChIP-seq of K562

Homo sapiens K562

Target: AFF1
Lab: Michael Snyder, Stanford
Project: ENCODE

ChIP-seq of K562

Homo sapiens K562

Target: AFF1
Lab: Michael Snyder, Stanford
Project: ENCODE

ChIP-seq of K562

Homo sapiens K562

Target: MNT
Lab: Michael Snyder, Stanford
Project: ENCODE

ChIP-seq of K562

Data access: <https://www.encodeproject.org/matrix/?type=Experiment>

Summary		Attribution	
Assay:	ChIP-seq	Lab:	Michael Snyder, Stanford
Target:	MTA1	Award:	U54HG006996 (Michael Snyder, Stanford)
Biosample summary:	<i>Homo sapiens</i> K562	Project:	ENCODE
Biosample Type:	cell line	External resources:	GEO:GSE105823
Replication type:	isogenic	Aliases:	michael-snyder:ChIPss-812
Description:	MTA1 ChIP-seq on human K562	Date submitted:	May 2, 2017
Nucleic acid type:	DNA	Date released:	May 17, 2017
Size range:	450-650		
Strand specificity:	Non-strand-specific		
Platform:	Illumina HiSeq 2000		
Controls:	ENCSR173USI		



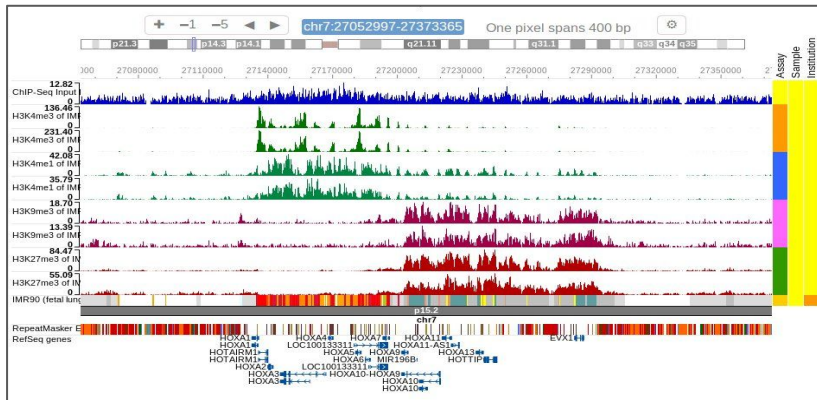
Supplementary file	Size	Download	File type/resource
GSE105823_ENCFF002ZRH_peaks_GRCh38.bed.gz	6.9 Mb	(ftp) (http)	BED
GSE105823_ENCFF056TVE_peaks_hg19.bigBed	12.4 Mb	(ftp) (http)	BIGBED
GSE105823_ENCFF132APO_fold_change_over_control_GRCh38.bigWig	1.9 Gb	(ftp) (http)	BIGWIG
GSE105823_ENCFF161QXX_optimal_idr_thresholded_peaks_GRCh38.bigBed	683.5 Kb	(ftp) (http)	BIGBED
GSE105823_ENCFF180XEU_conservative_idr_thresholded_peaks_hg19.bigBed	469.6 Kb	(ftp) (http)	BIGBED
GSE105823_ENCFF194ZGF_fold_change_over_control_hg19.bigWig	1.9 Gb	(ftp) (http)	BIGWIG
GSE105823_ENCFF334TNK_peaks_GRCh38.bigBed	12.4 Mb	(ftp) (http)	BIGBED
GSE105823_ENCFF362MSS_conservative_idr_thresholded_peaks_GRCh38.bigBed	470.0 Kb	(ftp) (http)	BIGBED
GSE105823_ENCFF407VAS_conservative_idr_thresholded_peaks_hg19.bed.gz	195.2 Kb	(ftp) (http)	BED
GSE105823_ENCFF433GGH_signal_p-value_GRCh38.bigWig	2.0 Gb	(ftp) (http)	BIGWIG
GSE105823_ENCFF445RPZ_signal_p-value_hg19.bigWig	1.8 Gb	(ftp) (http)	BIGWIG
GSE105823_ENCFF658ZHZ_peaks_hg19.bed.gz	6.9 Mb	(ftp) (http)	BED
GSE105823_ENCFF769GYG_optimal_idr_thresholded_peaks_hg19.bed.gz	326.6 Kb	(ftp) (http)	BED
GSE105823_ENCFF801KEW_optimal_idr_thresholded_peaks_GRCh38.bed.gz	327.8 Kb	(ftp) (http)	BED
GSE105823_ENCFF843MEU_conservative_idr_thresholded_peaks_GRCh38.bed.gz	193.9 Kb	(ftp) (http)	BED
GSE105823_ENCFF845IDA_optimal_idr_thresholded_peaks_hg19.bigBed	677.8 Kb	(ftp) (http)	BIGBED
GSE105823_ENCFF880QV_peaks_GRCh38.bed.gz	6.0 Mb	(ftp) (http)	BED
GSE105823/suppl/GSE105823_ENCFF334TNK_peaks_GRCh38.bigBed			

Available files:

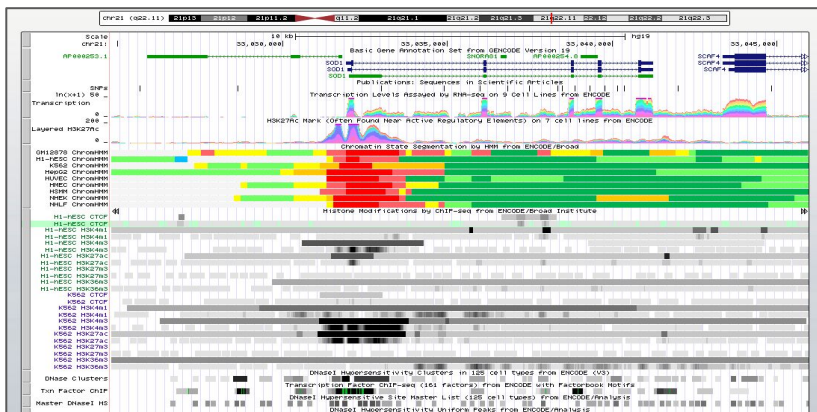
- Experiment protocols and QC
- Raw reads (Fastq)
- Alignment files (BED, SAM, BAM)
- Genomic coverage files (BIGWIG)
- Peaks coordinates (BED, BIGBED)

Data visualization and integration

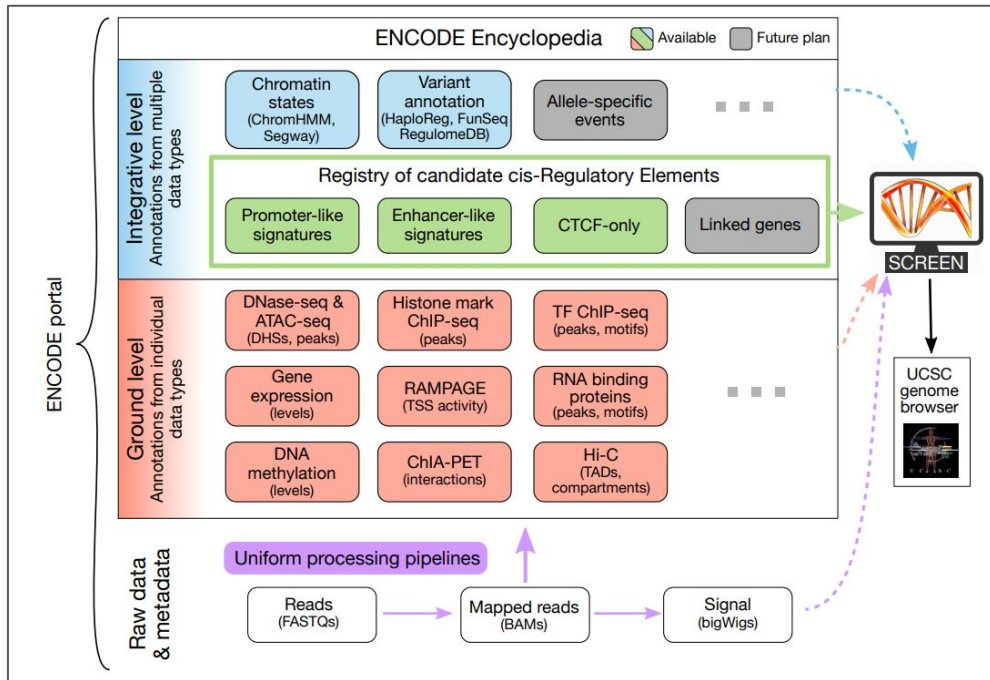
WashU Genome Browser



UCSC Genome Browser



Search Candidate cis-Regulatory Elements (SCREEN)



Review

Deciphering ENCODE

Adam G. Diehl¹ and Alan P. Boyle^{1,2,*}

The ENCODE project represents a major leap from merely describing and comparing genomic sequences to surveying them for direct indicators of function. The astounding quantity of data produced by the ENCODE consortium can serve as a map to locate specific landmarks, guide hypothesis generation, and lead us to principles and mechanisms underlying genome biology. Despite its broad appeal, the size and complexity of the repository can be intimidating to prospective users. We present here some background about the ENCODE data, survey the resources available for accessing them, and describe a few simple principles to help prospective users choose the data type(s) that best suit their needs, where to get them, and how to use them to their best advantage.

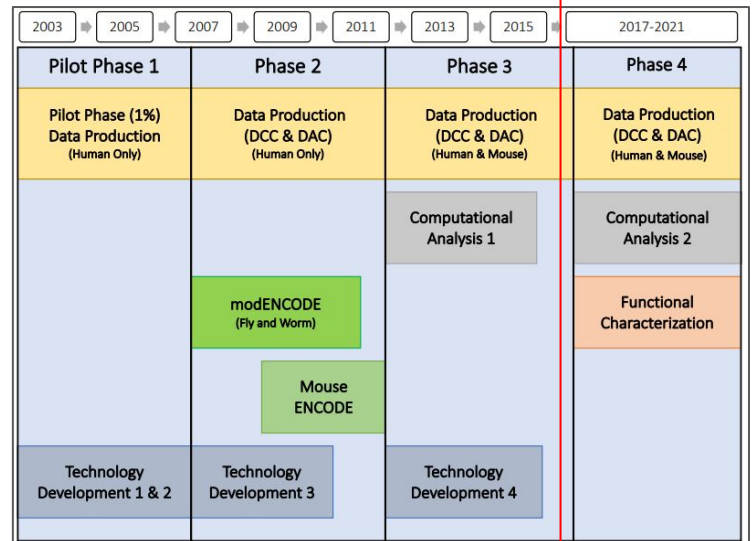
Trends Genet. 2016, 32(4), 238-249.

<https://doi.org/10.1016/j.tig.2016.02.002>

What else?

- [Cancer Cell Line Encyclopedia](#) (CCLE)
- [BD2K-LINCS project](#)
- [NCI60](#)
- [GDSC](#)
- Project [Achilles](#) and [DRIVE](#)

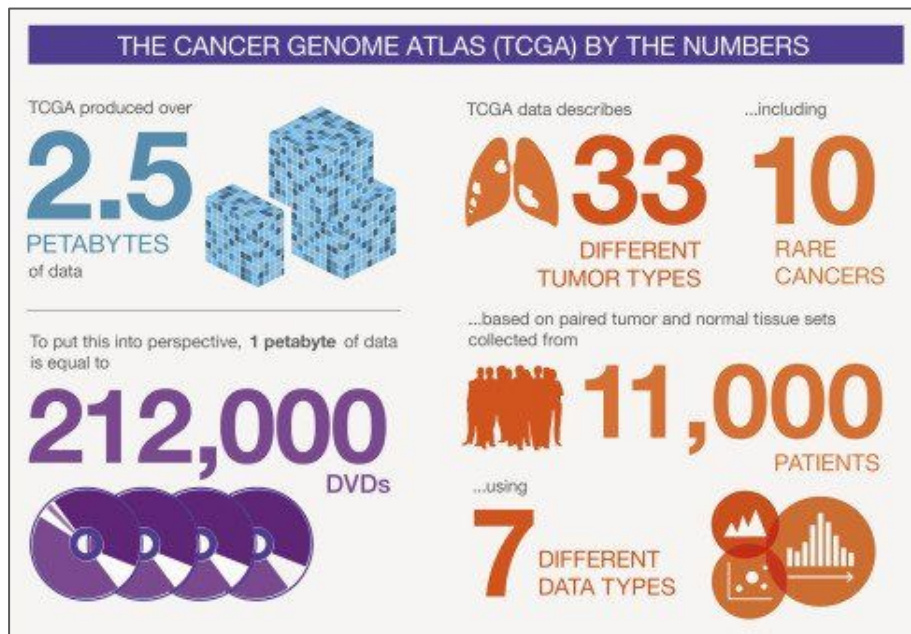
What's next?





THE CANCER GENOME ATLAS

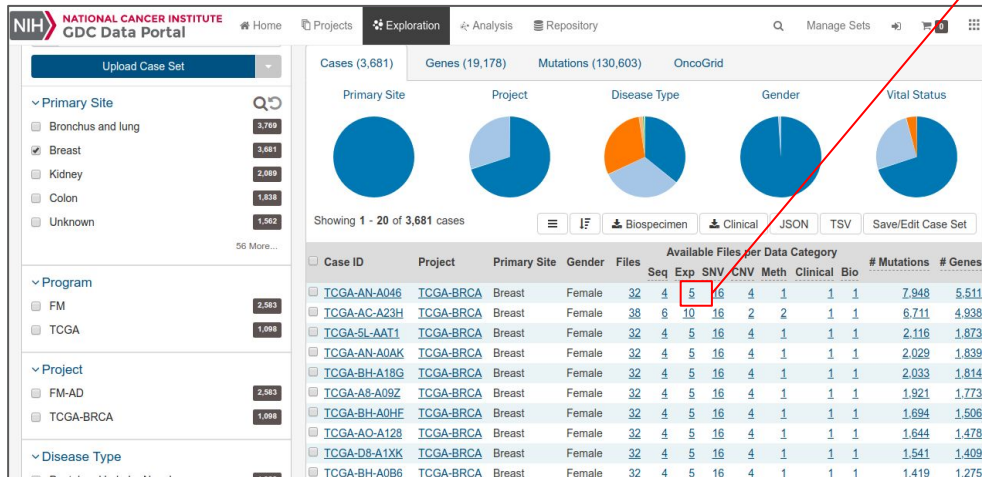
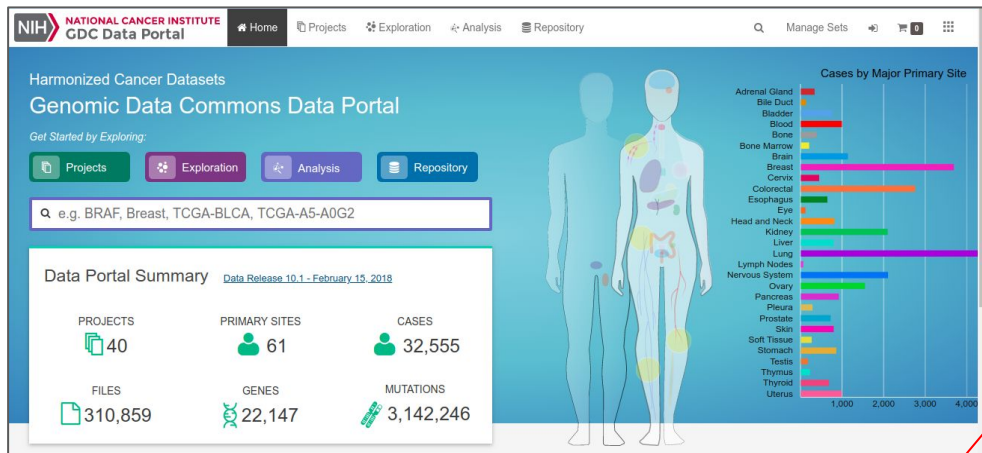
National Cancer Institute
National Human Genome Research Institute



[Publication list](#)

- **Level:** Multi omics - tissue level
- **Project:** TCGA
- **Aim:** Molecular characterization and classification of tumors and their vulnerability
- **Samples:** Healthy and primary cancer tissue samples
- **Technologies:** WES, WGS, RNA-Seq, protein array, DNA Methylation array
- **Data:** Processed and integrated

Data access: <https://portal.gdc.cancer.gov/>



Showing 1 - 5 of 5 files

Access	File Name	Cases	Project	Data Category	Data Format	File Size	Annotations
open	c98b8334-836a-4c3c-b149-b08a7256aa68.FPKM-UQ.txt.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	526.18 KB	0
open	c98b8334-836a-4c3c-b149-b08a7256aa68.FPKM.txt.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	524.2 KB	0
open	c98b8334-836a-4c3c-b149-b08a7256aa68.htseq_counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	254.53 KB	0
open	5002ba7d-8811-4141-89b8-d9487b533430_mirbase21_mirnas_quantificatio_n.txt	1	TCGA-BRCA	Transcriptome Profiling	TSV	50.08 KB	0
open	5002ba7d-8811-4141-89b8-d9487b533430_mirbase21_isoforms_quantificatio_n.txt	1	TCGA-BRCA	Transcriptome Profiling	TSV	217.36 KB	0

Show 20 entries

Available files:

- Gene / miRNA expression quantification
- VCF and MAF file
- CNV data
- Methylation data
- Protein level data
- Clinical data

Data visualization and integration

cBioPortal

Pancreas

Acinar Cell Carcinoma of the Pancreas

Acinar Cell Carcinoma of the Pancreas (Johns Hopkins, J Pathol 2014)

Cystic Tumor of the Pancreas

Cystic Tumor of the Pancreas (Johns Hopkins, PNAS 2011)

Pancreatic Adenocarcinoma

Pancreatic Adenocarcinoma (ICGC, Nature 2012)

Pancreatic Adenocarcinoma (OCMG, Nature 2016)

Pancreatic Adenocarcinoma (TCGA, Provisional)

Pancreatic Cancer (UTSW, Nat Commun 2015)

Pancreatic Neuroendocrine Tumor

Insulinoma (Shanghai, Nat Commun 2013)

Pancreatic Neuroendocrine Tumors (Johns Hopkins University, Science 2011)

Whole-Genome Sequencing of Pancreatic Neuroendocrine Tumors (Nature, 2017)

Select Genomic Profiles:

Mutations [?]

Putative copy-number alterations from GISTIC [?]

mRNA Expression z-Scores (RNA Seq V2 RSEM) [?]

Protein expression Z-scores (RPPA) [?]

Select Patient/Case Set:

Tumor Samples with sequencing and CNA data (149) x

To build your own case set, try out our enhanced Study View.

Enter Gene Set:

Advanced: Onco Query Language (OQL)

User-defined List x

Select from Recurrently Mutated Genes (MutSig) Select Genes from Recurrent CNAs (Gistic)

KRAS
TP53

All gene symbols are valid.

Submit Query

Modify Query Gene Set / Pathway is altered in 140 (94%) of queried samples

Pancreatic Adenocarcinoma (TCGA, Provisional)
Tumor Samples with sequencing and CNA data (149 samples) / 2 Genes

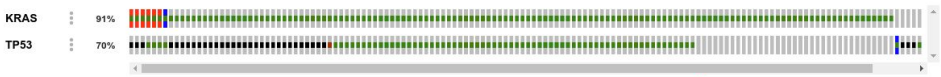
OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-Expression Enrichments Survival Network CN Segments Download Bookmark

Case Set: Tumor Samples with sequencing and CNA data (149 patients / 149 samples)

Altered in 140 (94%) of 149 sequenced cases/patients (149 total)

KRAS 91%

TP53 70%

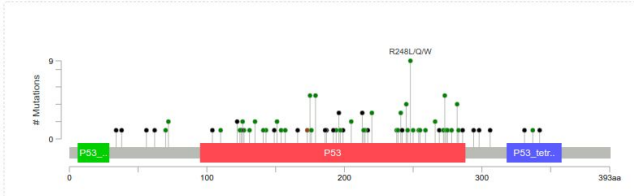


Genetic Alteration

- Inframe Mutation (putative driver)
- Missense Mutation (putative driver)
- Truncating Mutation (putative driver)
- ▬ Amplification
- ▬ Deep Deletion
- ▬ No alterations

OncoPrint Cancer Types Summary Mutual Exclusivity Plots Mutations Co-Expression Enrichments Survival Network CN Segments Download Bookmark

KRAS **TP53**



TP53

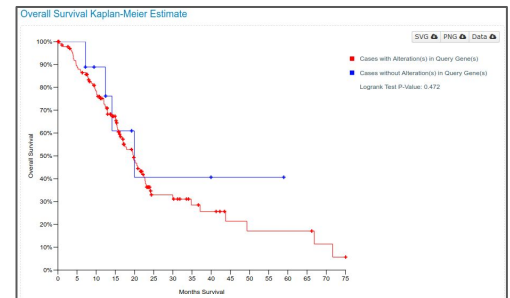
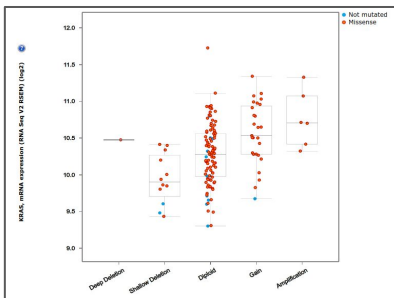
UniProt: P53_HUMAN

Transcript: ENST00000269305

Somatic Mutation Frequency: 70.5% [?]

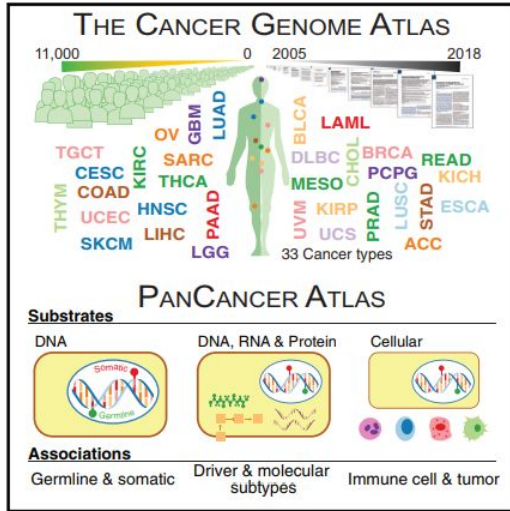
- Missense ● Truncating
- Inframe ● Other

View 3D Structure



Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics

Graphical Abstract



Authors

Li Ding, Matthew H. Bailey, Eduard Porta-Pardo, ..., David A. Wheeler, Gad Getz, The Cancer Genome Atlas Research Network

Correspondence

lding@wustl.edu (L.D.), wheeler@bcm.edu (D.A.W.), gadgetz@broadinstitute.org (G.G.)

In Brief

A synthesized view on oncogenic processes based on PanCancer Atlas analyses highlights the complex impact of genome alterations on the signaling and multi-omic profiles of human cancers as well as their influence on tumor microenvironment.

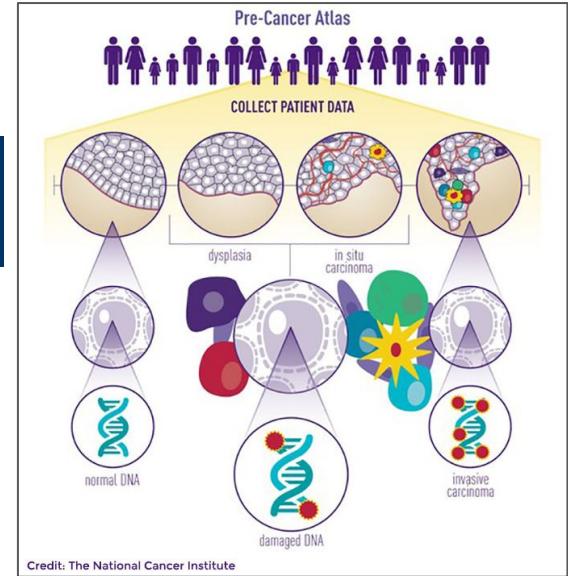
What else?

- [International Cancer Genome Consortium](#) (ICGC)
- [Pediatric Cancer Genome Project](#) (PGCP)
- [TARGET Project](#)

What's next?

CANCER MOONSHOT

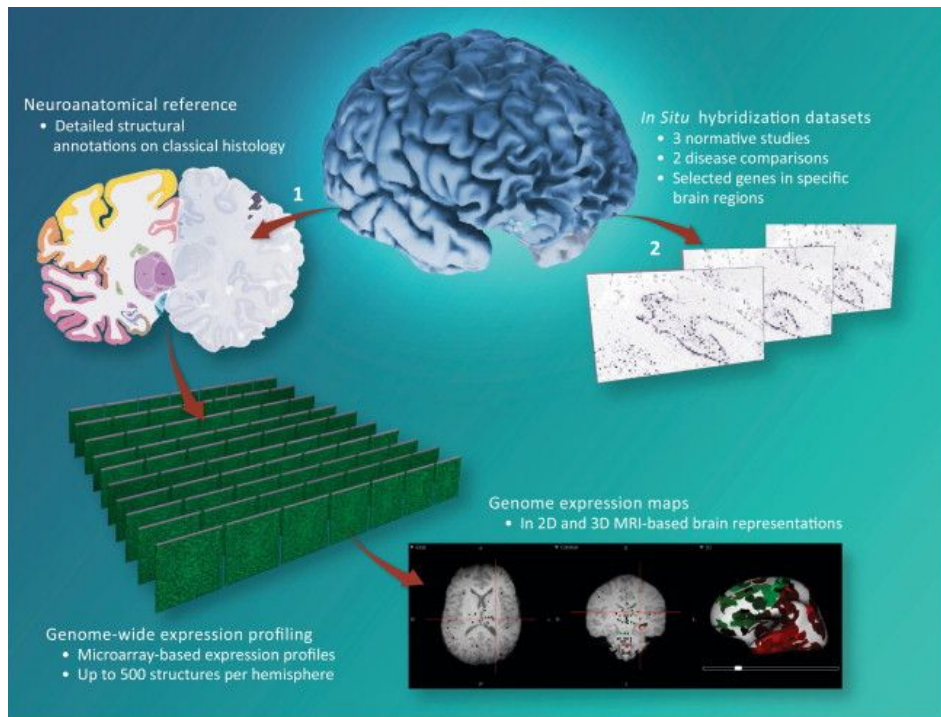
[Cancer Moonshot](#)



Cell. 2018, 173(2), 305-320

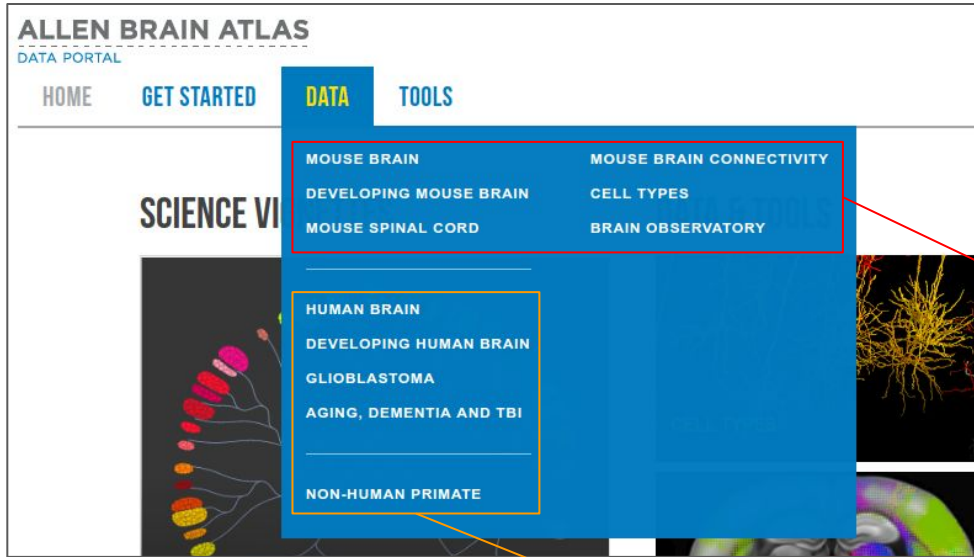
<https://doi.org/10.1016/j.cell.2018.03.033>

[Pre-Cancer Atlas](#)



- **Level:** Multi omics - organ level
- **Project:** [Allen Brain Atlas](#)
- **Aim:** Integrated analysis of cell line drug responsiveness
- **Samples:** Human and mouse brain biopsies
- **Technologies:** Microarray, ISH, RNA-Seq, MRI
- **Data:** Raw, processed and integrated

Data access: <http://www.brain-map.org/>



[Mouse Brain data](#): ISH data
[Developing Mouse Brain data](#): ISH data
[Mouse Spinal Cord](#): ISH data
[Mouse Brain Connectivity](#): Reconstruction using transgenic mice data
[Cell Type](#): RNA-Seq data, Morphology and Electrophysiology data
[Brain observatory](#): Physiological data transgenic mice

[Human Brain data](#): Microarray and RNA-Seq data, MRI scan
[Developing Human Brain data](#): Microarray, RNA-Seq, Exon microarray data
[Glioblastoma Atlas](#): RNA-Seq, Clinical data
[Aging, Dementia, TBI](#): RNA-Seq, Clinical data, IHC

[Non-human primate data](#): Microarray data

Data visualization and integration

Gene Search
 Differential Search
 Gene Classification
 Mouse Differential Search

Enter Gene Name, Gene Symbol, NCBI Accession Number or Entrez Gene ID

Show exact matches only

Structure:

- gray matter (GM)
- diencephalon (DiE)
- hypothalamus (Hy)
- tuberal region (TubR)
- dorsomedial hypothalamic nucleus (DM)

Gene Info:

Symbol: **ESR1**
 Name: **estrogen receptor 1**
 Probe: **A_23_P309739**
 Expression - z-score: 3.16201, log2 intensity: 3.99
 Related Data: **MOUSE HUMAN NHP**

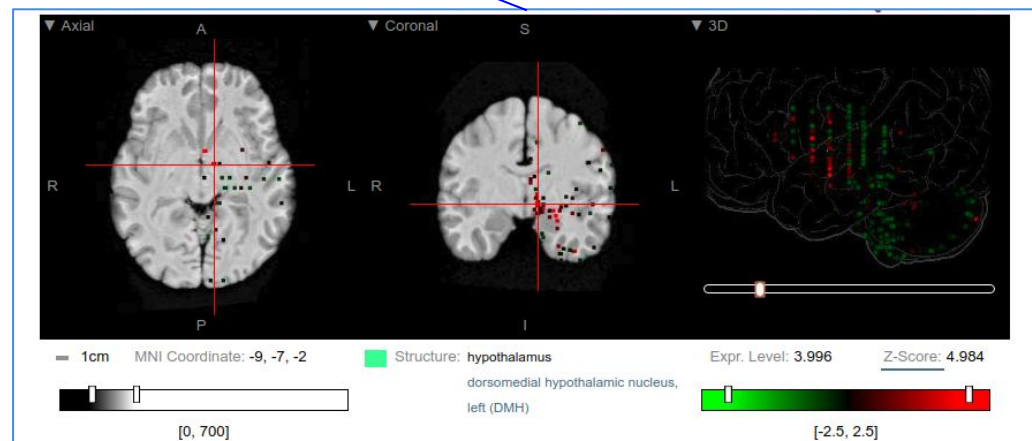
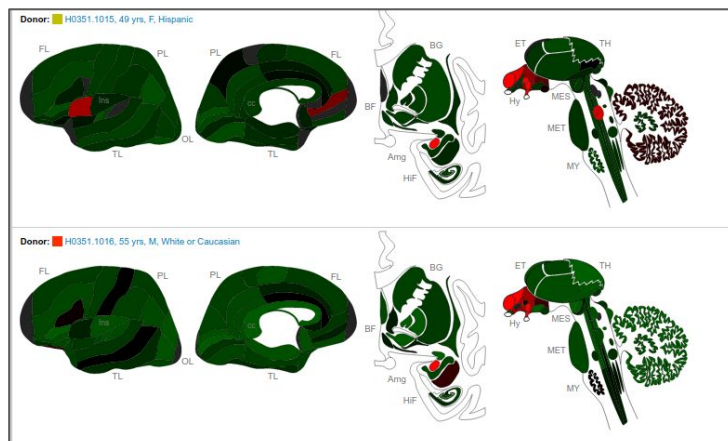
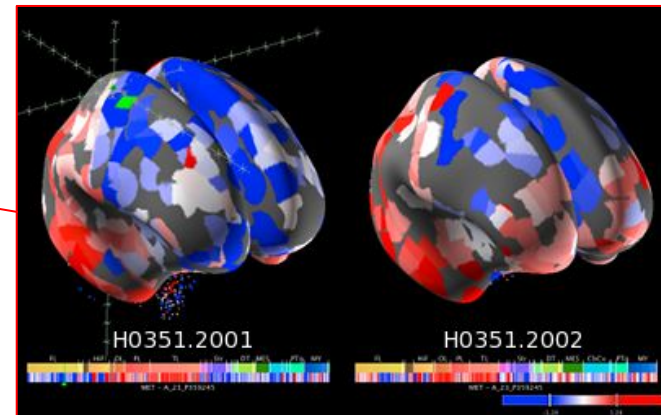
Donor:

- H0351.1012, 31 yrs, M, White or Caucasian

In:
For:
For this probe: **A_23_P309739**

1 - 4 of 4

Gene Symbol	Probe Name	Donor H0351.2002	V (Ve-V)
<input checked="" type="checkbox"/> ESR1	A_23_P309739		
<input type="checkbox"/> ESR1	A_24_P325215		
<input type="checkbox"/> ESR1	A_24_P383478		
<input type="checkbox"/> ESR1	CUST_255_PI416408490		



Trends in Neurosciences

Volume 35, Issue 12, December 2012, Pages 711-714

CellPress

Forum: Science & Society

The Allen Human Brain Atlas: Comprehensive gene expression mapping of the human brain

Elaine H. Shen ✉, Caroline C. Overly, Allan R. Jones

Show more

<https://doi.org/10.1016/j.tins.2012.09.005>

Get rights and content

The Allen Human Brain Atlas is a freely available multimodal atlas of gene expression and anatomy comprising a comprehensive 'all genes—all structures' array-based dataset of gene expression and complementary *in situ* hybridization (ISH) gene expression studies targeting selected genes in specific brain regions. Available via the Allen Brain Atlas data portal (www.brain-map.org), the Atlas integrates structure, function, and gene expression data to accelerate basic and clinical research of the human brain in normal and disease states.

Trend in Neuros. 2012, 35(12), 711-714
<https://doi.org/10.1016/j.tins.2012.09.005>

What else?

- [Human Connectome Project](#)
- [Human Brain Project](#)
- [iPOP](#) and [P100](#)

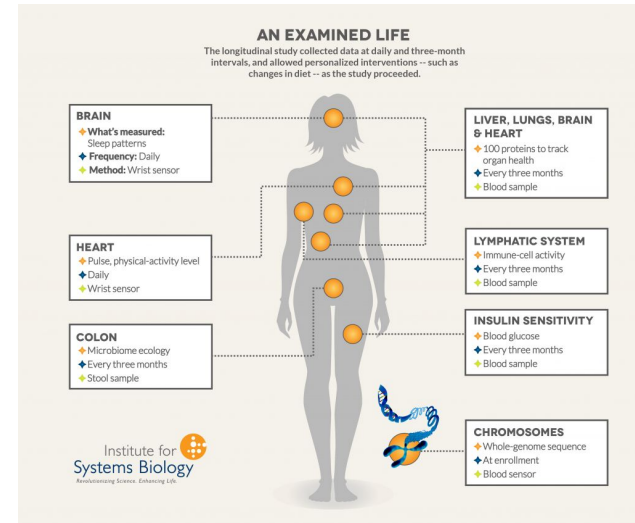
What's next?

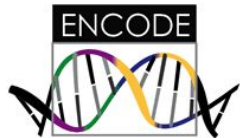
100 K

100K WELLNESS PROJECT

Institute for Systems Biology

[100K Wellness Project](#)



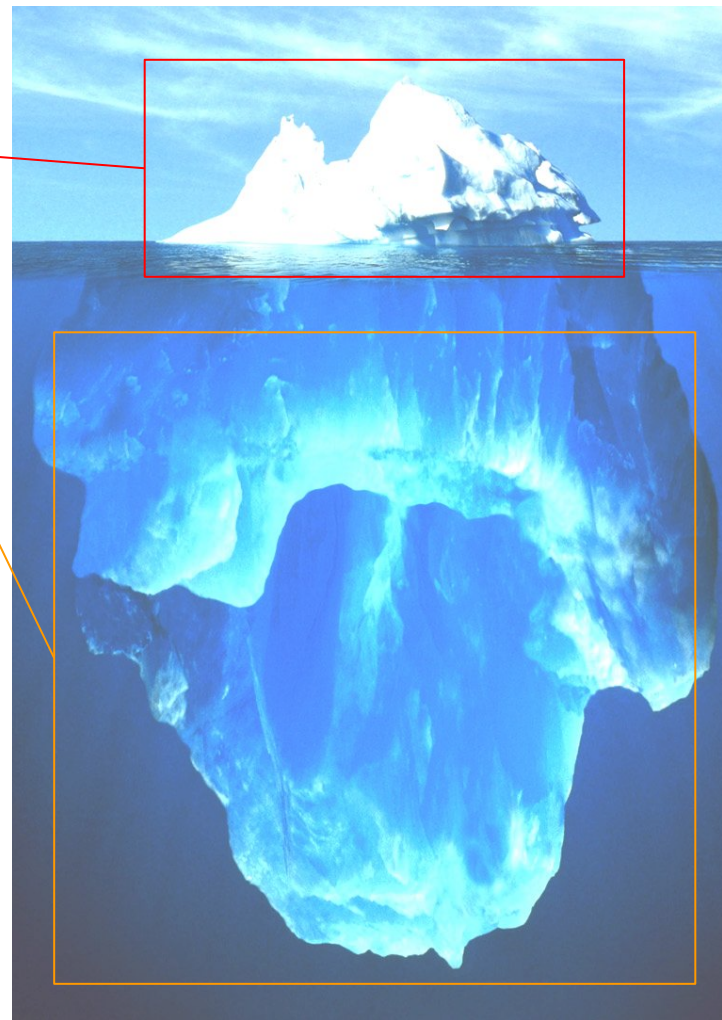


1000 Genomes
A Deep Catalog of Human Genetic Variation

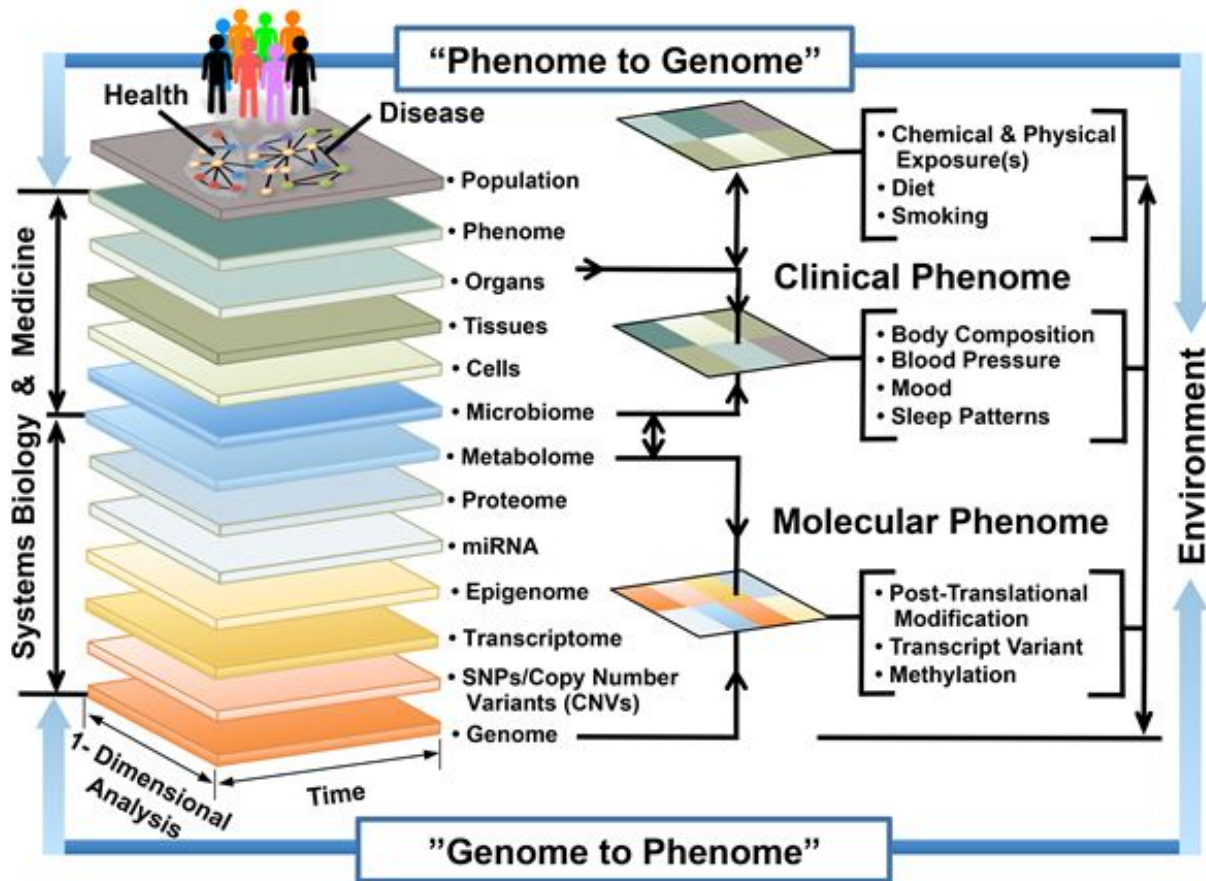


THE HUMAN PROTEIN ATLAS

NIH THE CANCER GENOME ATLAS
National Cancer Institute
National Human Genome Research Institute



- A large amount of high-throughput data derived from small scale omic project and research
- These data are stored in repositories like [Gene Expression Omnibus](#) (GEO) and [ArrayExpress](#) or provided as supplementary material of publications



- A large number of biological systems and phenotypes were studied using high-throughput technologies
- Most of these data are freely available and easy to retrieve and visualize
- Any biological studies can be improved by integrating these public omic information.
- Then exploit them and stay updated (e.g. pubmed “save” function)