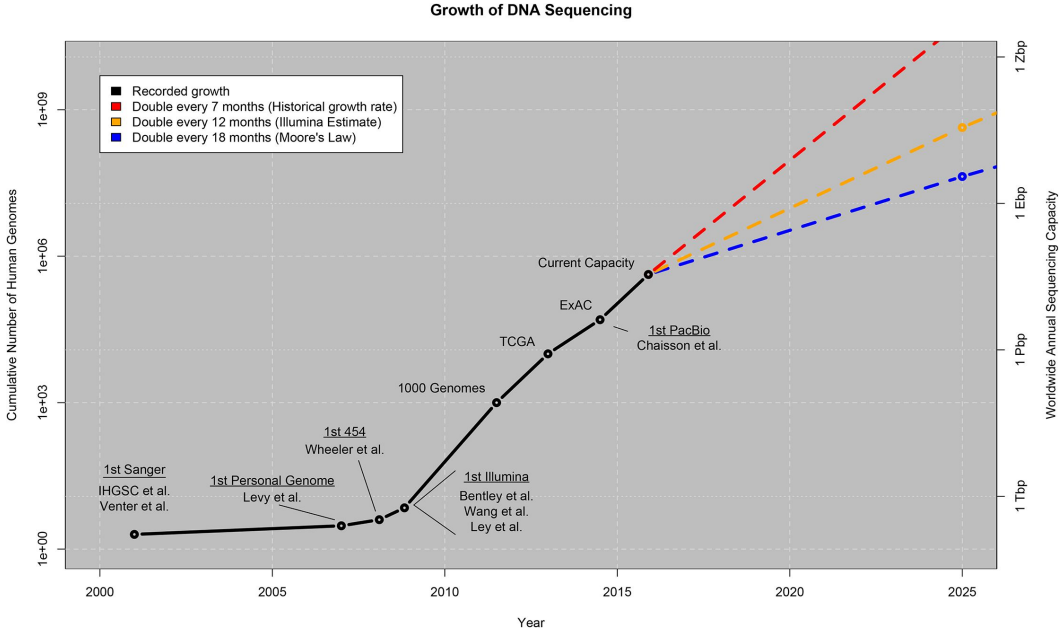
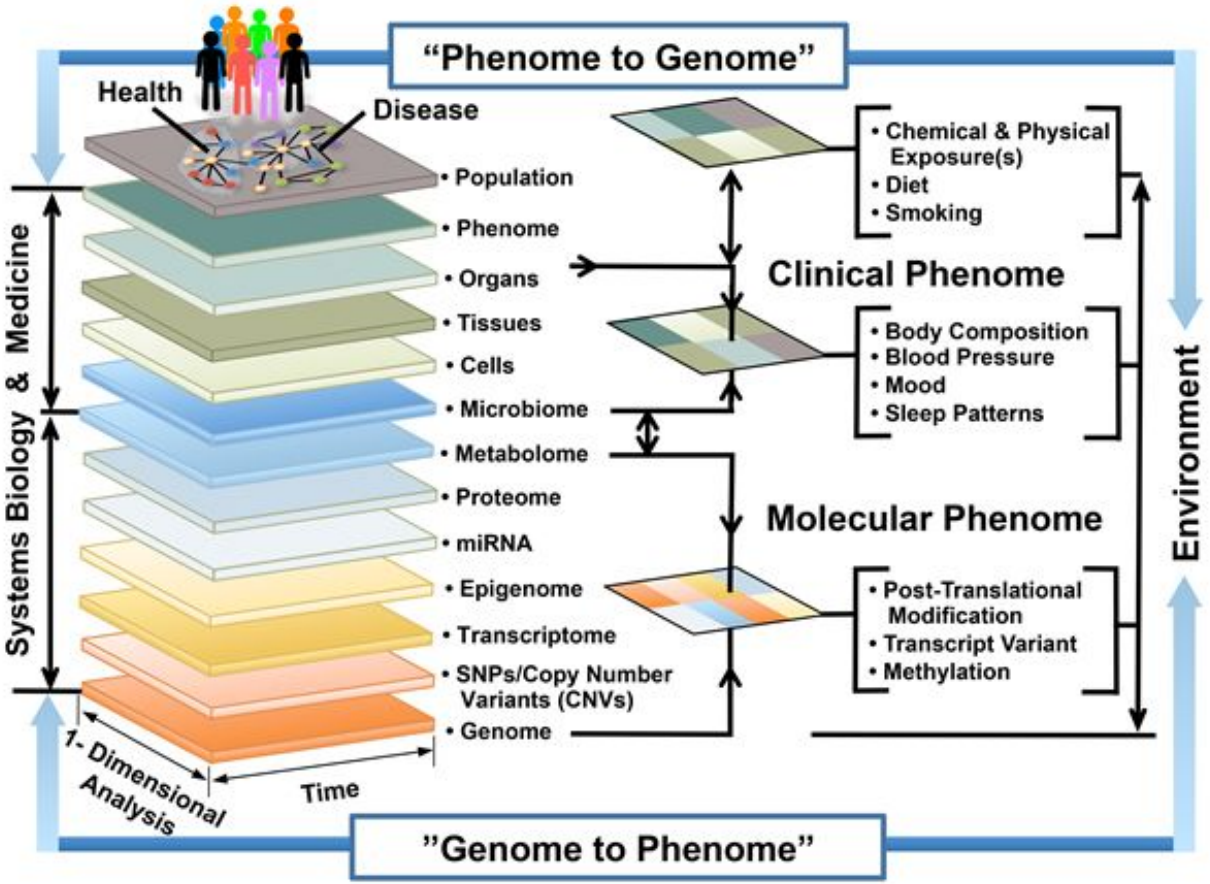
Public resources from Large Scale Omic Projects



From: *PLoS biology,* 13(7), 2015, e1002195.
https://doi.org/10.1371/journal.pbio.1002195

- 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: https://medicine.umich.edu/dept/dcmb/research

- 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_1
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, ...

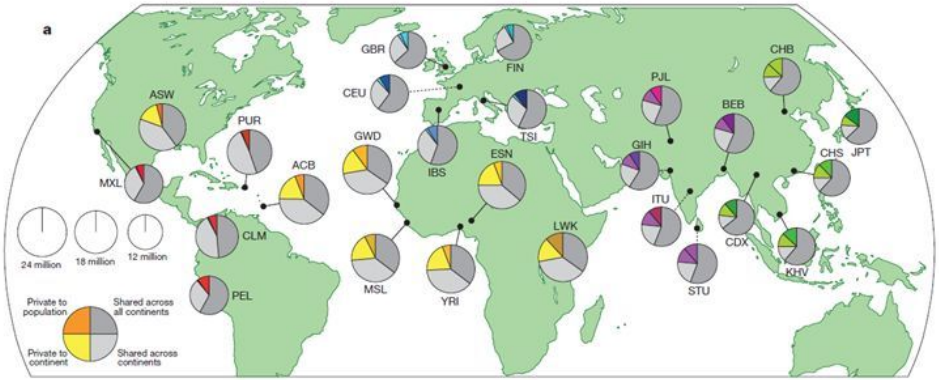
1000 Genomes Project

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

Nature Collection

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

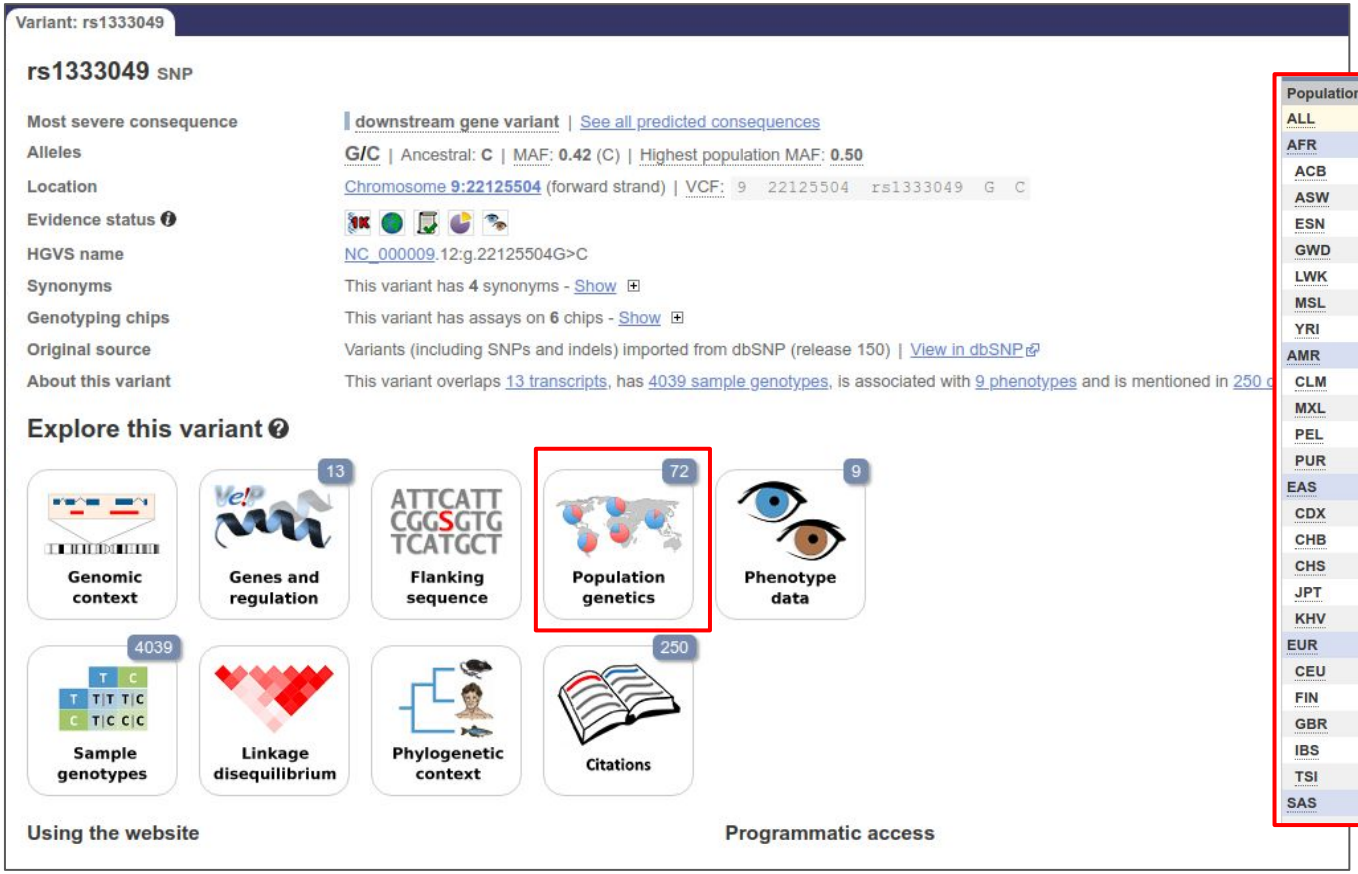
Sequence Alignment Map (SAM) Format

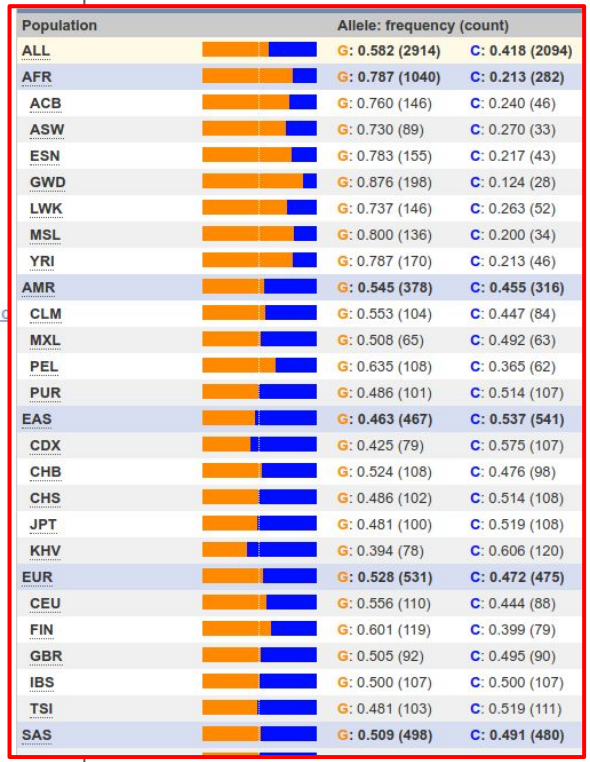
```
@HD VN:1.3 SO:coordinate
@SQ SN:ref LN:45
r001 163 ref  7 30 8M2I4M1D3M = 37  39 TTAGATAAAGGATACTG *
r002   0 ref  9 30 3S6M1P1I4M *  0   0 AAAAGATAAGGATA    *
r003   0 ref  9 30 5H6M       *  0   0 AGCTAA           *   NM:i:1
r004   0 ref 16 30 6M14N5M    *  0   0 ATAGCTTCAGC      *
r003  16 ref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





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.

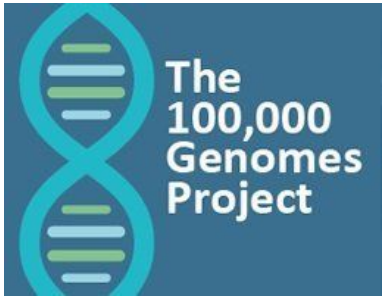
Brief. in Functional Genomics, 16(3), 2017, 163-70,
https://doi.org/10.1093/bfgp/elw027

What else?

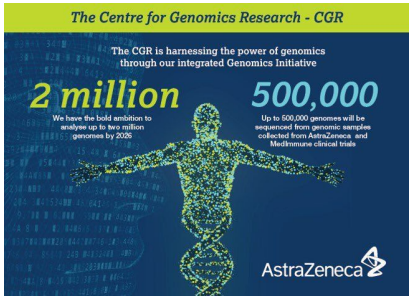
- Exome Aggregation Consortium (ExAC)
- European Genome-phenome Archive (EGA)
- The database of Genotypes and Phenotypes (dbGaP)

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

What's next?

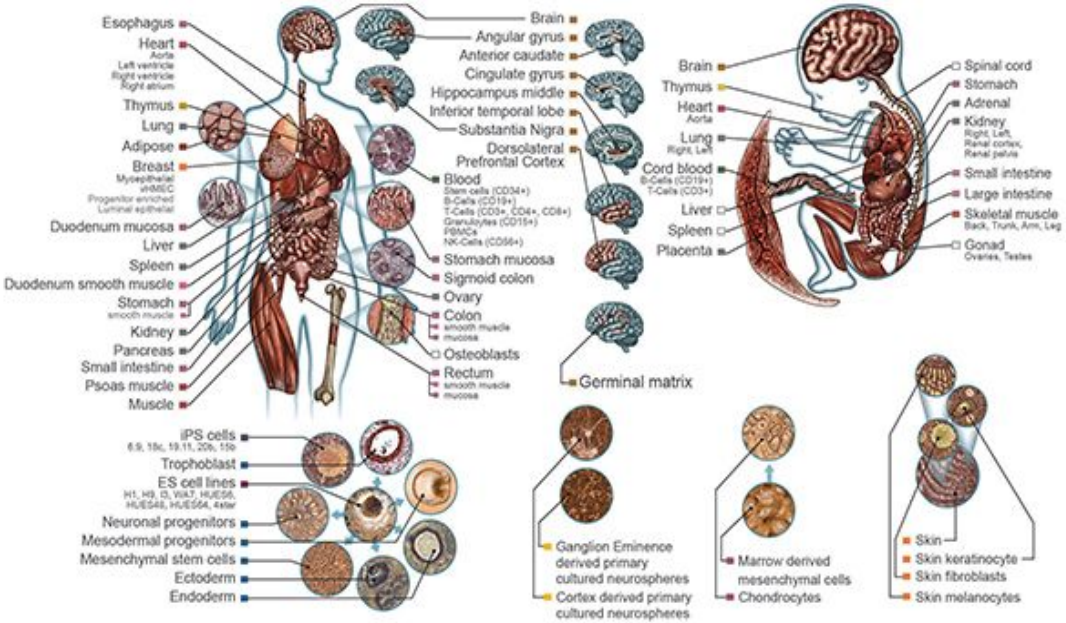


100,000 Genomes Project



AstraZeneca 2 million genomes

ROADMAP epigenomics PROJECT



- **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

Nature Collection

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

Processed data for single dataset:

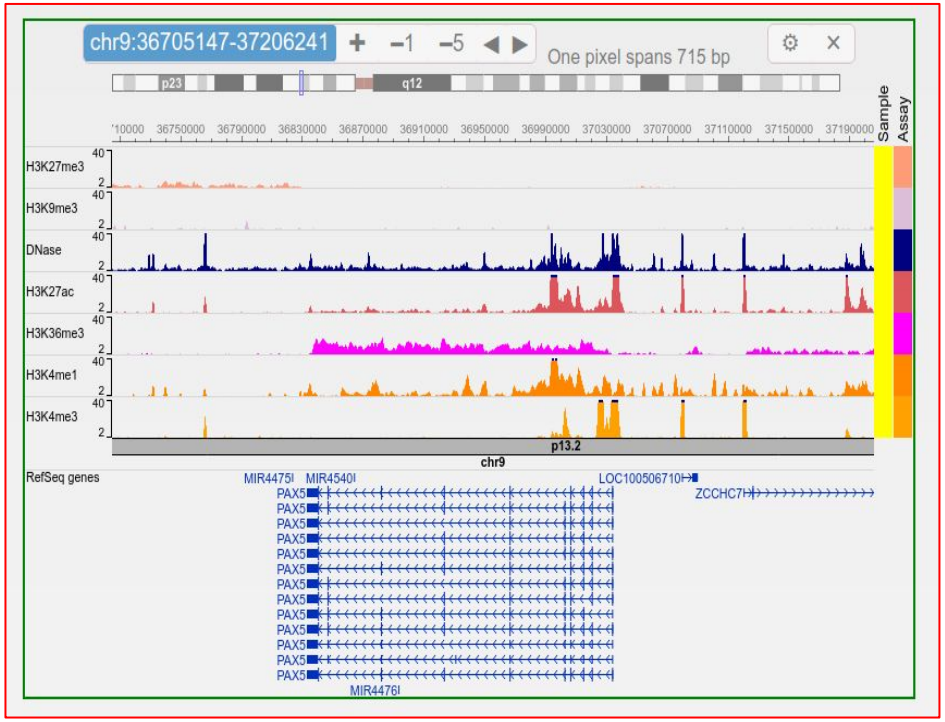
- Cell line, dataset Metadata
- Quality control results
- Read Alignments
- Signal peaks
- Genome-wide signal coverage
- Expression quantification

Integrated data from multiple datasets

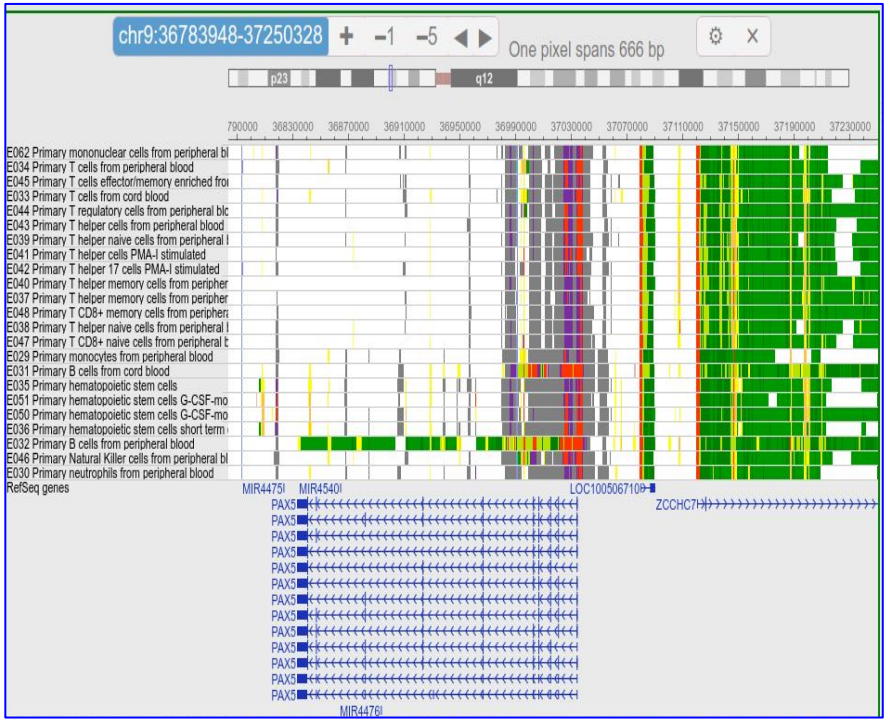
- Chromatin states
- Comparison between epigenomes
- Predicted promoters and enhancers
- Tissue-specific regulators
- Prediction of disease variant functional effect

Data visualization and integration

Processed data for single dataset / tissue



Integrated data from multiple datasets / tissues



[WashU Genome Browser]

ARTICLE

OPEN

doi:10.1038/nature14248

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-Moussavi[5], Pouya Kheradpour[1,2], Zhizhuo Zhang[1,2], Jianrong Wang[1,2], Michael J. Ziller[2,6], Viren Amin[7], John W. Whitaker[8], Matthew D. Schultz[9], Lucas D. Ward[1,2], Abhishek Sarkar[1,2], Gerald Quon[1,2], Richard S. Sandstrom[10], Matthew L. Eaton[1,2], Yi-Chieh Wu[1,2], Andreas R. Pfenning[1,2], Xinchen Wang[1,2,11], Melina Claussnitzer[1,2], Yaping Liu[1,2], Cristian Coarfa[7], R. Alan Harris[7], Noam Shoresh[2], Charles B. Epstein[2], Elizabeta Gjoneska[2,12], Danny Leung[8,13], Wei Xie[8,13], R. David Hawkins[8,13], Ryan Lister[9], Chibo Hong[14], Philippe Gascard[15], Andrew J. Mungall[5], Richard Moore[5], Eric Chuah[5], Angela Tam[5], Theresa K. Canfield[10], R. Scott Hansen[16], Rajinder Kaul[16], Peter J. Sabo[10], Mukul S. Bansal[1,2,17], Annaick Carles[18], Jesse R. Dixon[8,13], Kai-How Farh[2], Soheil Feizi[1,2], Rosa Karlic[19], Ah-Ram Kim[1,2], Ashwinikumar Kulkarni[20], Daofeng Li[21], Rebecca Lowdon[21], GiNell Elliott[21], Tim R. Mercer[22], Shane J. Neph[10], Vitor Onuchic[7], Paz Polak[2,23], Nisha Rajagopal[8,13], Pradipta Ray[20], Richard C. Sallari[1,2], Kyle T. Siebenthall[10], Nicholas A. Sinnott-Armstrong[1,2], Michael Stevens[21,42], Robert E. Thurman[10], Jie Wu[24,25], Bo Zhang[21], Xin Zhou[21], Arthur E. Beaudet[26], Laurie A. Boyer[11], Philip L. De Jager[2,23,27], Peggy J. Farnham[28], Susan J. Fisher[29], David Haussler[30], Steven J. M. Jones[5,31,32], Wei Li[33], Marco A. Marra[5,32], Michael T. McManus[34], Shamil Sunyaev[2,23,27], James A. Thomson[35,41], Thea D. Tlsty[15], Li-Huei Tsai[2,12], Wei Wang[8], Robert A. Waterland[36], Michael Q. Zhang[20,37], Lisa H. Chadwick[38], Bradley E. Bernstein[2,39,40]§, Joseph F. Costello[14]§, Joseph R. Ecker[9]§, Martin Hirst[5,18]§, Alexander Meissner[2,6]§, 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

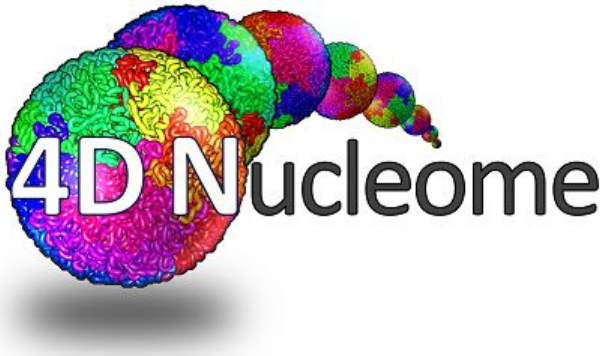
What else?

- Blueprint
- CEEHRC

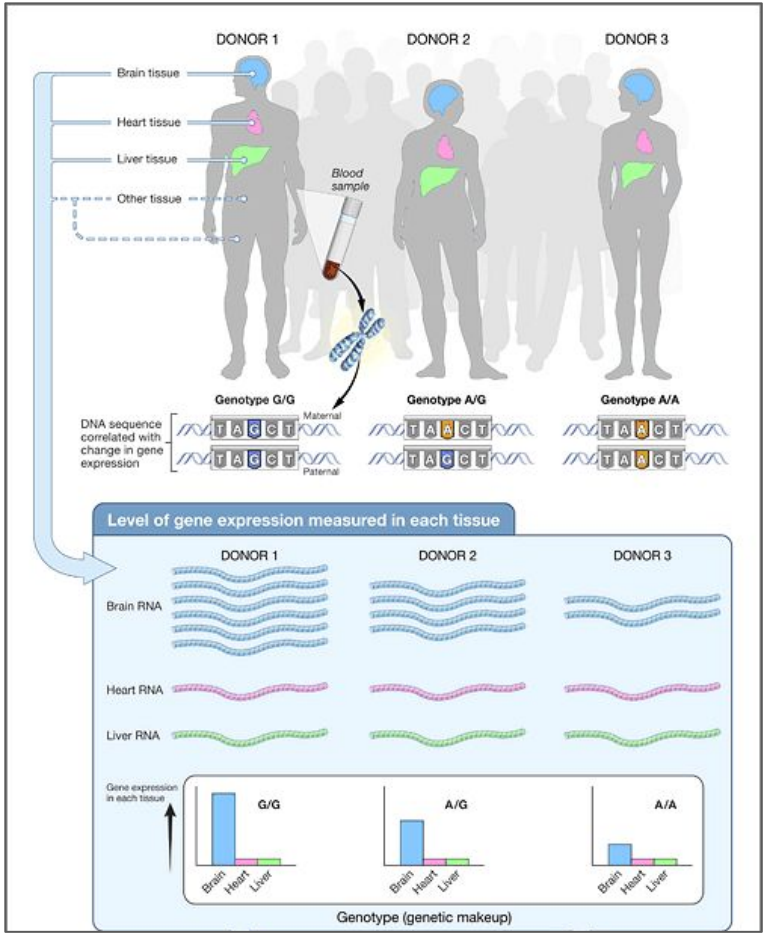
> (Cell special issue)

- 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

Nature Collection

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.

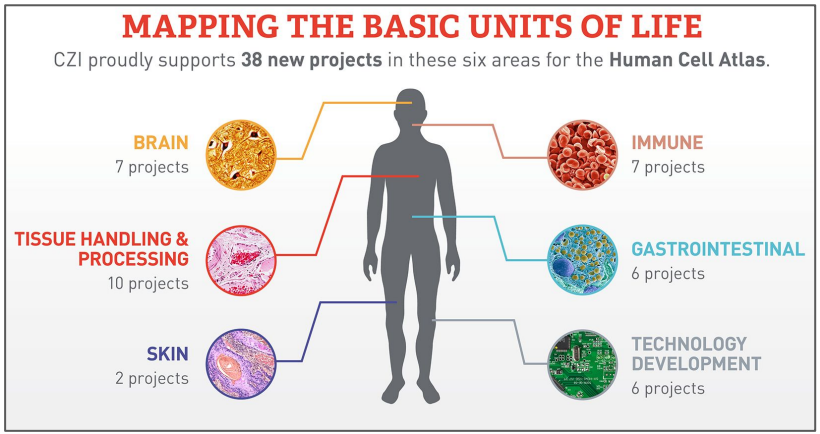
Nature, 550, 2017, 204-213.
http://doi.org/10.1038/nature24277

What else?

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

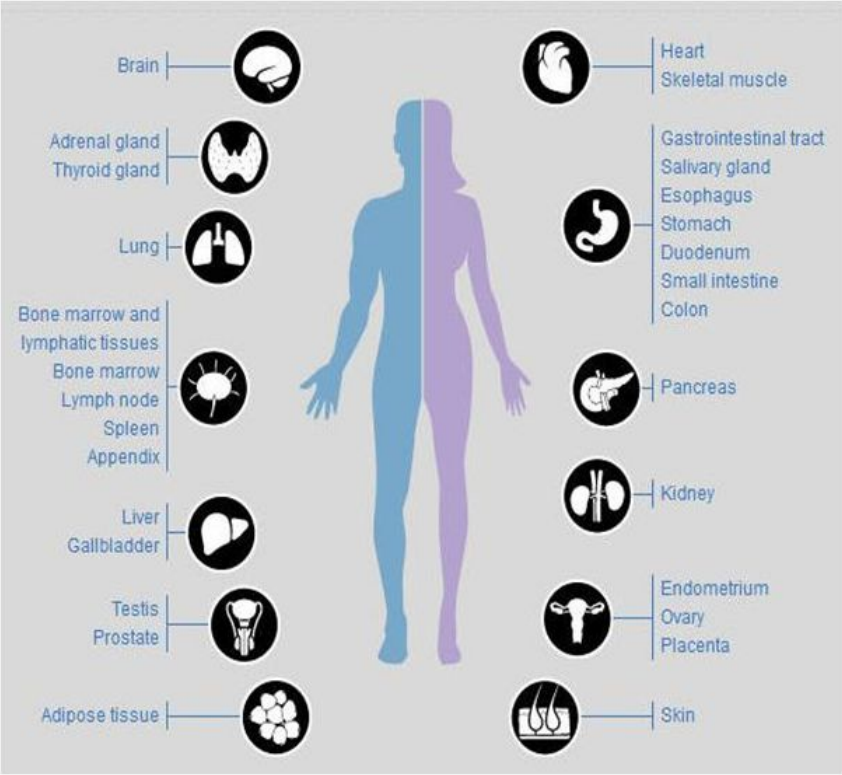
What's next?

HUMAN CELL ATLAS



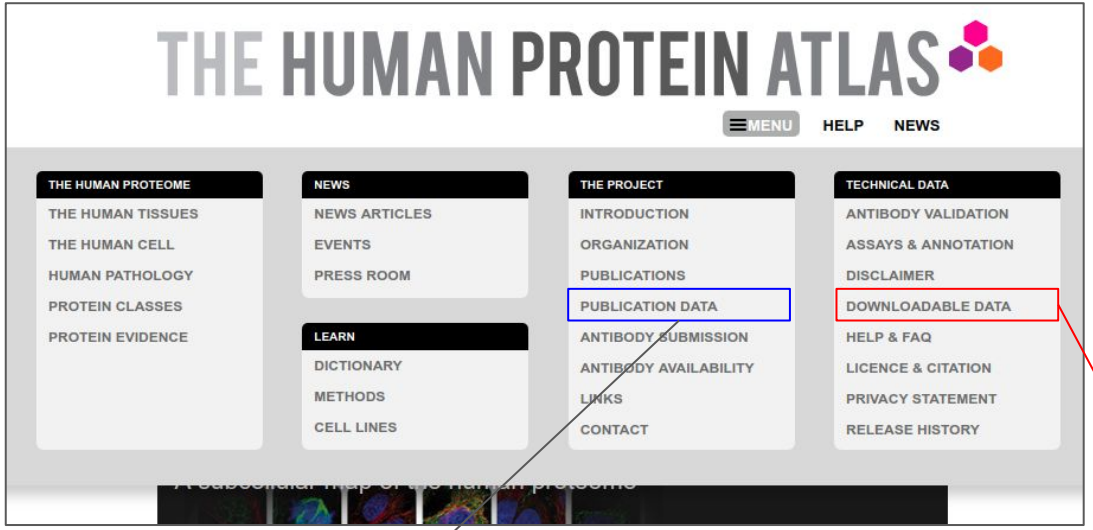
Human Cell Atlas

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



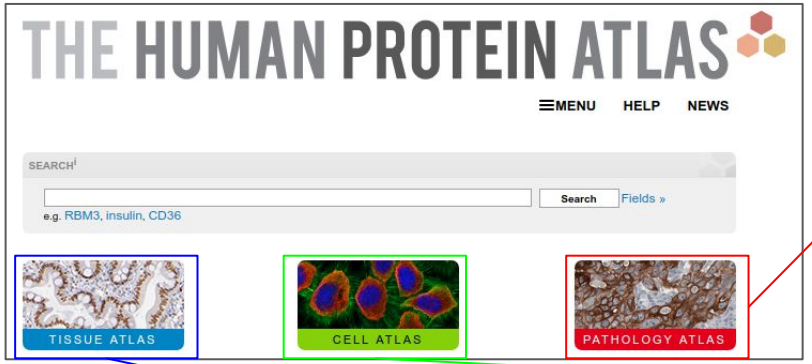
[Publication Data]

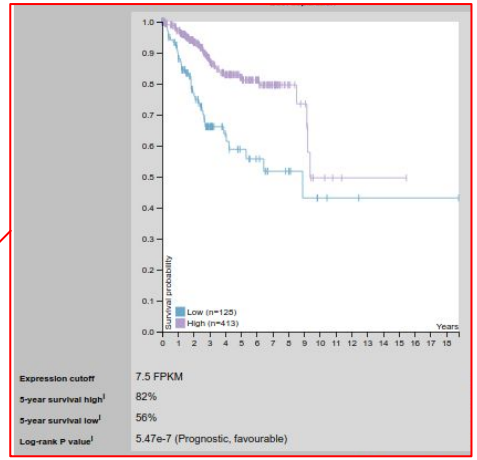
- Antibody response
- Protein array expression data - Tissue
- Protein array expression data - Cells
- Immunofluorescence expression data
- RNA isoform expression data

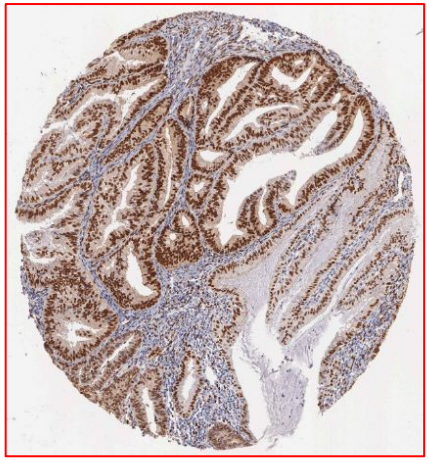
[Downloadable Data]

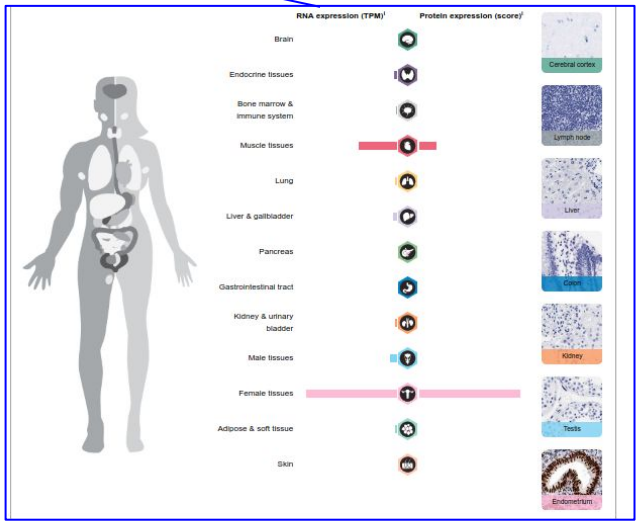
- Normal tissue data (level and reliability)
- Pathology data (level, reliability, correlation with survival)
- Subcellular localization data (reliability, level, between cell variations, cell cycle dependency)
- RNA-Seq data (gene/isoform level in TPM)

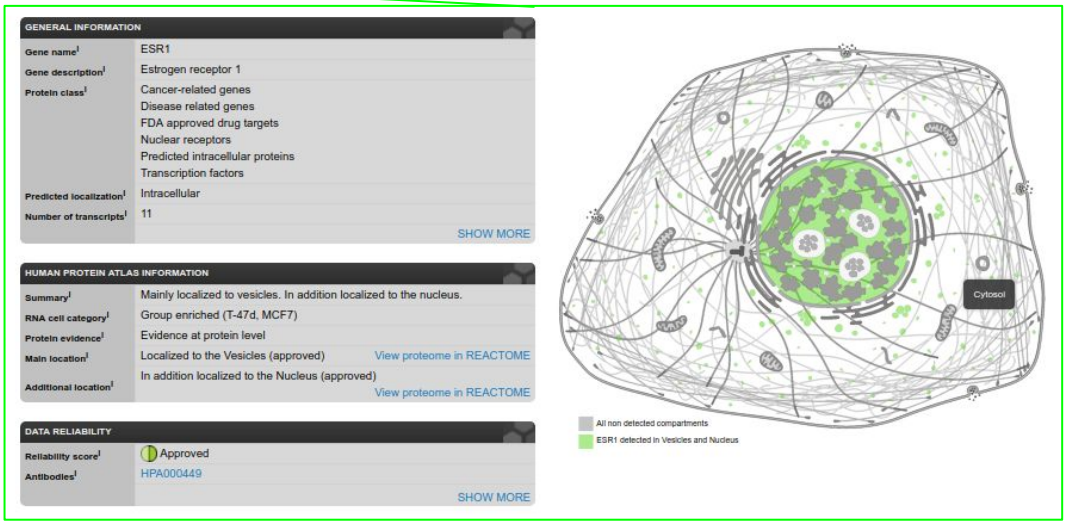
Data visualization and integration

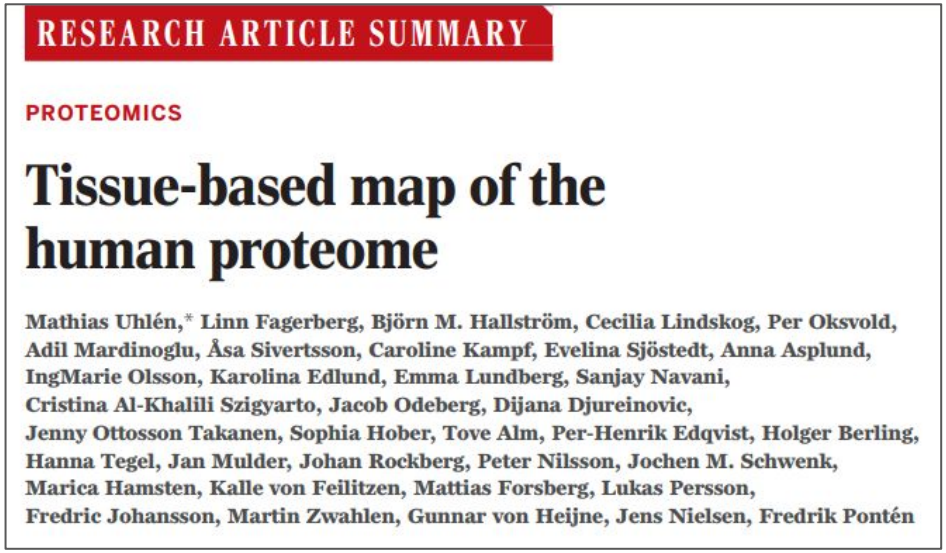












RESEARCH ARTICLE SUMMARY

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 Sjöstedt, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szigyarto, 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?

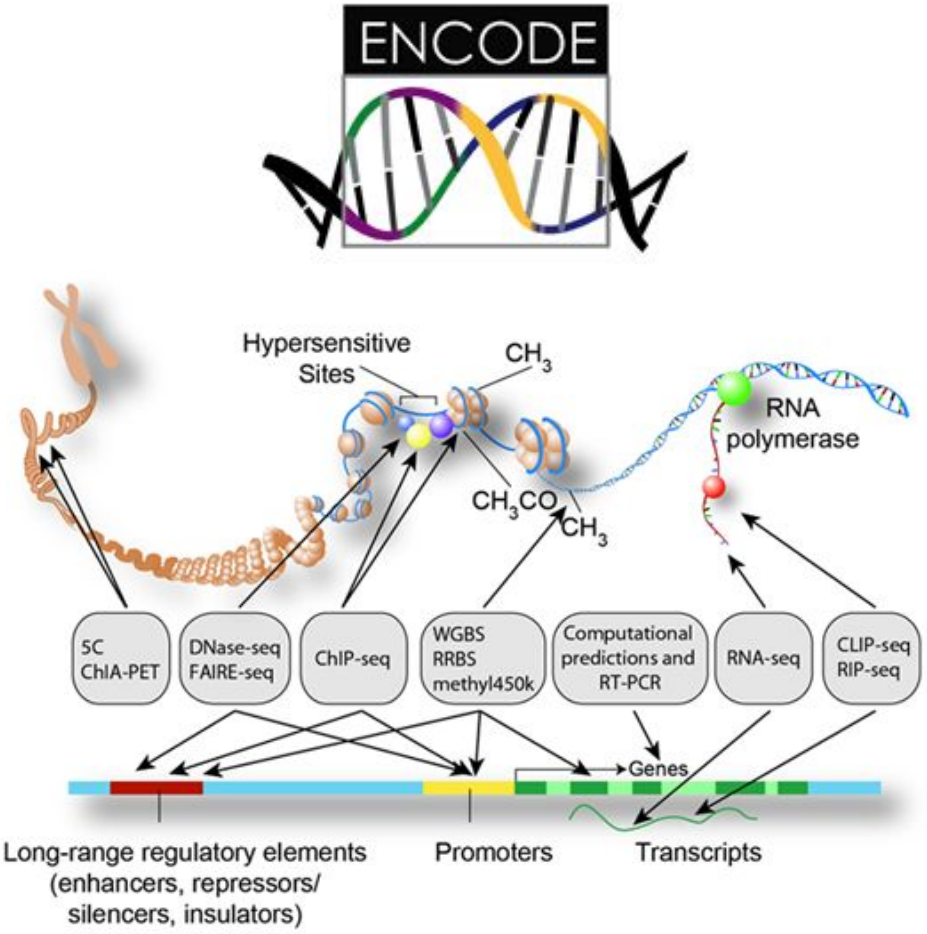
- Human Proteome Map
- Human Metabolome Project
- Clinical Proteomic Tumor Analysis Consortium (CPTAC)

What's next?

HUPO
HUMAN PROTEOME ORGANIZATION
translating the code of life

C-HPP
Chromosome-based Human Proteome Project (C-HPP)

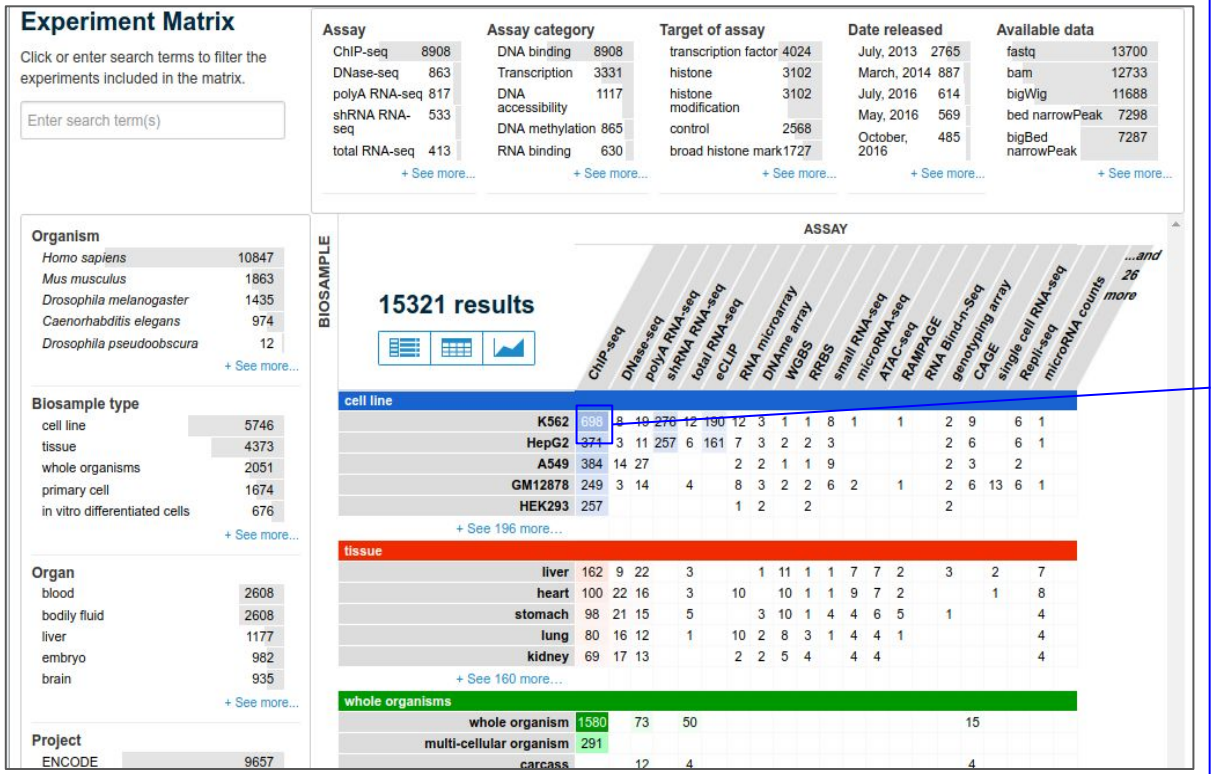
B/D HPP
Biology/Disease Human Proteome Project (B/D-HPP)

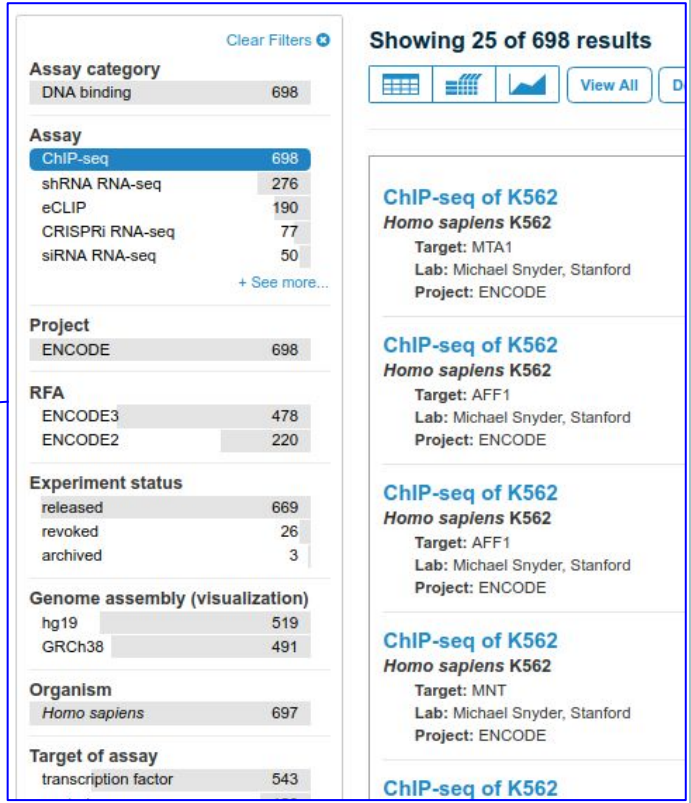


Nature Collection

- **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





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:	*Homo sapiens* 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
[illegible]	[illegible]	(ftp)(http)	BED

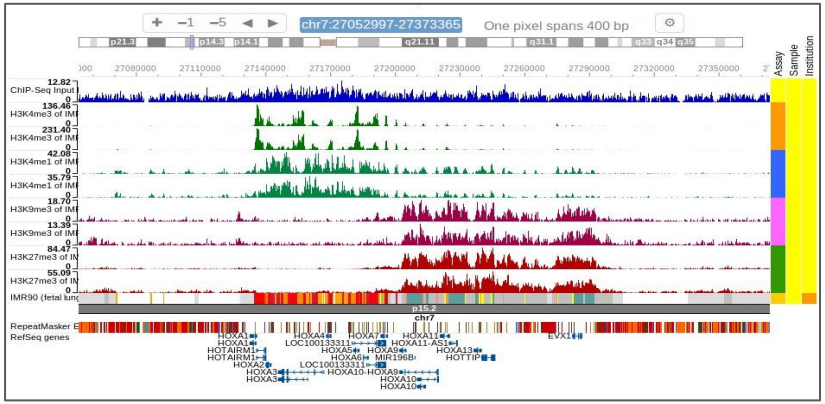
SE105nnn/GSE105823/suppl/GSE105823_ENCFF334TNK_peaks_GRCh38.bigBed

Available files:

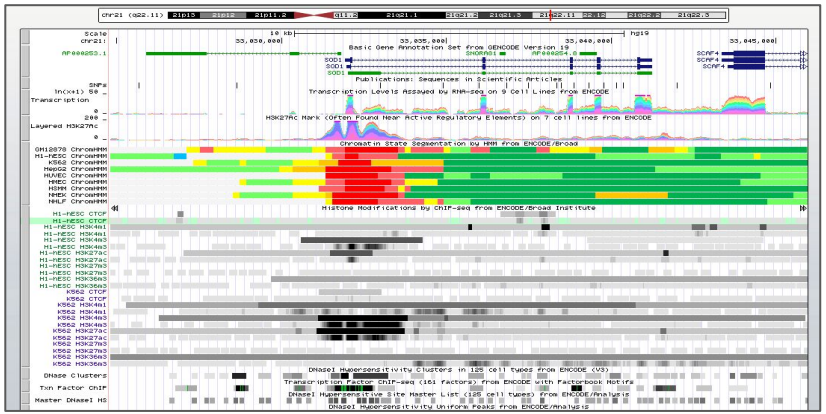
- Experiment protocols and QC
- Raw reads (Fastq)
- Alignment files (BED, SAM, BAM)
- Genomic coverage files (BIGWIG)
- Peaks coordinates (BED, BIGBED)

Data visualization and integration

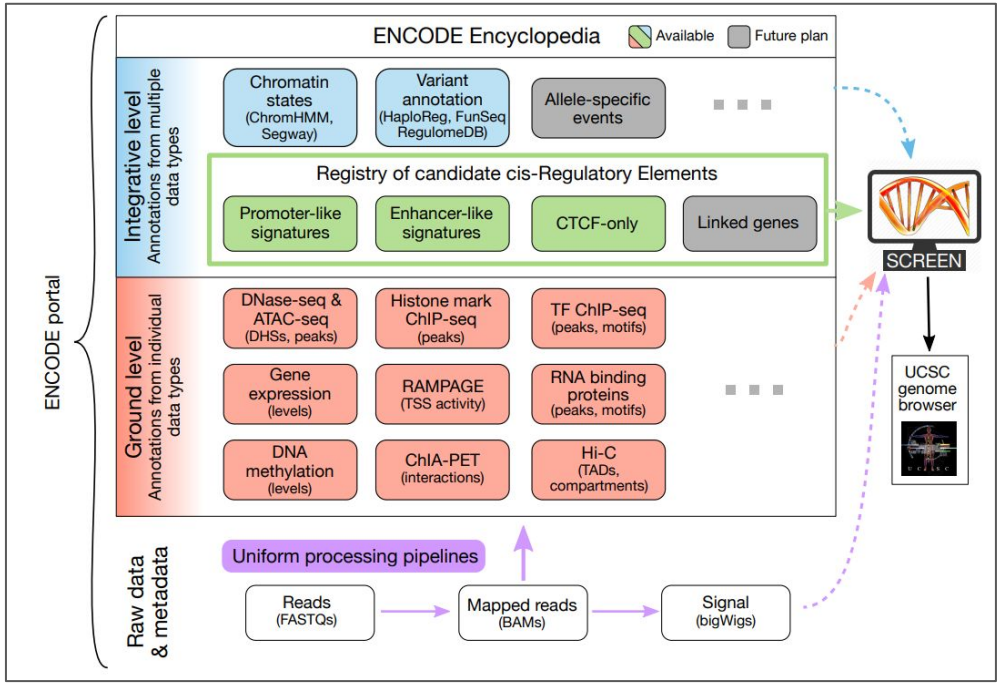
WashU Genome Browser



UCSC Genome Browser



Search Candidate cis-Regulatory Elements by ENCODE (SCREEN)



Review

Deciphering ENCODE

Adam G. Diehl[1] and Alan P. Boyle[1,2,*]

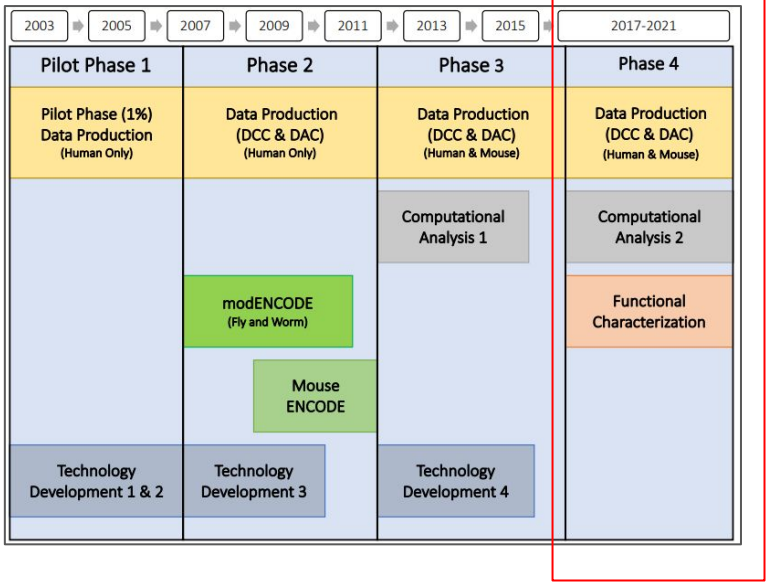
The ENCODE project represents a major leap from merely describing and comparing genomic sequences to surveying them for direct indicators of function. The astounding quantity of data produced by the ENCODE consortium can serve as a map to locate specific landmarks, guide hypothesis generation, and lead 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

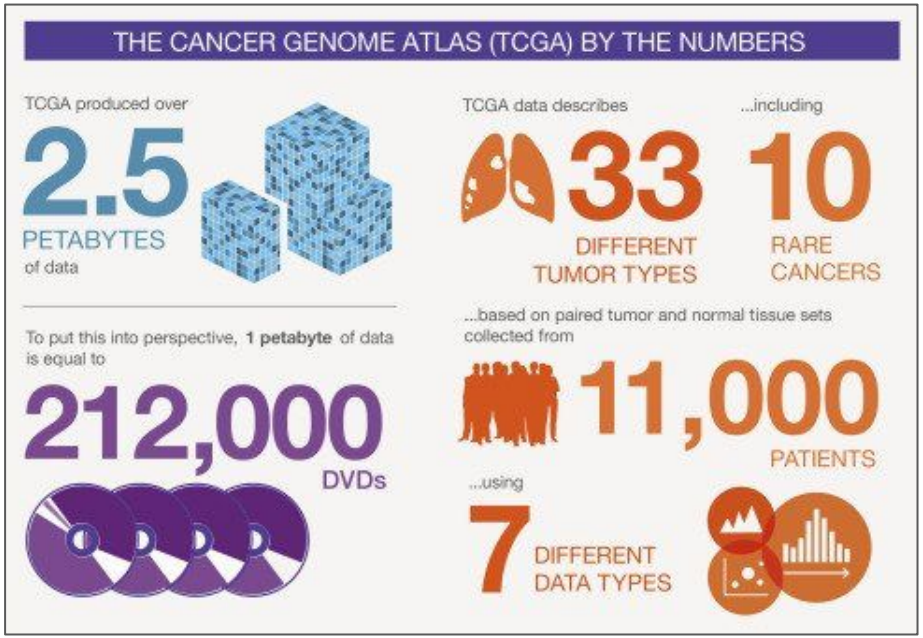
What else?

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

What's next?



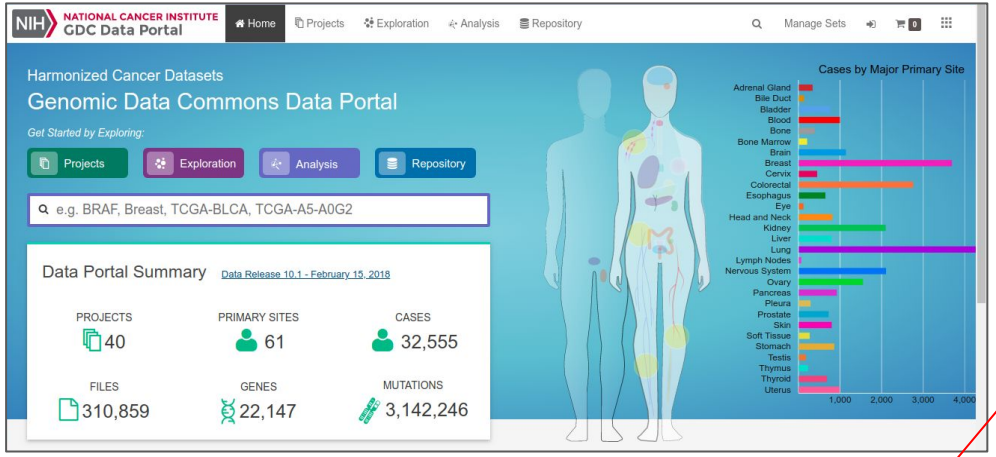
NIH **THE CANCER GENOME ATLAS**
National Cancer Institute
National Human Genome Research Institute

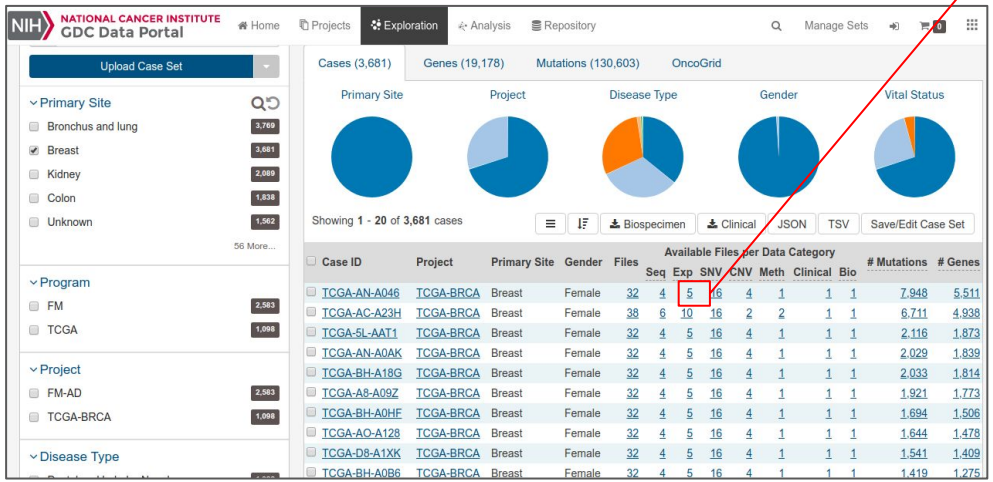


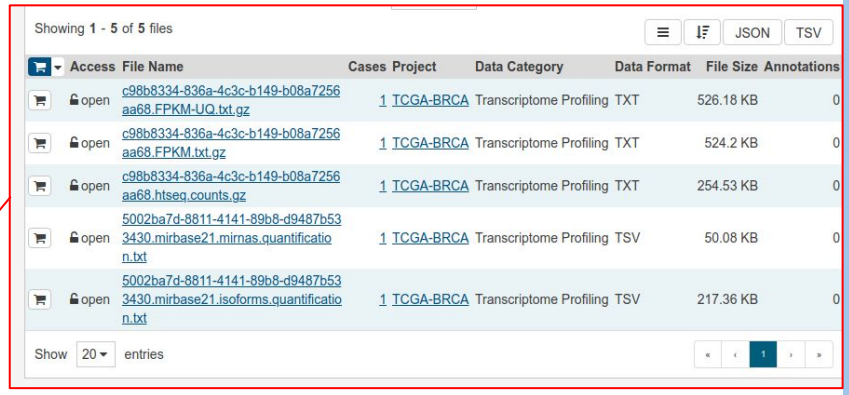
[Publication list]

- **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/





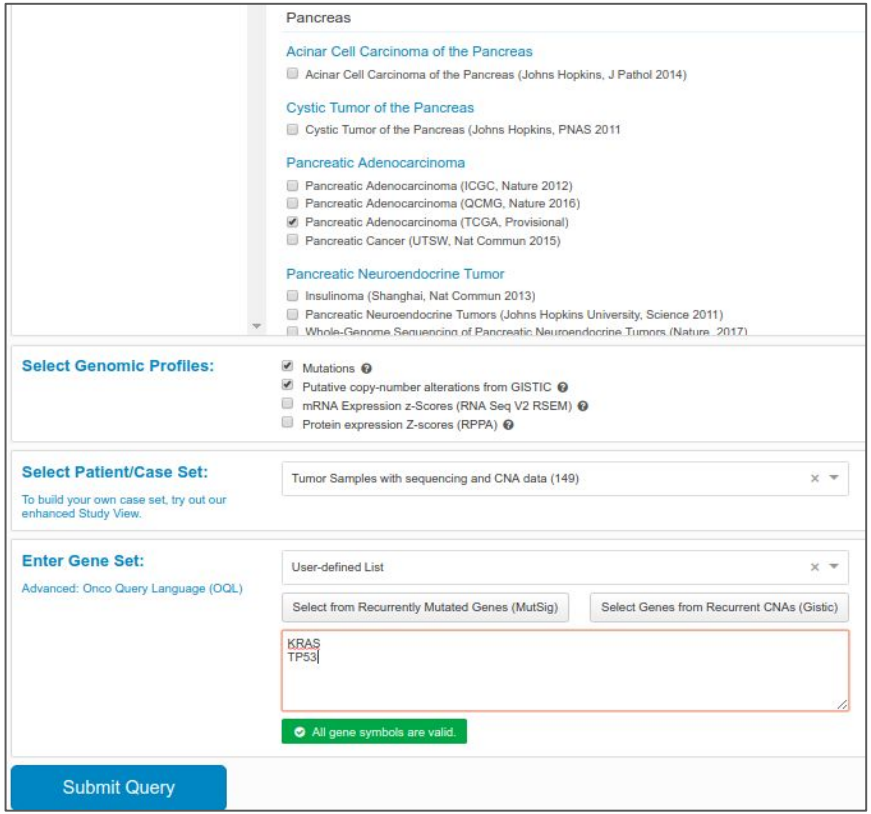


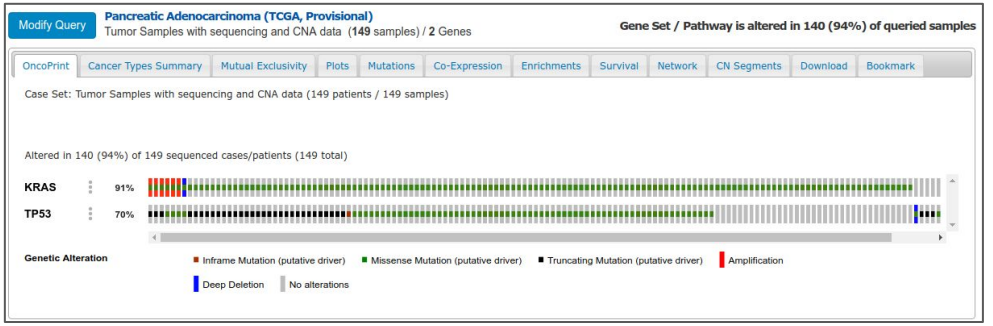
Available files:

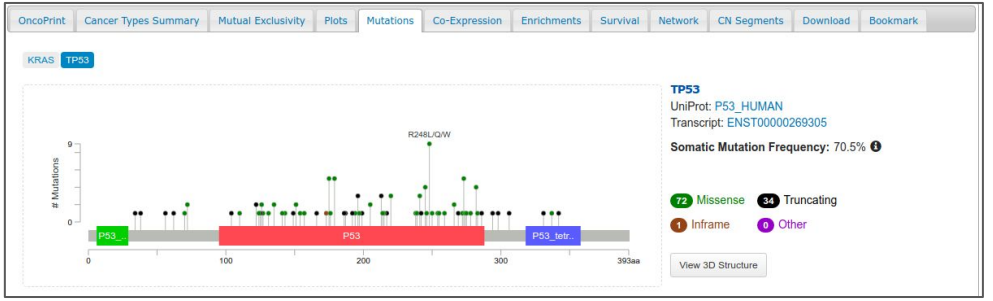
- Gene / miRNA expression quantification
- VCF and MAF file
- CNV data
- Methylation data
- Protein level data
- Clinical data

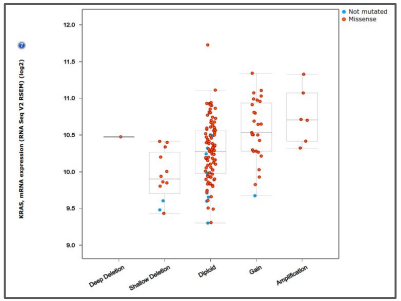
Data visualization and integration

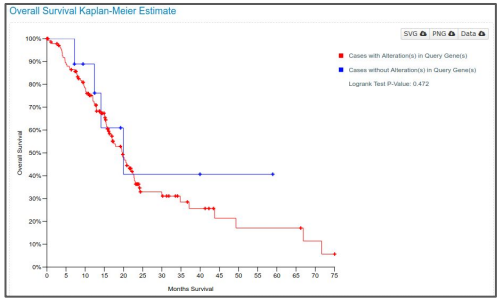
[cBioPortal](#)

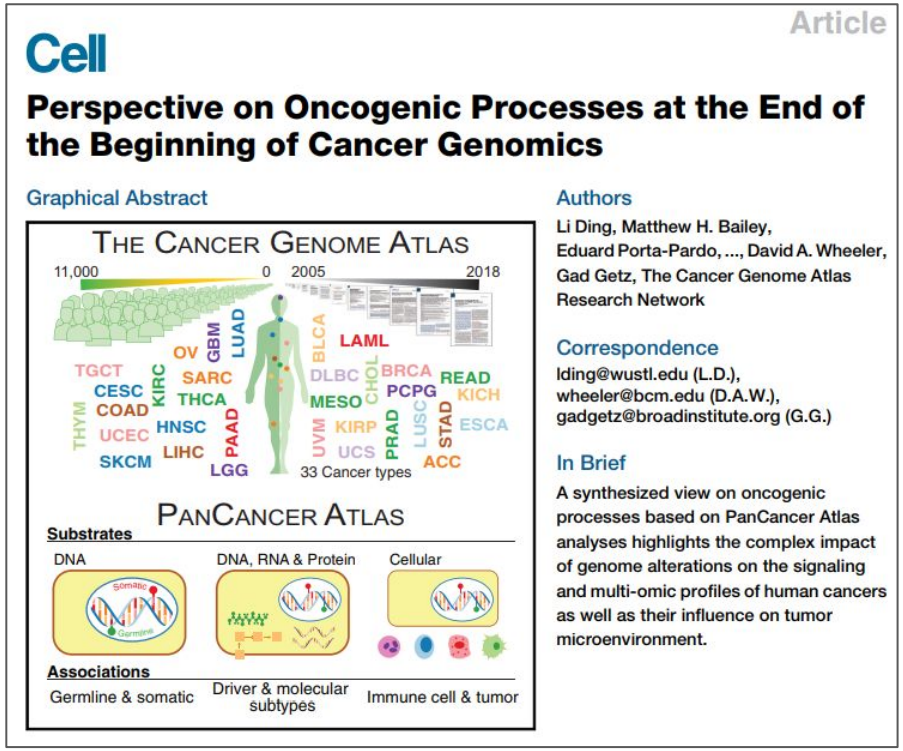












Article

Cell

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

Graphical Abstract

THE CANCER GENOME ATLAS

11,000 0 2005 2018

OV GBM LUAD BLCA LAML TGCT KIRC SARC DLBC CHOL BRCA READ CESC THCA MESO PCPG KICH THYM COAD HNSC UCEC KIRP LUSC STAD ESCA PAAD UVM PRAD LIHC SKCM UCS ACC LGG

33 Cancer types

PANCANCER ATLAS

Substrates

DNA DNA, RNA & Protein Cellular

Somatic Germline

Associations

Germline & somatic Driver & molecular subtypes Immune cell & tumor

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.

Cell. 2018, 173(2), 305-320
https://doi.org/10.1016/j.cell.2018.03.033

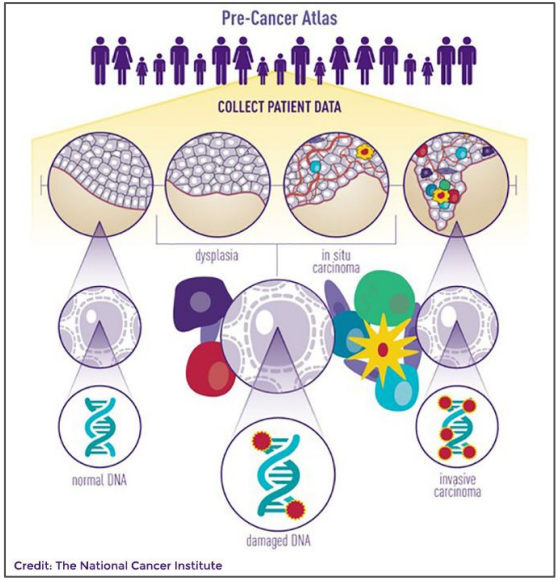
What else?

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

What's next?

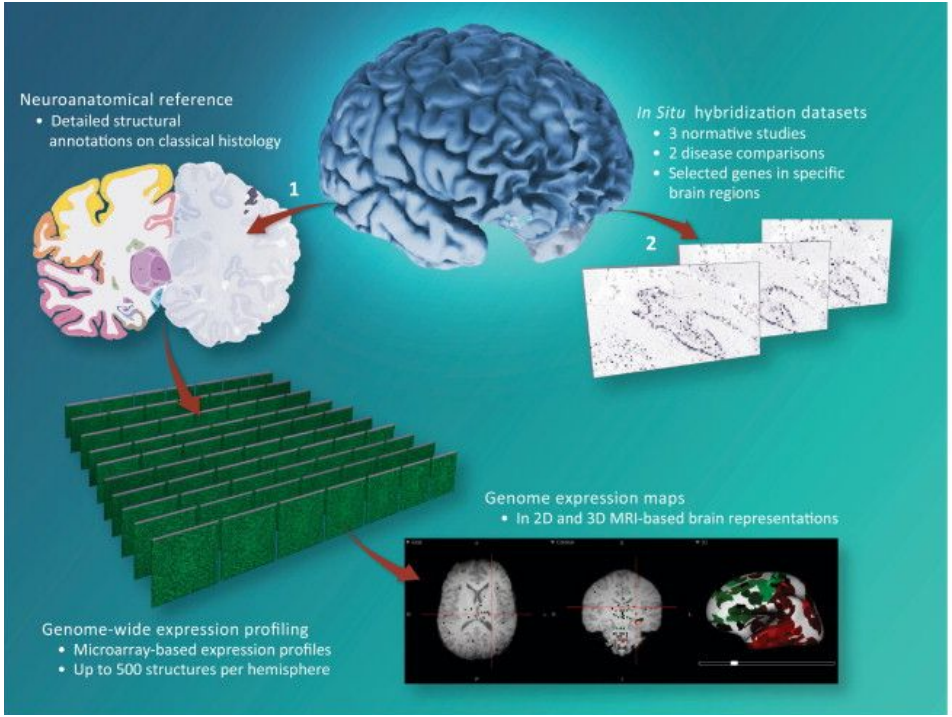
CANCER MOONSHOT

Cancer Moonshot



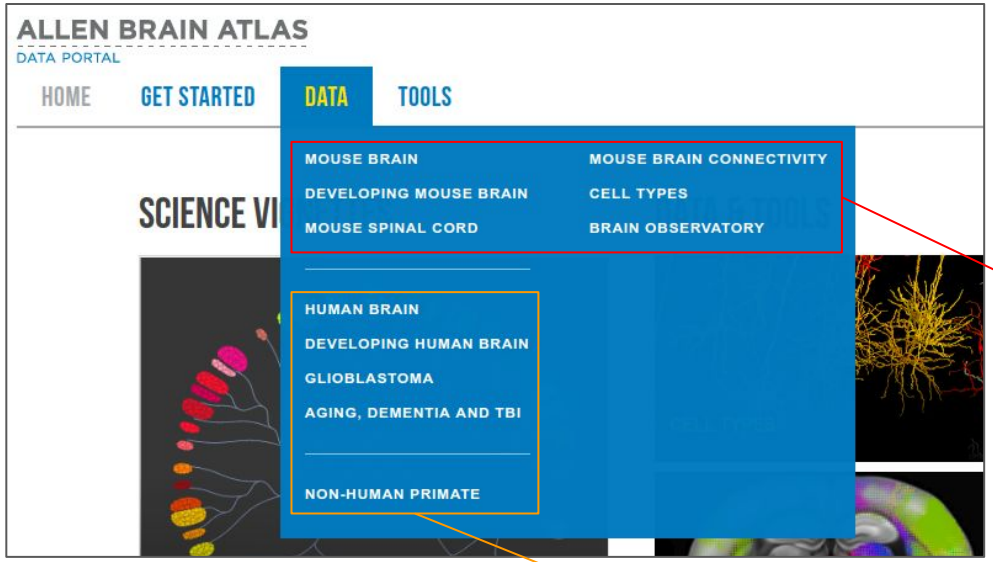
Pre-Cancer Atlas

ALLEN INSTITUTE *for* BRAIN SCIENCE
Fueling Discovery



- **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/

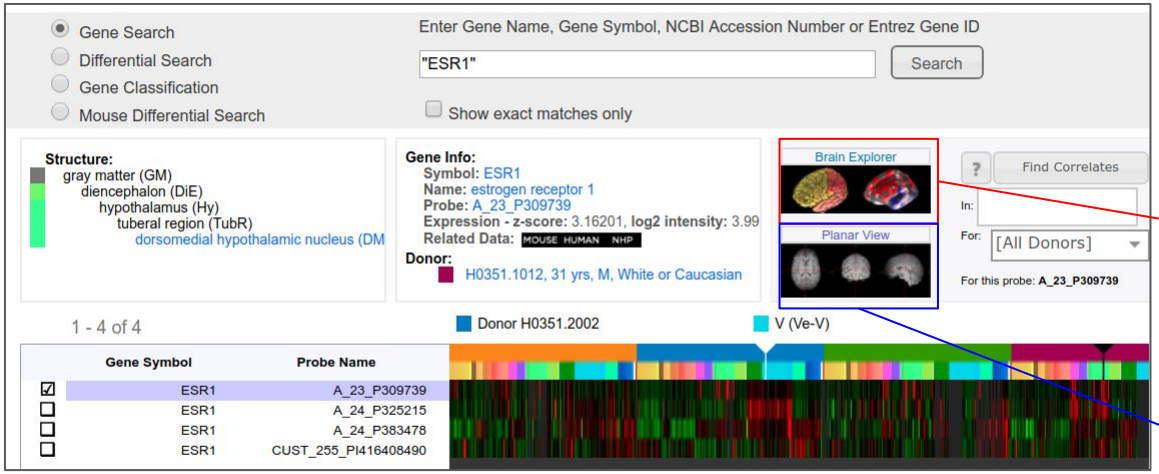


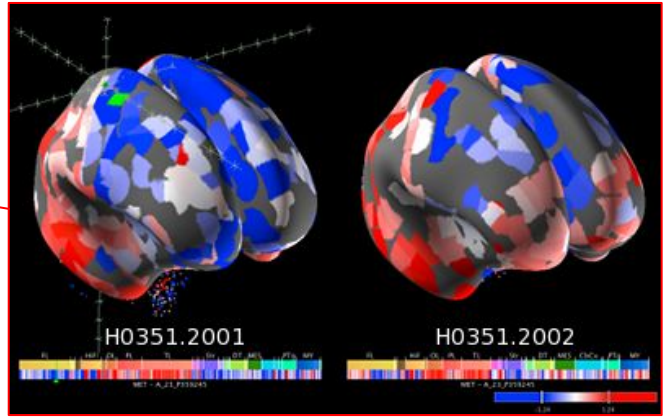
Mouse Brain data: ISH data
Developing Mouse Brain data: ISH data
Mouse Spinal Cord: ISH data
Mouse Brain Connectivity: Reconstruction using transgenic mice data
Cell Type: RNA-Seq data, Morphology and Electrophysiology data
Brain observatory: Physiological data transgenic mice

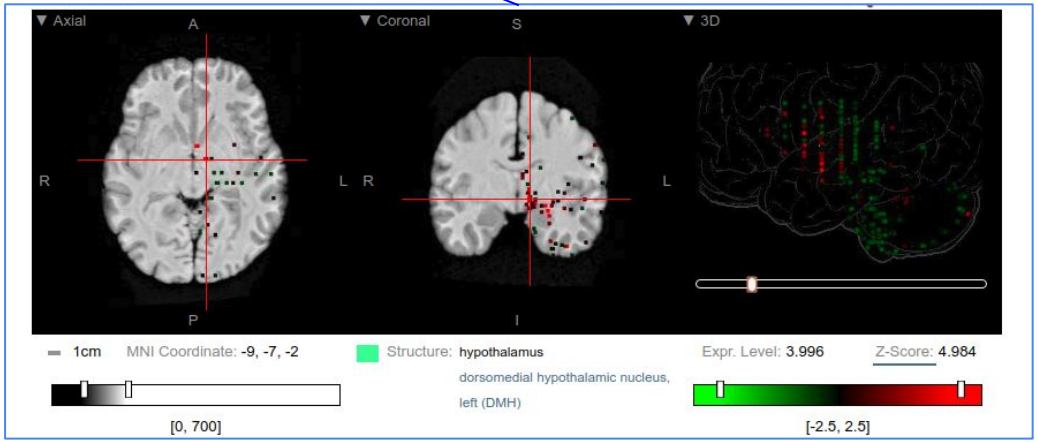
Human Brain data: Microarray and RNA-Seq data, MRI scan
Developing Human Brain data: Microarray, RNA-Seq, Exon microarray data
Glioblastoma Atlas: RNA-Seq, Clinical data
Aging, Dementia, TBI: RNA-Seq, Clinical data, IHC

Non-human primate data: Microarray data

Data visualization and integration







Trends in Neurosciences
CellPress
Volume 35, Issue 12, December 2012, Pages 711-714

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

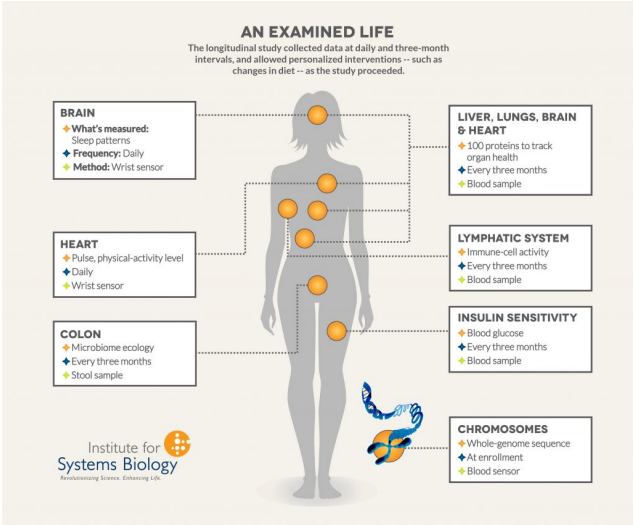
What else?

- Human Connectome Project
- Human Brain Project
- iPOP and P100

What's next?

100K
100K WELLNESS PROJECT
Institute for Systems Biology

100K Wellness Project



ENCODE

1000 Genomes
A Deep Catalog of Human Genetic Variation

GTEx

ALLEN INSTITUTE for BRAIN SCIENCE
Fueling Discovery

THE HUMAN PROTEIN ATLAS

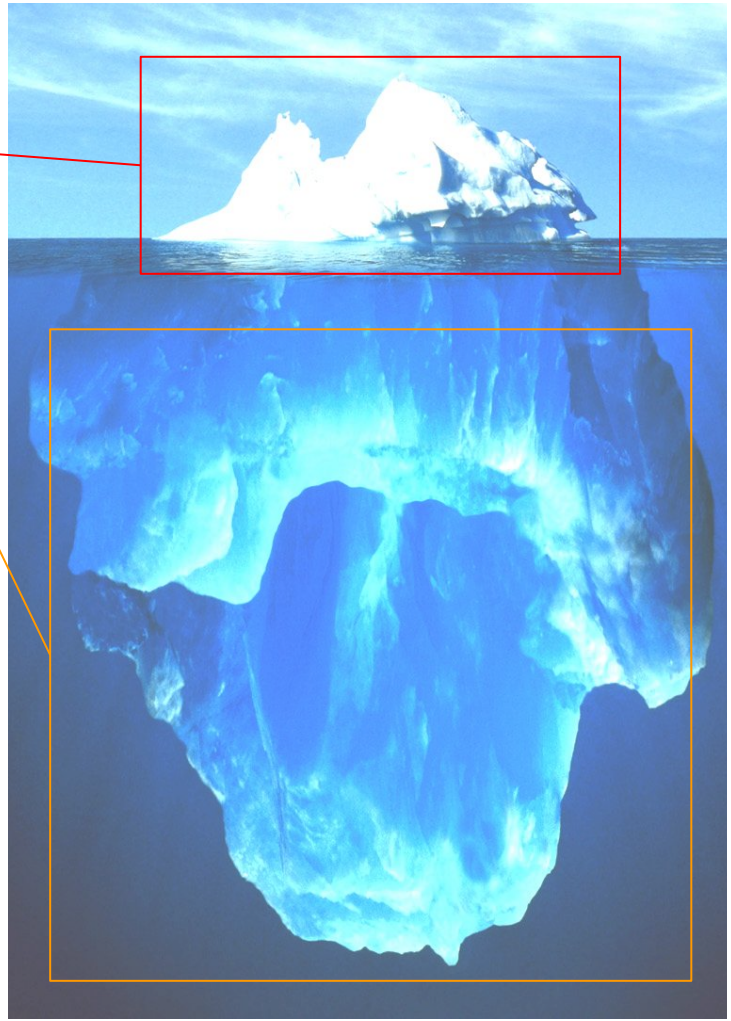
NIH THE CANCER GENOME ATLAS
National Cancer Institute
National Human Genome Research Institute

ROADMAP epigenomics PROJECT

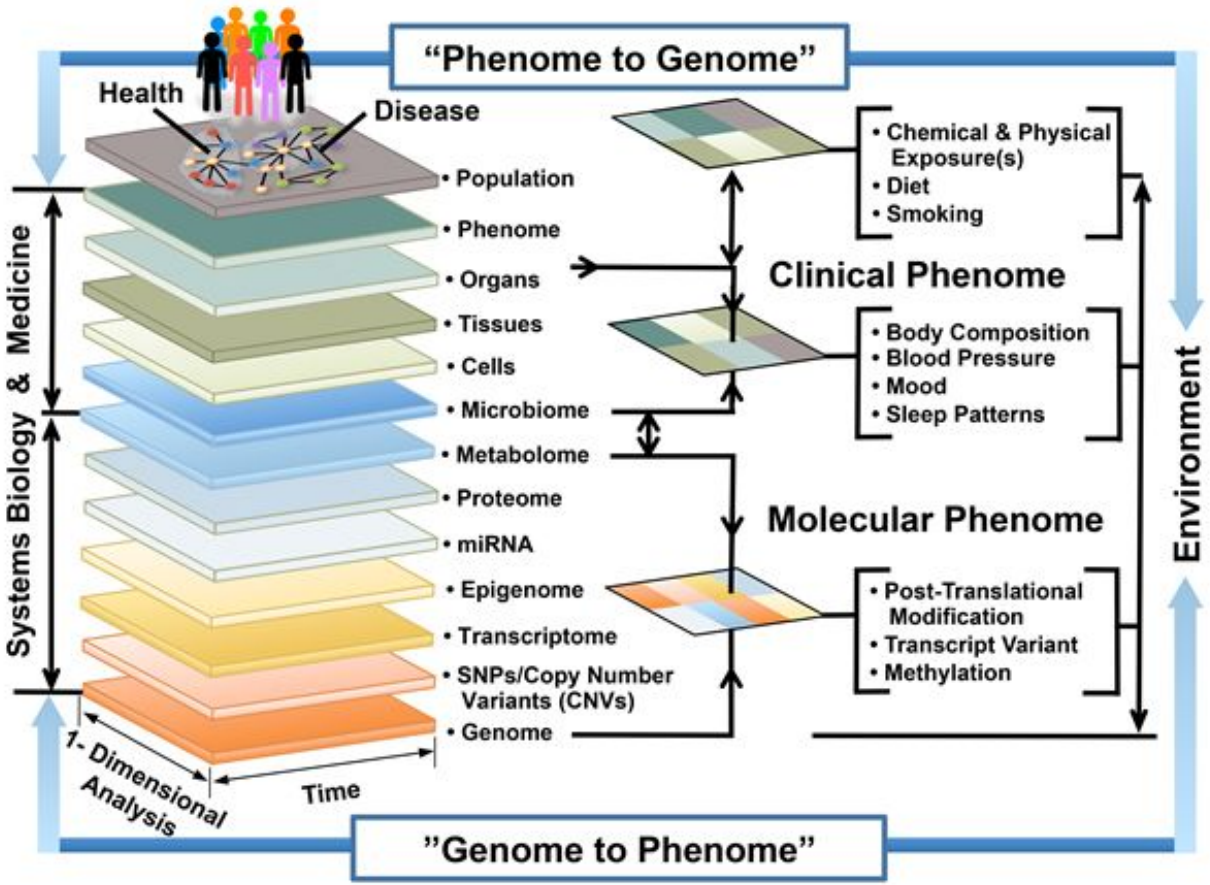
GEO
Gene Expression Omnibus

ArrayExpress

PubMed



- 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



From: *https://medicine.umich.edu/dept/dcmb/research*

- 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)