

Topic 6. Functional Genomics

Functional genomics: ENCODE, FANTOM, GTEx, Roadmap Epigenomics, Blueprint, IHEC. Functional genomics databases at EMBL-EBI: ArrayExpress and Expression Atlas. GO enrichment analysis.

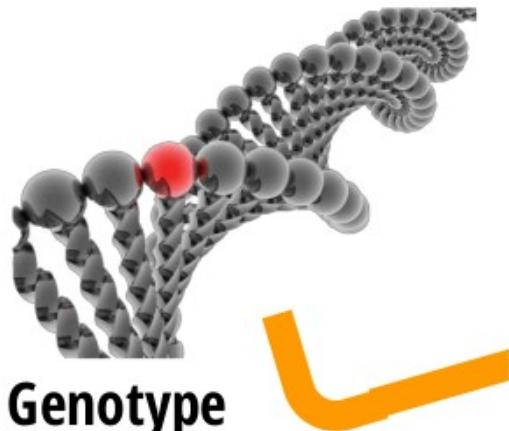
The Genotype to Phenotype challenge

Topic 5 – Genes & Genomes

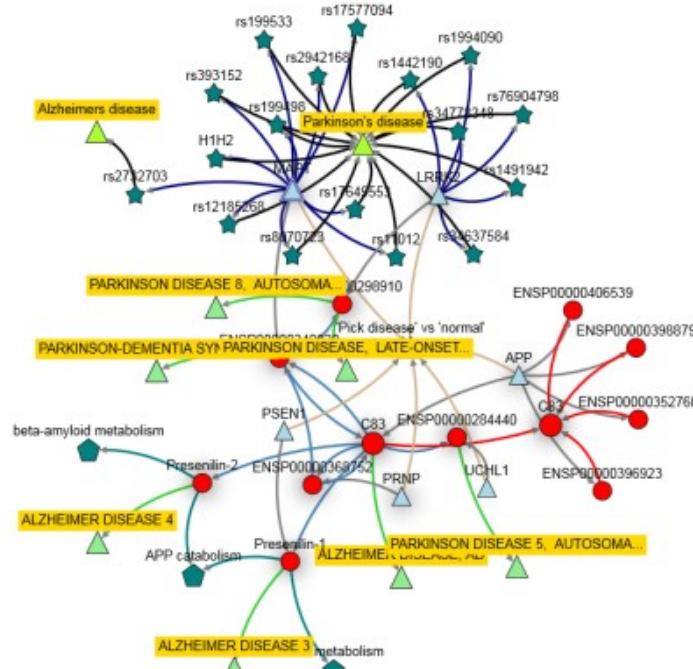
Topic 6 – Functional genomics

Topic 7 – Networks & Pathways

Topic 8 – Phenotypes & Diseases

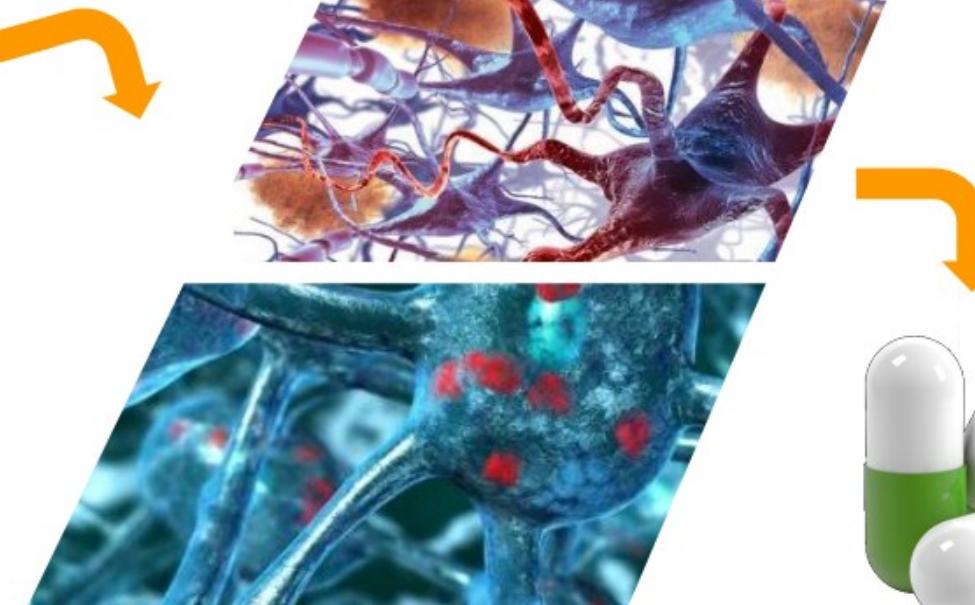


Genotype
GWAS, QTL



Biological Knowledge Discovery

Data selection, processing, transformation,
integration, interpretation



Phenotype

Alzheimer, Parkinson



Drugs

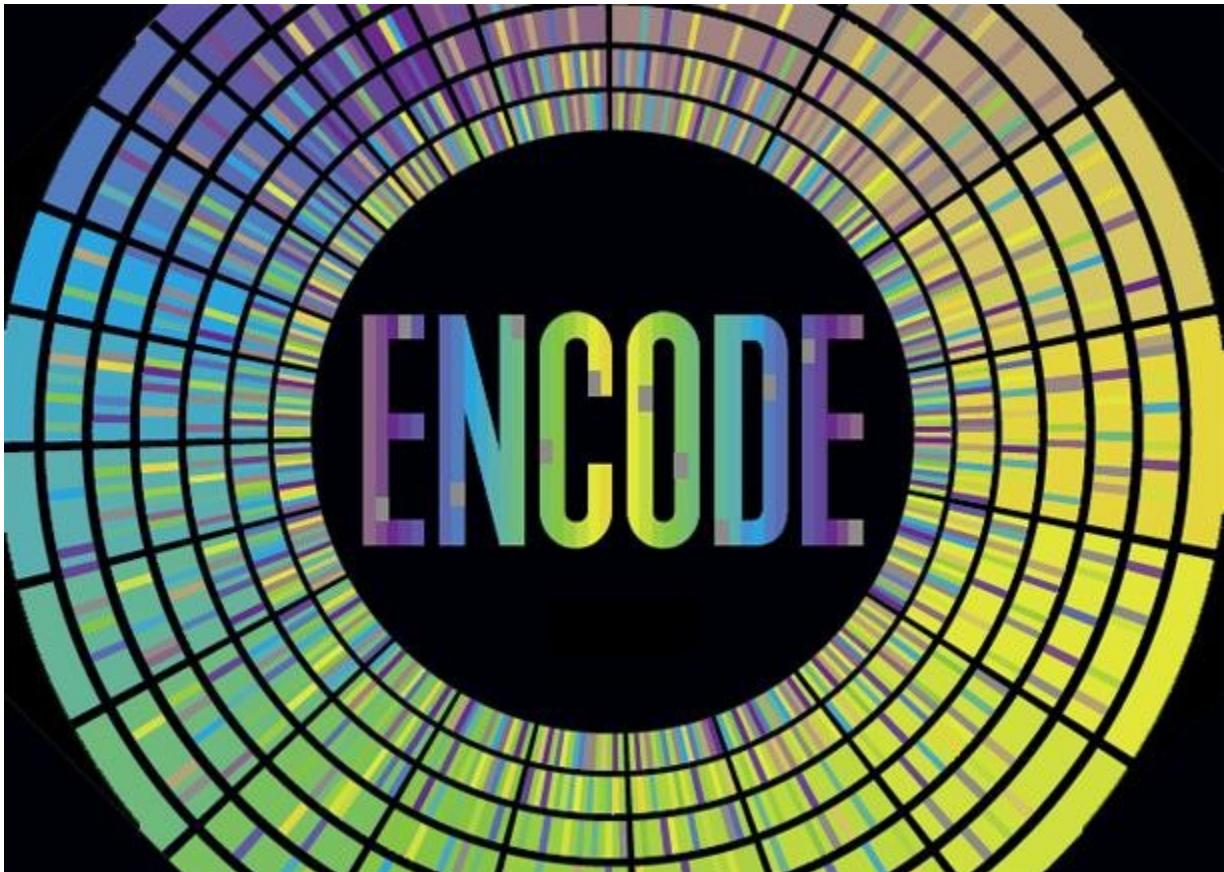
Precision medicine

What is Functional Genomics?

The aim of **functional genomics** is to understand the complex **relationship between genotype and phenotype on a global (genome-wide) scale** in an attempt to answer relevant biological questions such as:

- When and where are genes expressed?
- How do gene expression levels differ in various cell types and states?
- What are the functional roles of different genes and in what cellular processes do they participate?
- How are genes **regulated**? Where are the active gene promoters in a particular cell type?
- How do genes and gene products **interact**?
- How does gene expression change in various diseases or following a treatment?

ENCODE



- 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 **catalog all 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.

<https://www.encodeproject.org/>

Pilot phase

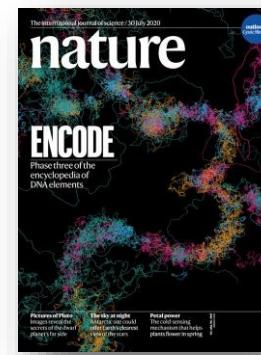
2003-2007

**1% of the human genome,
~30 Mb, 44 regions**

Production phase

2007-2020

Whole genome

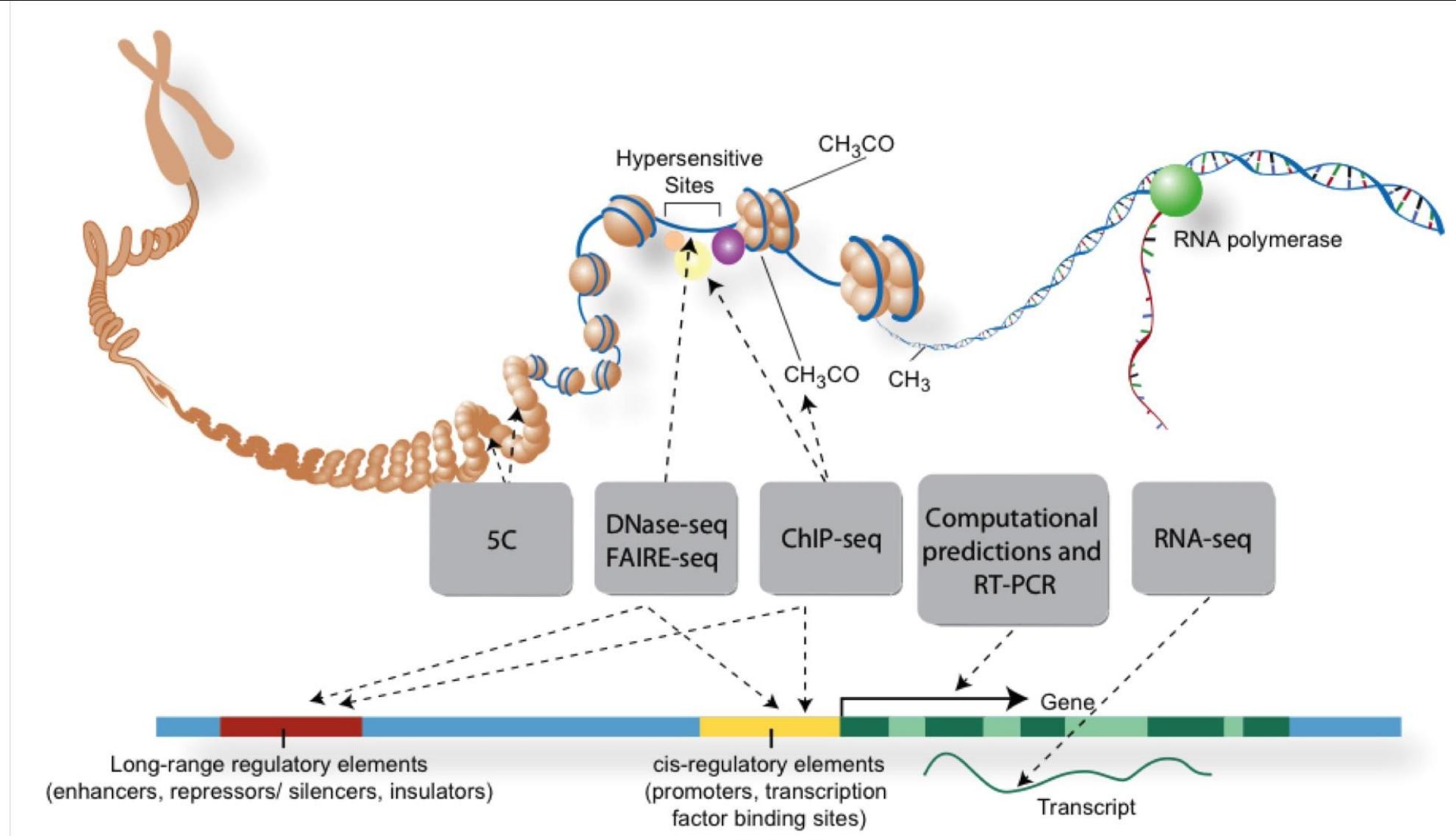


Phase III

ENCODE Project Consortium (July 2020)

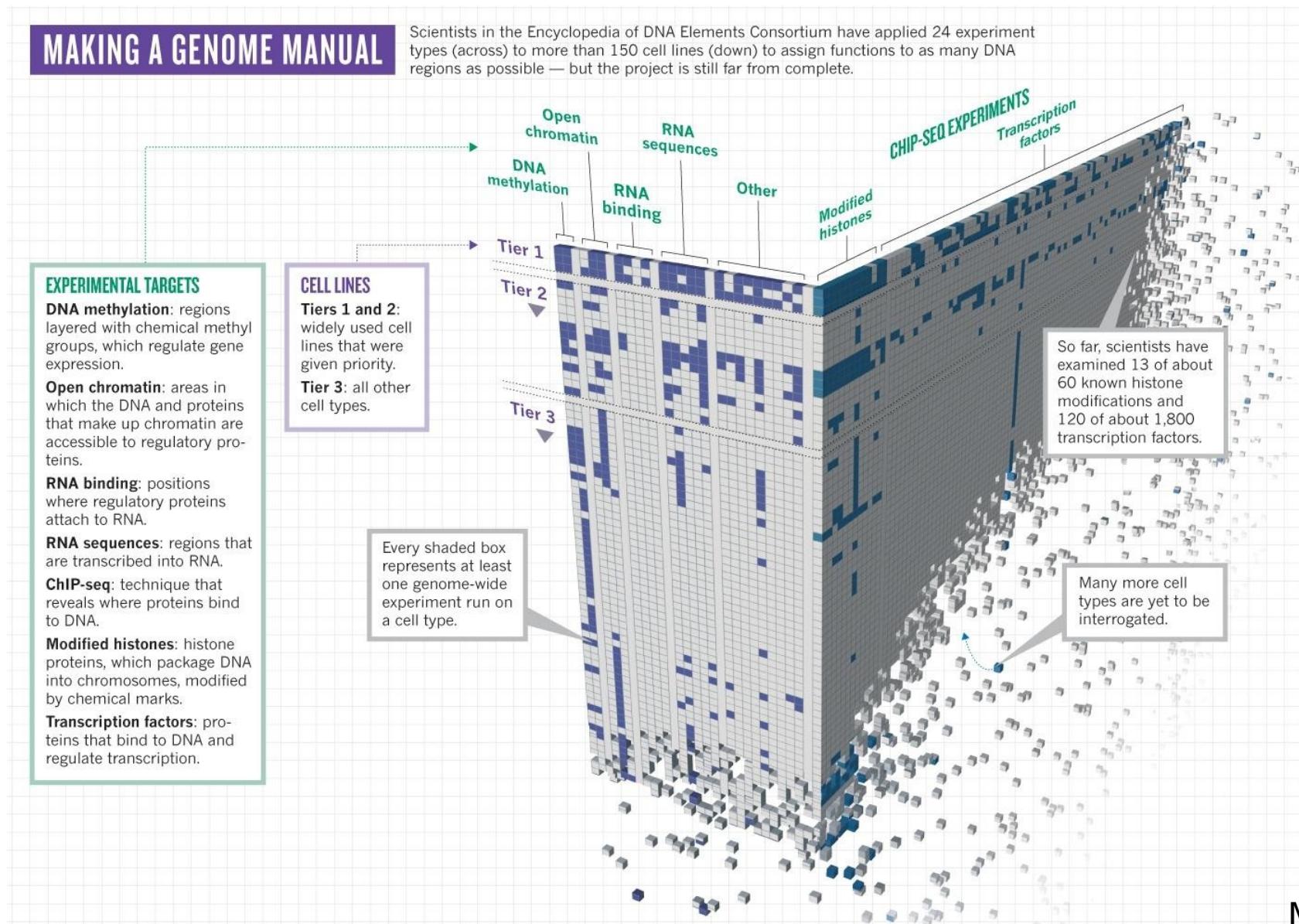
Nature 583: Special issue.

ENCODE: assays and methods



ENCODE Project Consortium (2011) PLoS Biology 9:e1001046

ENCODE: assays and methods



Maher (2012) Nature

ENCODE: MAPT gene



Region Search

Enter any one of human Gene name, Symbol, Synonyms, Gene ID, HGNC ID, coordinates, rsid, Ensemble ID

MAPT

- MAPT-IT1 (Homo sapiens)
 - MAPT-AS1 (Homo sapiens)
 - MAPT (Homo sapiens)
 - STH (Homo sapiens)
- : (chr17:45894382-46028334) +/- 2kb

Homo sapiens ▾

Search

Datasets

Genome Browser

Showing 25 of 3722

Assay	
ChIP-seq	2037
DNase-seq	1331
ATAC-seq	354

Biosample term

Target

CTCF	364
POLR2A	60
POLR2AphosphoS5	41
RAD21	37
JUN	27

TF ChIP-seq in K562

Homo sapiens K562

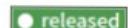
Target: FOXM1

Lab: Michael Snyder, Stanford

Project: ENCODE

Experiment

ENCSR429QPP



3

TF ChIP-seq in MCF-7

Homo sapiens MCF-7

Target: SIN3A

Lab: Michael Snyder, Stanford

Project: ENCODE

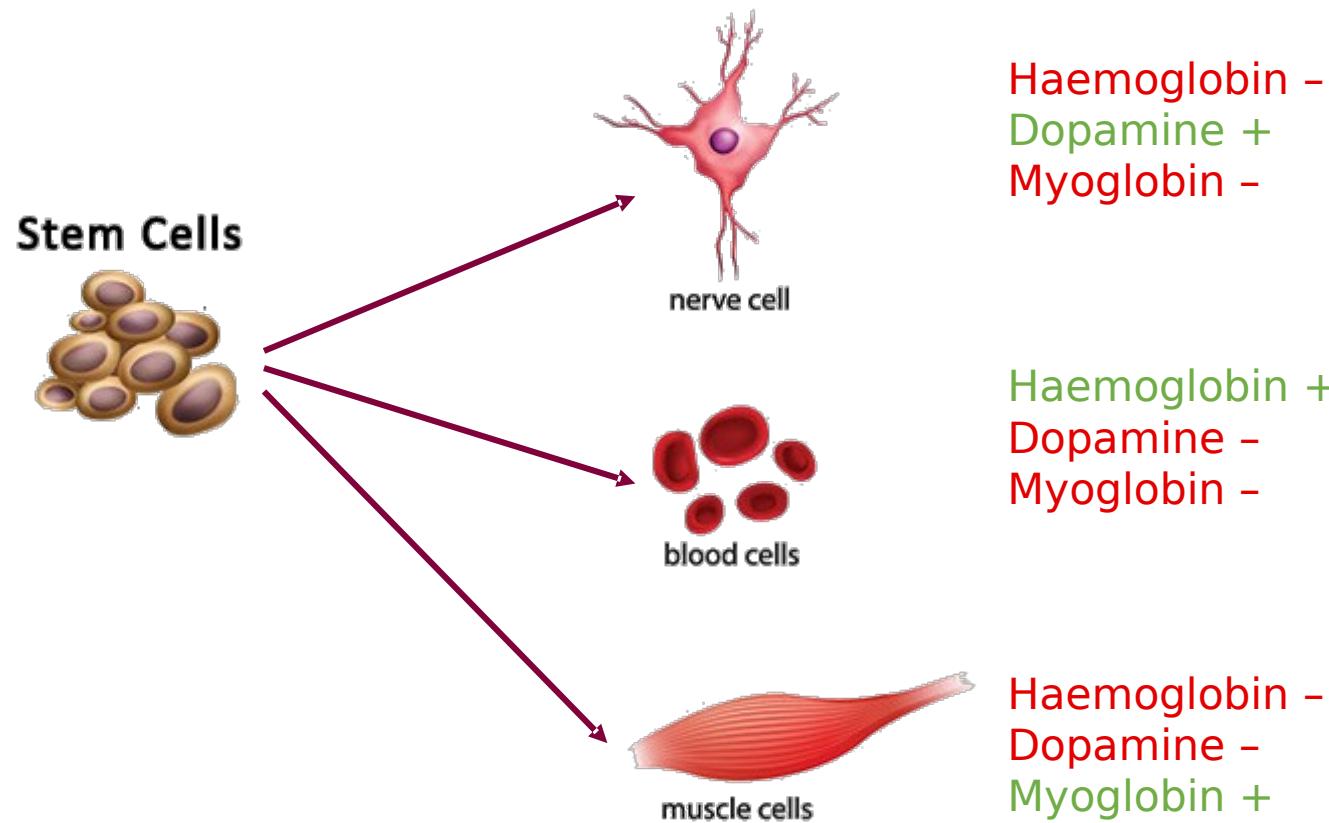
Experiment

ENCSR468LUO



[https://www.encodeproject.org/region-search/?region=MAPT%20\(Homo%20sapiens\)&annotation=chr17:45894382-46028334&genome=GRCh38](https://www.encodeproject.org/region-search/?region=MAPT%20(Homo%20sapiens)&annotation=chr17:45894382-46028334&genome=GRCh38)

FANTOM5



- ~400 distinct cell types BUT one single genome
- Each cell type expresses a restricted subset of genes
- FANTOM5 have systematically investigated what are the sets of genes used in virtually all cell types across the human body, and the genomic regions which determine where the genes are read from

<http://fantom.gsc.riken.jp/5/>
Kawaji et al. (2017) Scientific Data

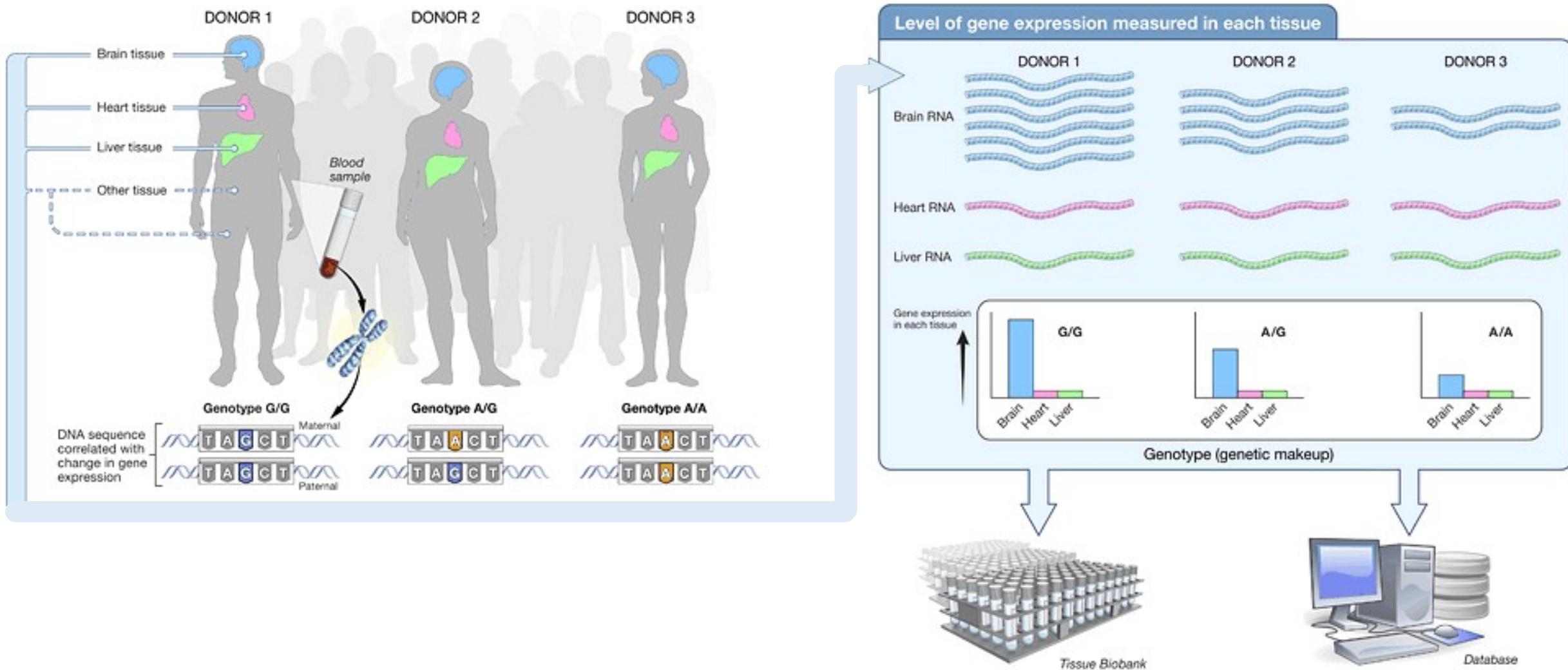
GTEX Portal

Fase III
GTEx Consortium (September 2020)
Science 369: Special issue.

- The **Genotype-Tissue Expression (GTEx)** project aims to study human gene expression and regulation and its relationship to genetic variation. By analyzing global RNA expression within individual tissues and treating the expression levels of genes as quantitative traits, variations in gene expression that are highly correlated with genetic variation can be identified as **expression quantitative trait loci**, or **eQTLs**.

<https://www.gtexportal.org>

Lonsdale et al. (2013) Nature Genetics 45:580-585



<https://www.gtexportal.org>

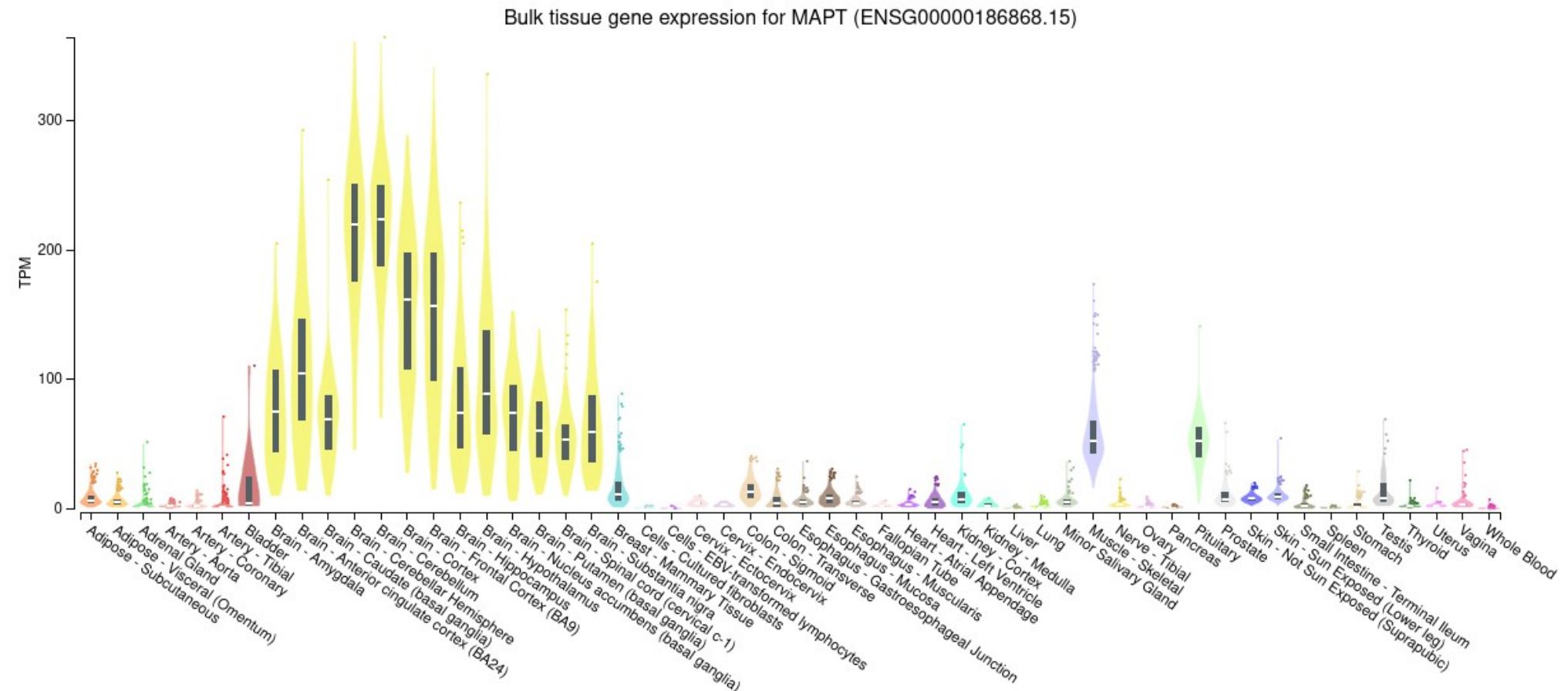
Lonsdale et al. (2013) Nature Genetics 45:580-585.

GTEx: MAPT gene

Bulk tissue gene expression for MAPT (ENSG00000186868.15)

Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)

Data processing and normalization 



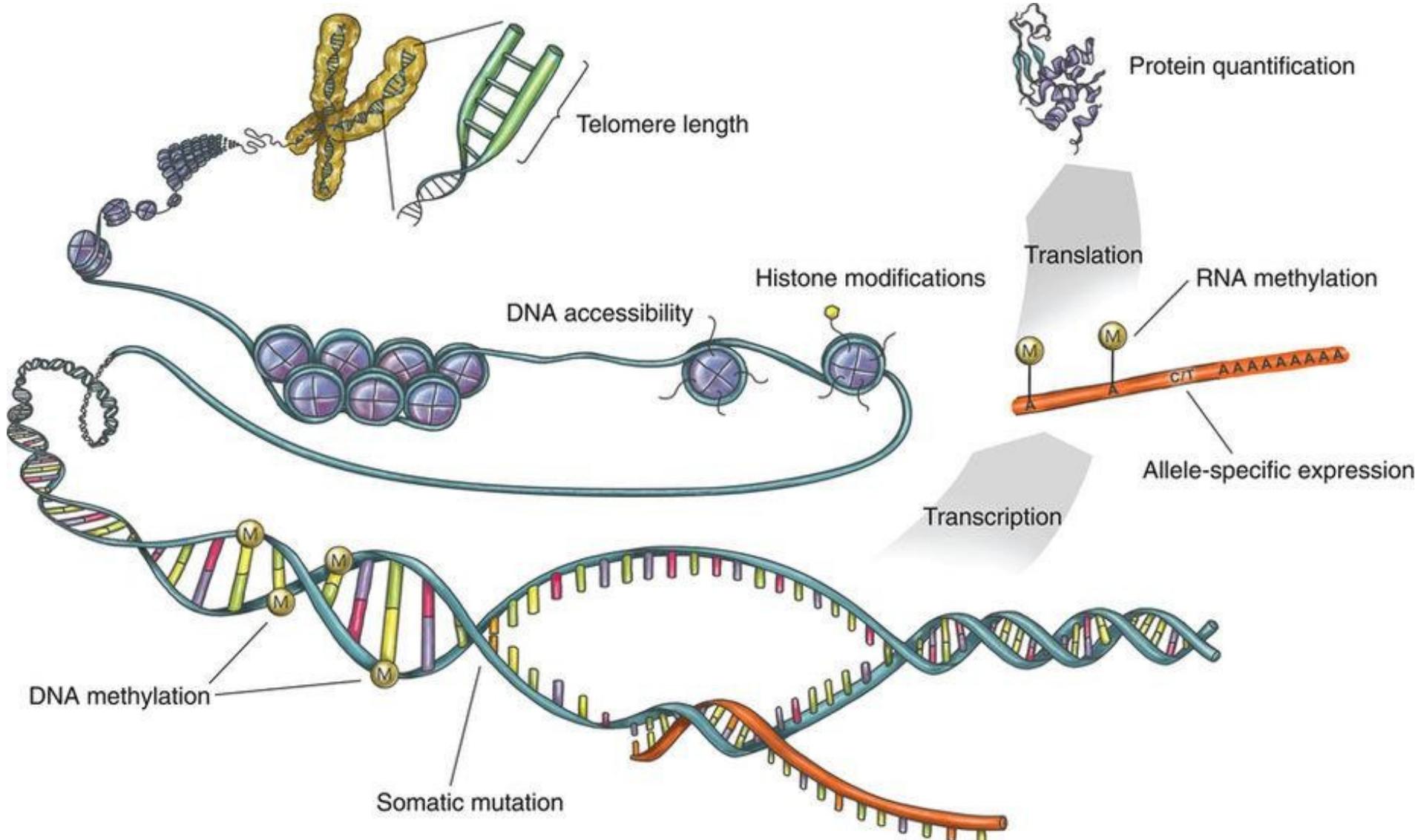
<https://www.gtexportal.org/home/gene/MAPT>

GTEx: MAPT gene

Isoform Expression of MAPT: ENSG00000186868.15 microtubule associated protein tau [Source:HGNC Symbol;Acc:HGNC:6893]

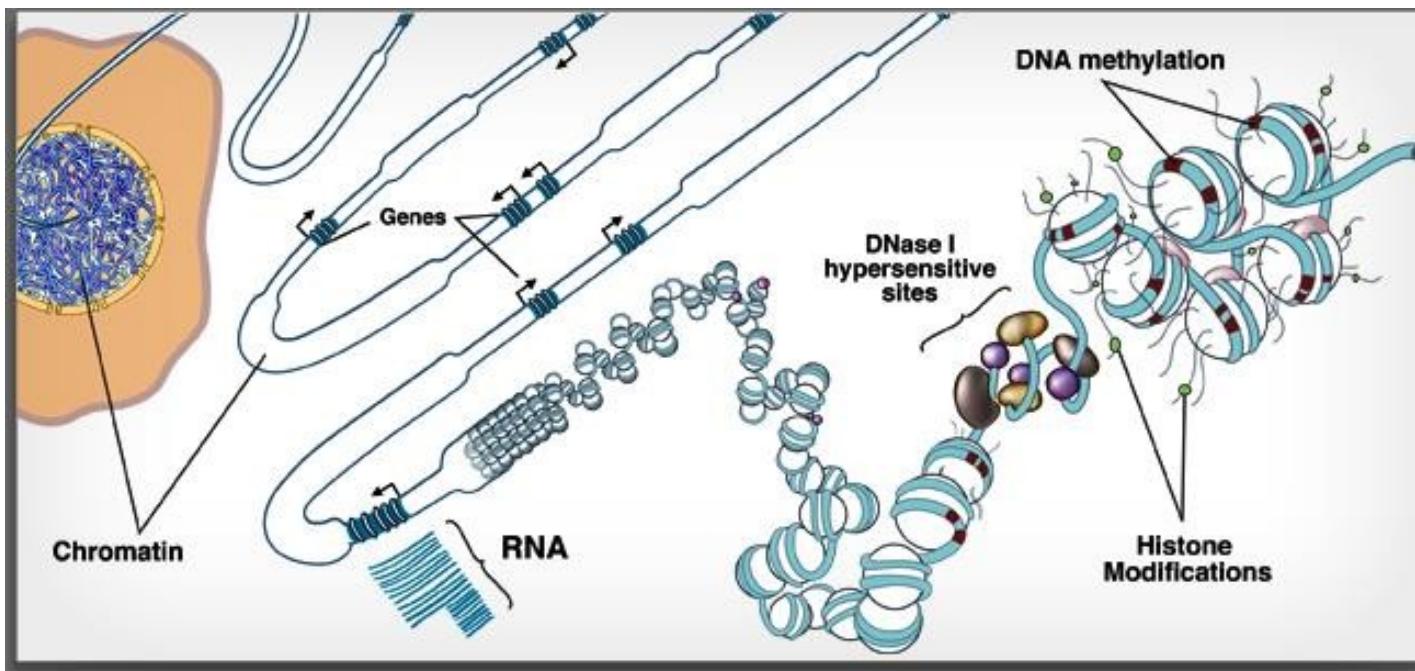


Enhancing GTEx (eGTEx)



eGTEx Project (2017) Nature Genetics.

Roadmap Epigenomics

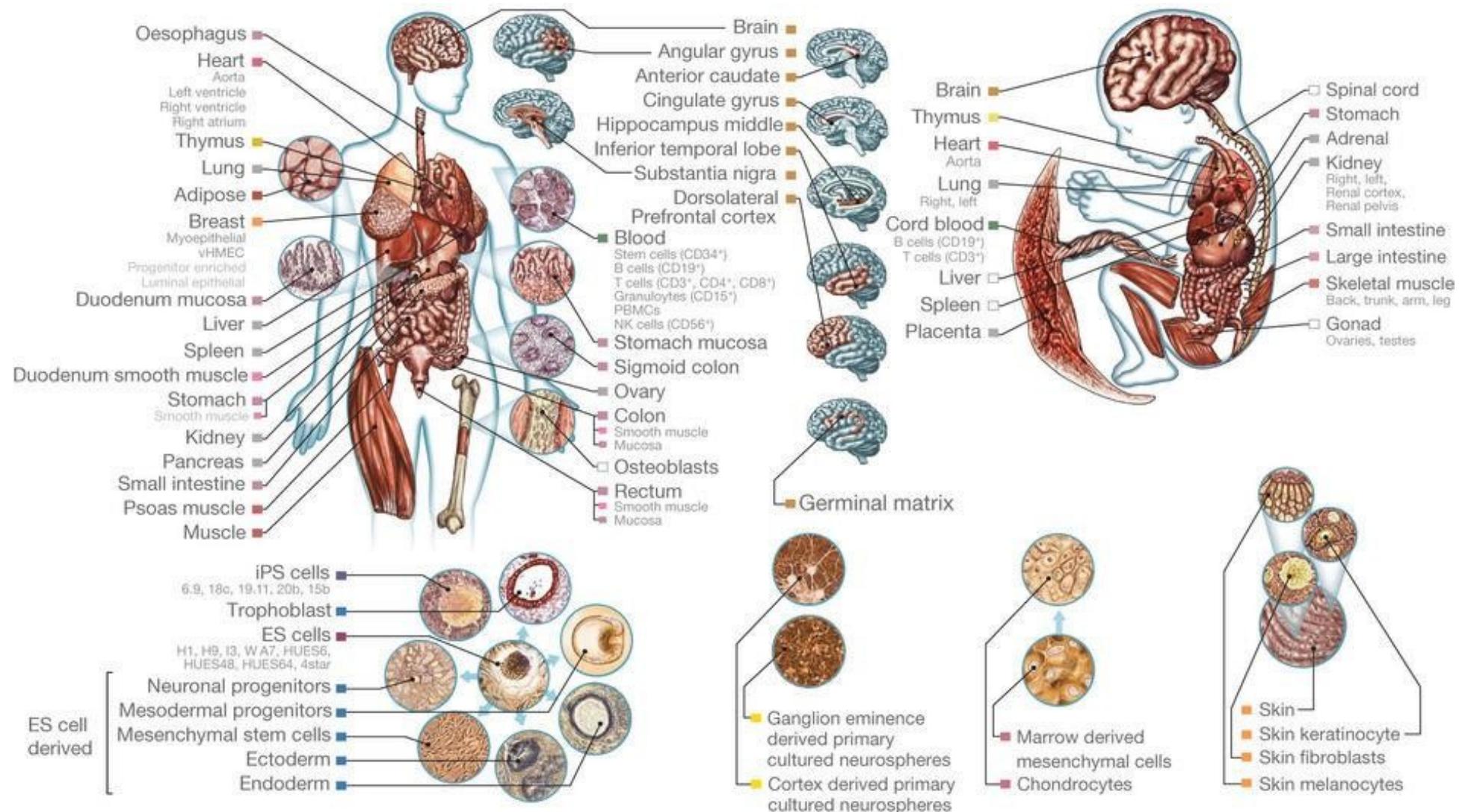


<http://www.roadmapepigenomics.org/>

Roadmap Epigenomics Consortium (2015) Nature 518:317-330.

- The NIH Roadmap Epigenomics Mapping Consortium was launched with the goal of producing a public resource of **human epigenomic data** (DNA methylation, histone modifications, chromatin accessibility and small RNA transcripts) in **stem cells and primary ex vivo tissues** to catalyze basic biology and disease-oriented research.
- The Consortium expects to deliver a collection of **111 normal epigenomes** that will provide a framework or reference for comparison and integration within a broad array of future studies.

Roadmap Epigenomics: 111 reference human epigenomes



<http://www.roadmapepigenomics.org/>

Roadmap Epigenomics Consortium (2015) Nature 518:317-330.

Roadmap Epigenomics: MAPT gene

ROADMAP
epigenomics
PROJECT

HOME PARTICIPANTS BROWSE DATA PROTOCOLS COMPLETE EPIGENOMES TOOLS PUBLICATIONS

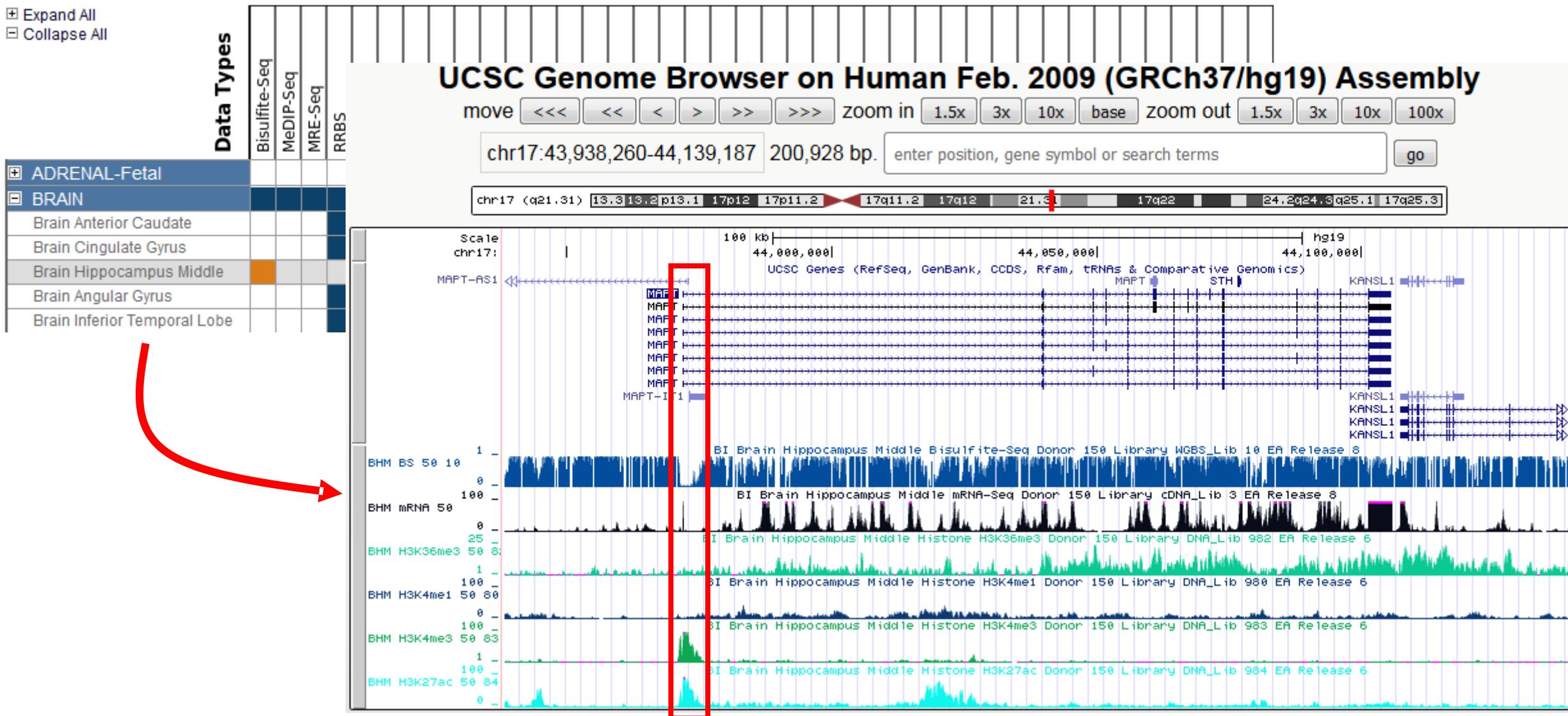
Data

DATA TABLE	ADULT	BRAIN	FETAL	STEM CELLS
<input checked="" type="checkbox"/> Expand All <input type="checkbox"/> Collapse All	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> <input checked="" type="checkbox"/> Bisulfite-Seq <input type="checkbox"/> MeDIP-Seq <input type="checkbox"/> MRE-Seq <input type="checkbox"/> RRBS <input checked="" type="checkbox"/> DNase <input type="checkbox"/> DGF <input type="checkbox"/> mRNA-Seq <input type="checkbox"/> smRNA-Seq <input type="checkbox"/> ChIP-Input <input type="checkbox"/> H3K27me3 <input type="checkbox"/> H3K36me3 <input type="checkbox"/> H3K4me1 <input type="checkbox"/> H3K4me3 <input type="checkbox"/> H3K9ac <input type="checkbox"/> H3K9me3 <input type="checkbox"/> H3K27ac <input type="checkbox"/> H2AK5ac <input type="checkbox"/> H2AK9ac <input type="checkbox"/> H2AZ <input type="checkbox"/> H2BK120ac <input type="checkbox"/> H2BK12ac <input type="checkbox"/> H2BK15ac <input type="checkbox"/> H2BK20ac <input type="checkbox"/> H3K14ac <input type="checkbox"/> H3K18ac <input type="checkbox"/> H3K23ac <input type="checkbox"/> H3K4ac <input type="checkbox"/> H3K4me2 <input type="checkbox"/> H3K56ac <input type="checkbox"/> H3K79me1 <input type="checkbox"/> H3K79me2 <input type="checkbox"/> H4K20me1 <input type="checkbox"/> H4K5ac <input type="checkbox"/> H4K8ac <input type="checkbox"/> H4K12ac <input type="checkbox"/> H4K91ac <input type="checkbox"/> H3K23me2 <input type="checkbox"/> H2BK5ac <input type="checkbox"/> H3K9me1 <input type="checkbox"/> H3T11oh </div> </div>			

<http://www.roadmapepigenomics.org/data/tables/all>

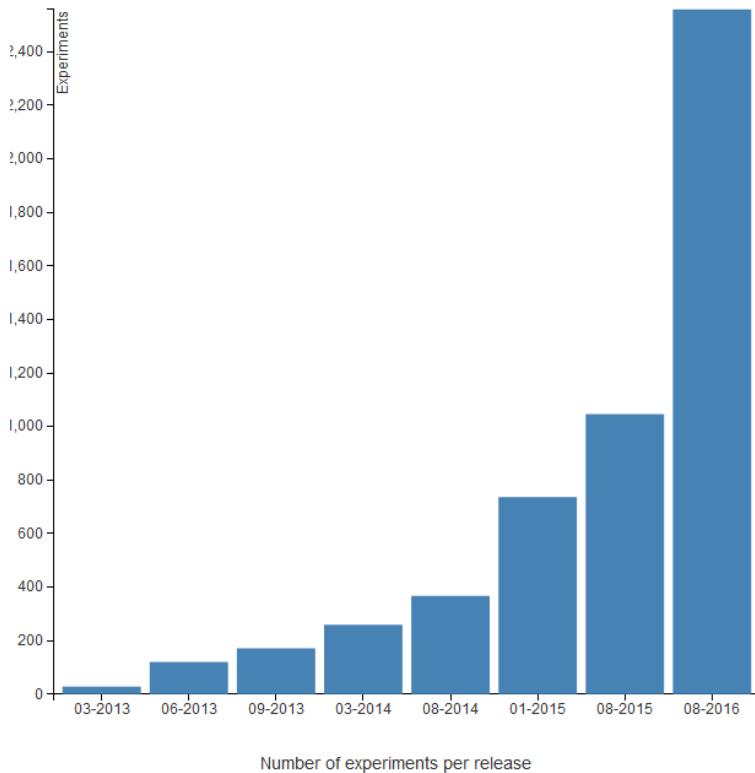
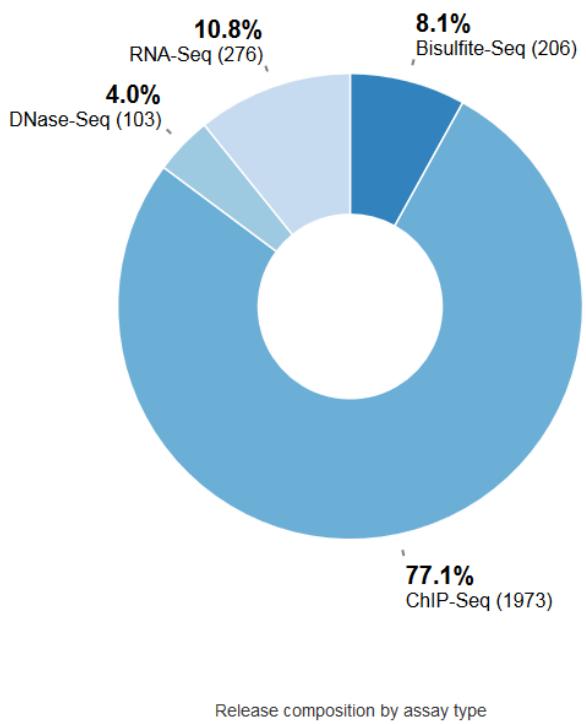
Roadmap Epigenomics: MAPT gene

[+] Expand All
[-] Collapse All



http://www.epigenomebrowser.org/cgi-bin/hgTracks?db=hg19&position=chr17%3A43938260-44139187&hgsid=201579_xne7Ya8ytQLfcMXjzaLHbRBAtodV

Blueprint



- The goal of the BLUEPRINT project is to apply highly sophisticated functional genomics analysis on a clearly defined set of **primarily human samples from healthy and diseased individuals** and to provide **at least 100 reference epigenomes** to the scientific community. This resources-generating activity is complemented by research into blood-based diseases, including common leukemias and autoimmune disease (Type 1 Diabetes).

<http://www.blueprint-epigenome.eu/>

Roadmap Epigenomics Consortium (2015) Nature 518:317-330.

International Human Epigenome Consortium (IHEC)

[Home](#)[Data Grid](#)[Download](#)[Community](#)[Help](#)[About](#)[IHEC Main Site](#)

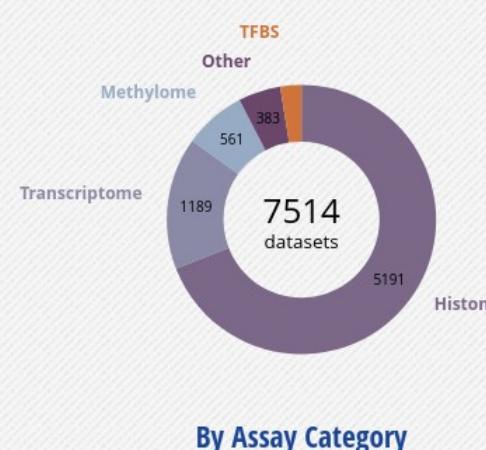
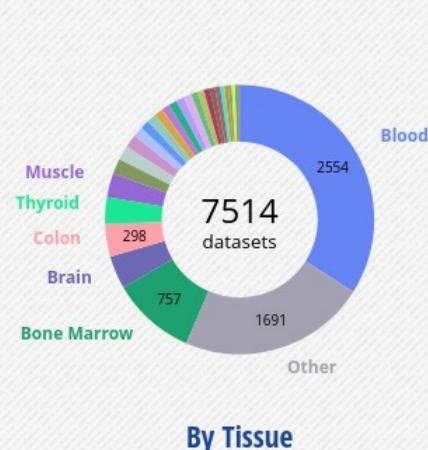
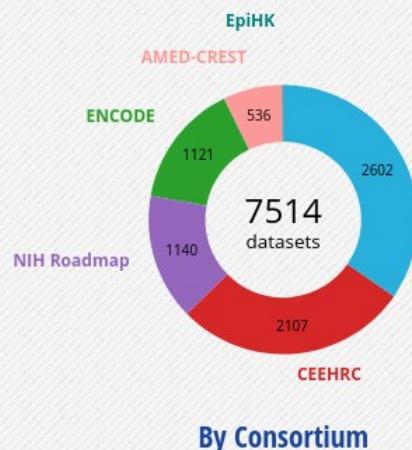
Welcome to the IHEC Data Portal

You may select IHEC datasets in these charts to view them in the Data Grid. Alternatively you can download or display them in a Genome Browser.

Genome : **Human (hg38)**

Display : **Datasets**

Build : **2020-10**



- IHEC makes available comprehensive sets of reference epigenomes relevant to health and disease.

- The IHEC Data Portal can be used to view, search and download data already released by the different IHEC-associated projects

<http://epigenomesportal.ca/ihec/>

IHEC Special Edition, Volumes 1 & 2 (2016) Cell 167

IHEC: MAPT gene

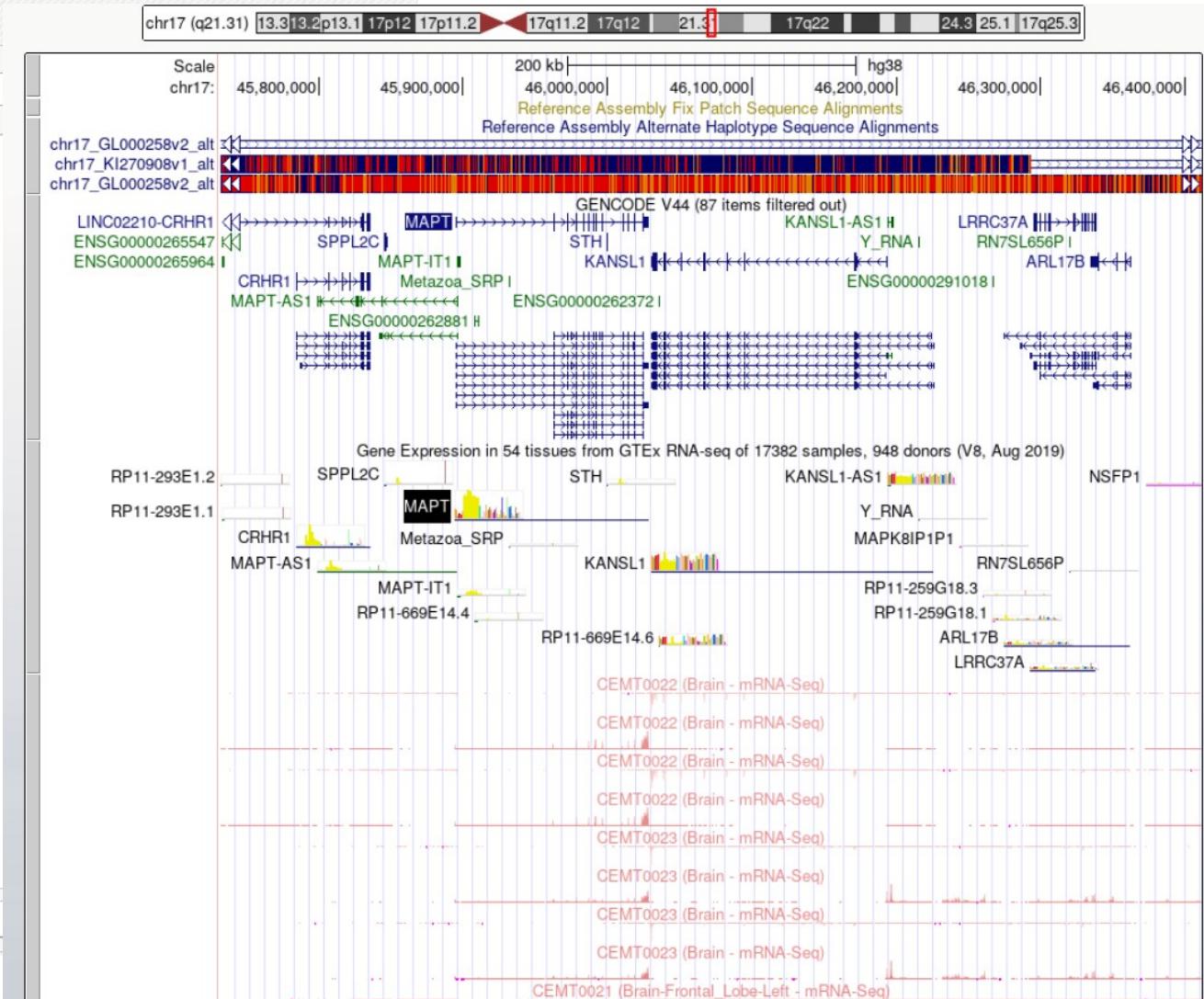
Data Grid



<https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr17>

IHEC: MAPT gene

Data Grid



<https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr17:10600000-10700000>

Omics Discovery Index (OmicsDI)

OmicsDI Browse Submit Data Databases API Help + Login

Organism, repository, gene, tissue, accession

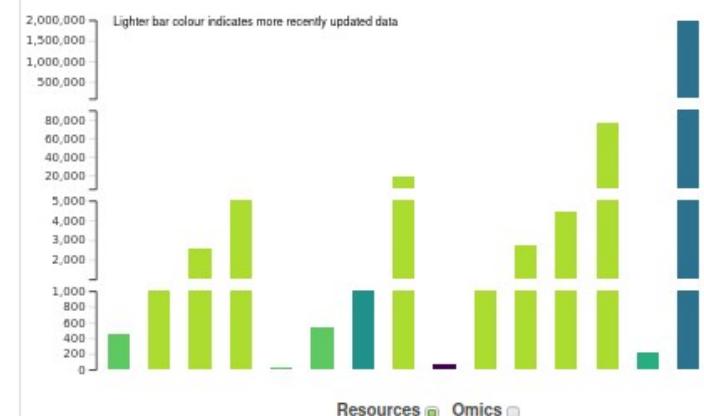
Advanced Search

revealed first potential mass
mechanisms through related further
pathways including disease
target dataset samples more
patients expression novel
studies sequencing analysis sample
derived molecular regulation

Description Sample Data



Tissues Organisms Diseases



Resources Omics

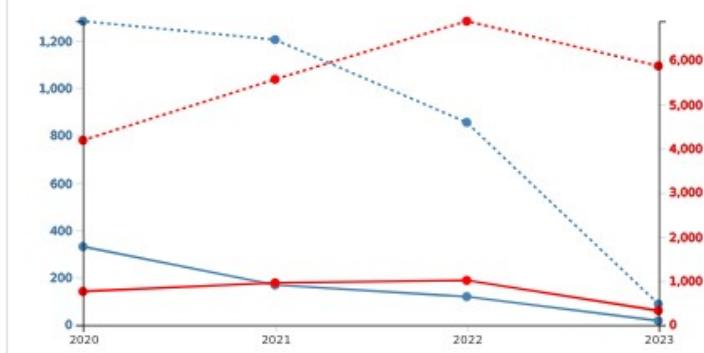
Latest Datasets

- Oct 29 '23 Proteomic analysis of small extracellular vesicles purified by three methods
- Oct 28 '23 The Liquid Chromatography/Tandem Mass Spectrometry-Based Quantification of Angiotensin II Receptor Blocker (ARB) Metabolites in Human Urine
- Oct 28 '23 Identification of SENP5 target proteins by endogenous SUMO2/3 IP-MS
- Oct 28 '23 HA-GABARAPL1 immunoprecipitation with siRNA-mediated NPM1 depletion
- Oct 28 '23 Identification of SENP3 and SENP5 target proteins by endogenous SUMO2/3 IP-MS
- Oct 28 '23 Identification of UTP14A interactors by endogenous UTP14A IP-MS
- Oct 28 '23 Two novel Bartonella (sub)species isolated from edible dormice (*Glis glis*): Hints of cultivars
- Oct 28 '23 Whole cell proteome of OCI-AML3 FKBP12(F36V) cells with dTAG-13 treatment
- Oct 28 '23 Whole cell proteome of U2OS cells or Saos-2 cells depleted for SENP3
- Oct 27 '23 iCstr1116FB23: Genome-scale metabolic model of *Corynebacterium striatum* strain FD...

Most Accessed Datasets

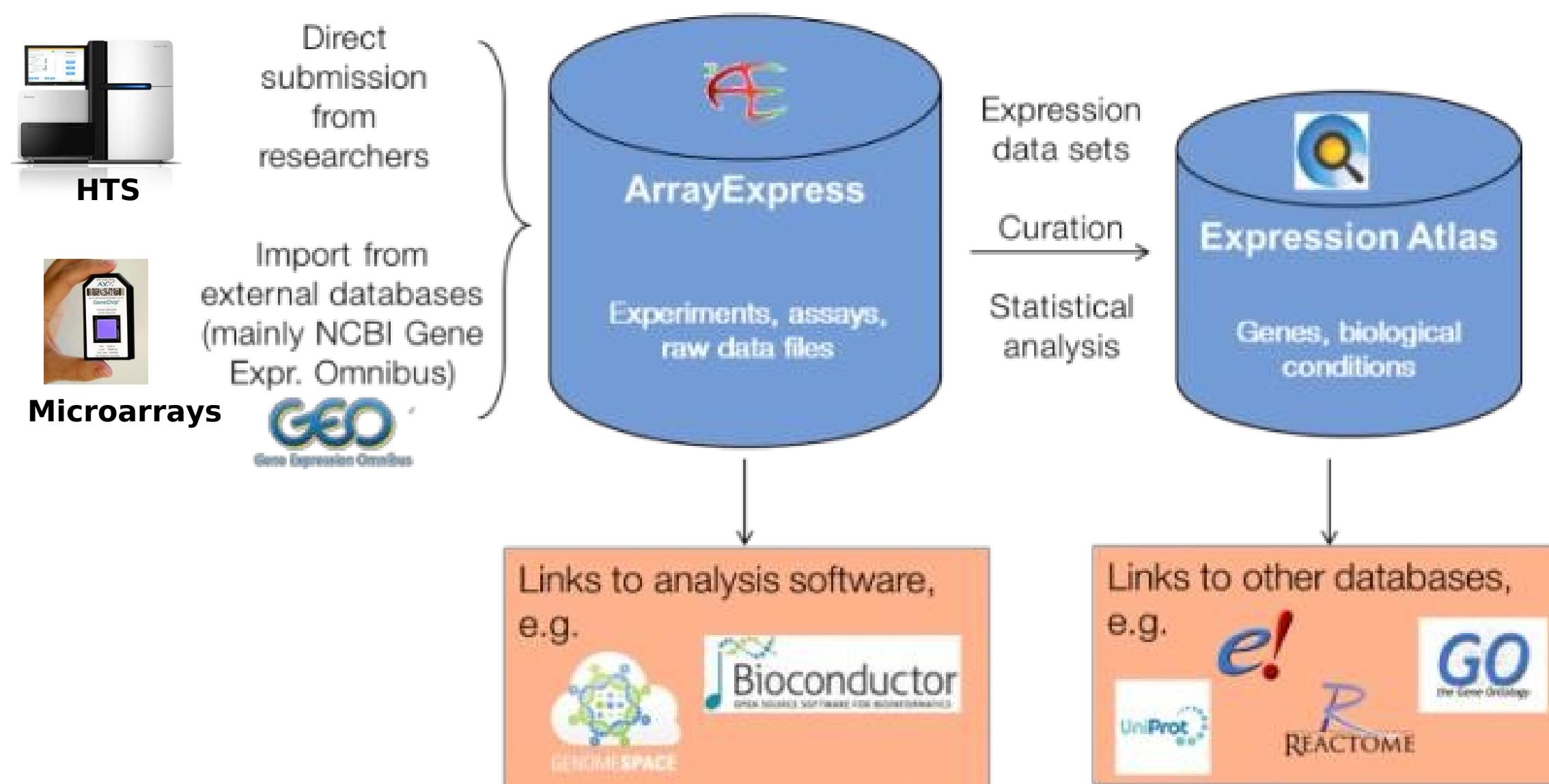
- 8806 58C
- 15787 Kholodenko1999 - EGFR signaling
- 9440 A metabolomic study of urinary changes in type 2 diabetes in human compared to the control
- 9189 Transcription profiling of *Streptococcus gordonii* wildtype and Brf- deletion and insertion mutant
- 8982 WTCCC case-control study for Bipolar Disorder
- 8833 E-MTAB-2037
- 12772 Ndairou2020 - early-stage transmission dynamics of COVID-19 in Wuhan
- 5373 Identification of proteins interacting with lncRNA Braveheart in mouse using protein microarray
- 4678 Cytokine interactome alterations promote oropharyngeal Epstein-barr virus (EBV) replication
- 4365 Elowitz2000 - Repressilator
- 3997 Iridovirus and Microsporidian Linked to Honey Bee Colony Decline
- 3987 Swainston2016 - Reconstruction of human metabolic network (Recon 2.2)

New Datasets Per Year



<http://www.omicsdi.org>

Functional genomics at EMBL-EBI: ArrayExpress and Expression Atlas

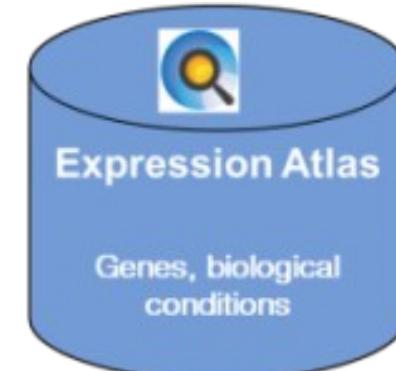


Functional genomics at EMBL-EBI: ArrayExpress and Expression Atlas



Raw cooking ingredients

Create your own dishes or
make a meal out of them



CANTEEN MENU

No time/skills/tools

Canteen with readily available
pre-defined dishes

BioStudies – one package for all the data supporting a study

The BioStudies database holds descriptions of biological studies, links to data from these studies in other databases at EMBL-EBI or outside, as well as data that do not fit in the structured archives at EMBL-EBI. The database can accept a wide range of types of studies described via a simple format. It also enables manuscript authors to submit supplementary information and link to it from the publication.



Functional genomics data

BIOSTUDIES / ARRAYEXPRESS

ArrayExpress - Functional Genomics Data

The functional genomics data collection (ArrayExpress), stores data from high-throughput functional genomics experiments, and provides data for reuse to the research community.

In line with community guidelines, a study typically contains metadata such as detailed sample annotations, protocols, processed data and raw data. Raw sequence reads from high-throughput sequencing studies are brokered to the European Nucleotide Archive (ENA), and links are provided to download the sequence reads from ENA. Data can be submitted to the ArrayExpress collection through its dedicated submission tool, Annotare. For more information about submissions, see our [submission guide](#).



[Browse ArrayExpress](#)



[Submit an Experiment](#)

<https://www.ebi.ac.uk/arrayexpress/>

Expression Atlas



Expression Atlas

Gene expression across species and biological conditions

Query single cell expression

To Single Cell Expression Atlas ➔

[Home](#) [Browse experiments](#) [Download](#) [Release notes](#) [FAQ](#) [Help](#) [Licence](#) [About](#) [Support](#)

Search across **66 species**, 4,424 studies, 158,401 assays

Ensembl 104, Ensembl Genomes 51, WormBase ParaSite 15, EFO 3.10.0

Search Gene set enrichment

Gene / Gene properties

Enter gene query...

Examples: REG1B, zinc finger, O14777 (UniProt), GO:0010468 (regulation of gene expression)

Search Clear

Animals Plants Fungi

Species

Any

Biological conditions

Enter condition query...

Examples: lung, leaf, valproic acid, cancer

What is Expression Atlas

Expression Atlas is a resource to query gene and protein expression data across species and biological conditions and to visualise downstream analysis results to explore co-expression. Queries can be either in a **baseline context**, e.g. find genes expressed in the macaque brain, or in a **differential context**, e.g. find genes that are up or downregulated in response to auxin in *Arabidopsis*. Expression Atlas contains thousands of selected microarray and RNA-sequencing data that are manually curated and annotated with ontology terms, checked for high quality and re-analysed using standardised methods.



Homo sapiens

1502 experiments

Mus musculus

1265 experiments

Rattus norvegicus

185 experiments

Drosophila melanogaster

149 experiments

Gallus gallus

39 experiments

Caenorhabditis elegans

33 experiments

<https://www.ebi.ac.uk/gxa/home>

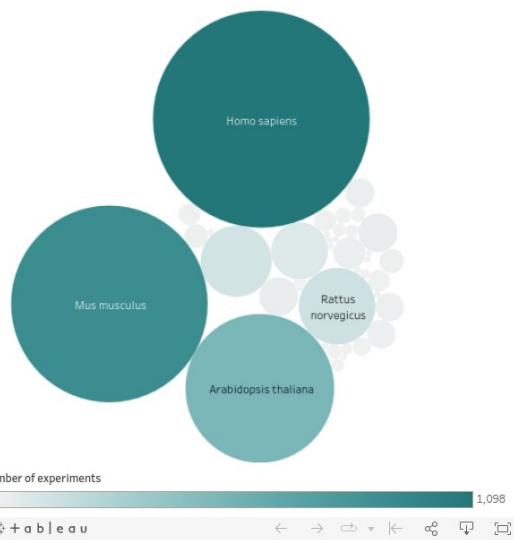
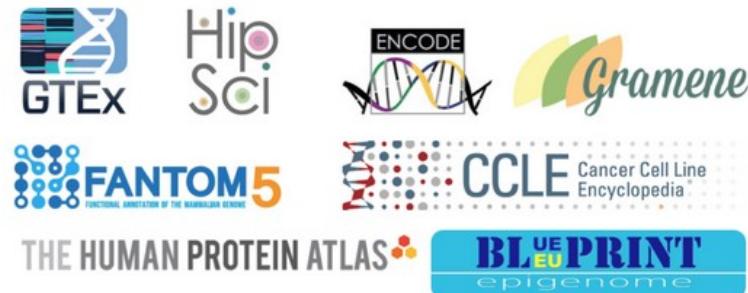
Expression Atlas (<https://www.ebi.ac.uk/gxa/about.html>)



All data are re-analysed using standardised methods

Expression Atlas has re-analysed more than 3,000 experiments, from which more than 500 are RNA-seq experiments. Microarray raw data are analysed using different packages from Bioconductor depending on the array platform used to perform the experiment.

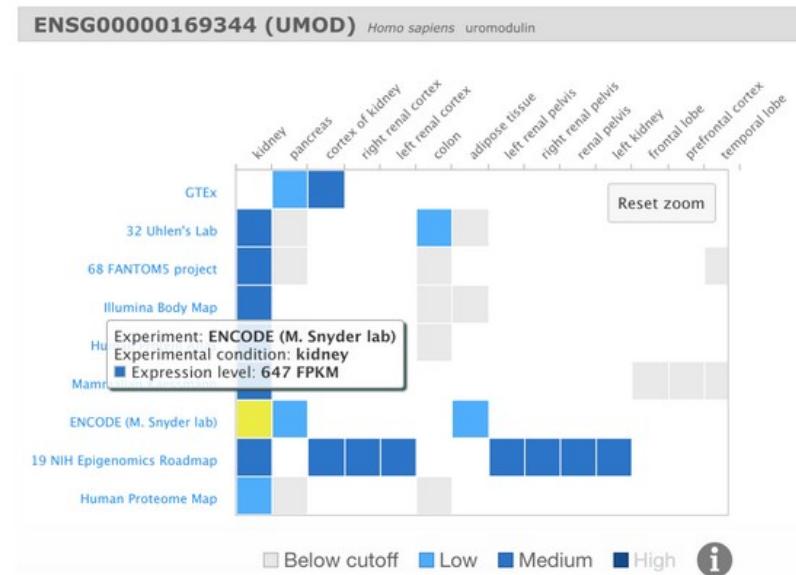
RNA-seq data are analysed using the iRAP pipeline. RNA-seq experiments in Expression Atlas include large landmark studies such as GTEx, CCLE, ENCODE or HipSci.



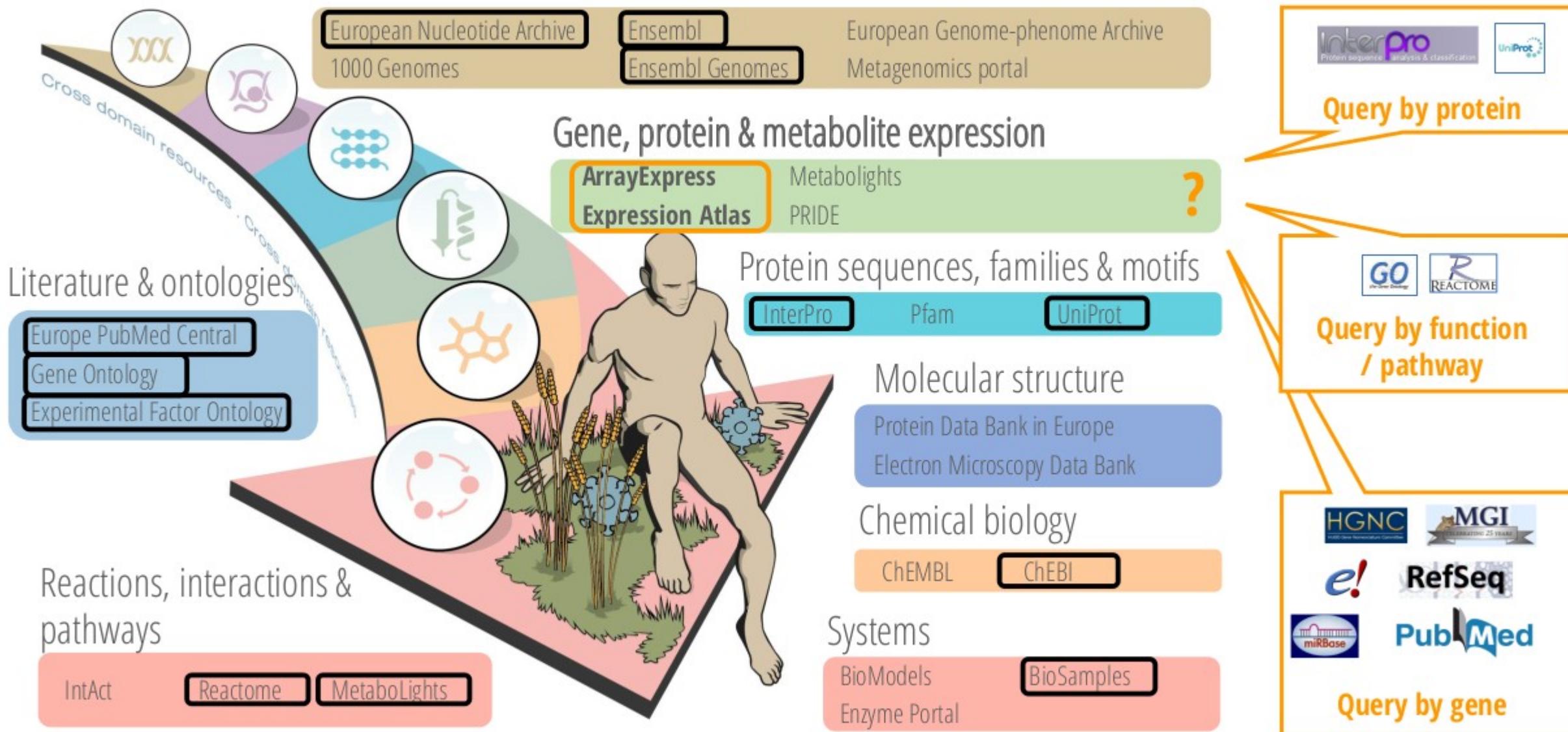
Easy interpretation of gene expression results through heatmap visualisation

Expression Atlas represents gene expression results using heatmaps. In a heatmap, gene expression values are converted into a colour-scale image, providing a visual representation of gene expression levels across different biological conditions.

In Expression Atlas we also generate heatmaps showing gene expression levels of a particular gene across different experiments and different biological conditions.



Expression Atlas



Expression Atlas

What questions does the **Expression Atlas** address?

Where is the **MAPT** gene expressed in the human body?



What genes are expressed in normal **brain**?

Under what conditions does human **MAPT** change expression levels?

Expression Atlas

Baseline	Differential
<p>Gene expression in healthy/untreated conditions</p> <p><i>E.g. human liver, ENCODE cell lines, rice root</i></p>	<p>Changes in gene expression in “comparisons” of experiment conditions</p> <p><i>E.g. mutant vs <u>wild type</u>, drought vs <u>normal watering</u>, drug treated vs <u>untreated</u></i></p>



Expression Atlas release 39 – July 2023

Data statistics

- Ensembl **104** / Ensembl Genomes **51** / WormBase ParaSite **15** gene annotations.
array design probe set mappings.
- This release contains **4424** datasets (**158401** assays), in particular:
 - 1406** RNA-seq and **93** proteomics datasets, of which **340** report baseline gene expression, across a total of **47** different organisms;
 - 1007** datasets reporting expression in plants;
 - 4084** datasets studying samples in **14283** differential comparisons across **51** different organisms.
 - 66** overall species count

Expression Atlas: MAPT gene

ensg00000186868 (MAPT) *Homo sapiens* microtubule associated protein tau

Baseline expression

Differential expression

MAPT information

Show anatomograms

Filter your results

Homo sapiens

Organism part

Age

Cell line

Cell type

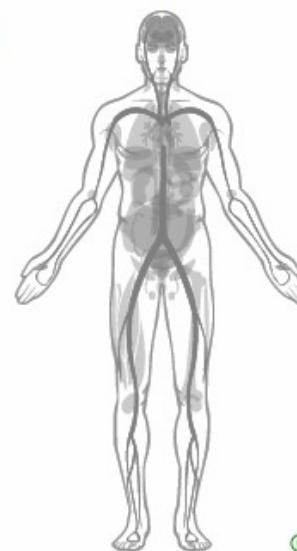
Developmental stage

Disease

Individual

Sex

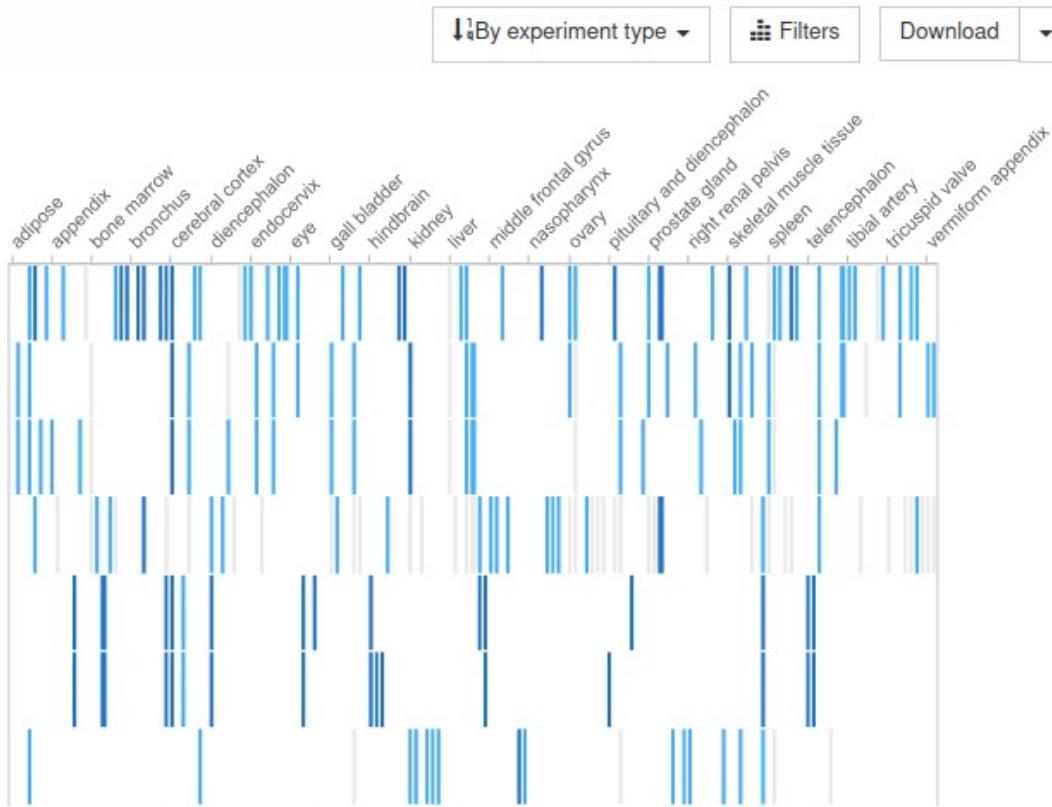
Time



Organism part

Showing 36 experiments:

- T GTEX
- T 32 Uhlen's Lab
- T Hallstrom et al., 2014 - Organism part
- T 68 FANTOM5 project - adult
- T HDBR developing brain - 9 post ...
- T HDBR developing brain - 12 post ...
- T 19 NIH Epigenomics Roadmap



[https://www.ebi.ac.uk/gxa/genes/ensg00000186868?bs=%7B%22homo%20sapiens%22%3A%5B%22ORGANISM_PART%22%5D%7D&ds=%7B%22kingdom%22%3A%5B%22%5D%7D&experiments=36&filters=&highlighted_experiments=19&highlighted_genes=ensg00000186868&highlighted_probes=&highlighted_tissues=&organism=homo_sapiens&organism_part=whole_body&organism_part_label=Whole%20body&probe_type=none&sort_by=experiment&sort_order=asc&tissue_label=Whole%20body](https://www.ebi.ac.uk/gxa/genes/ensg00000186868?bs=%7B%22homo%20sapiens%22%3A%5B%22ORGANISM_PART%22%5D%7D&ds=%7B%22kingdom%22%3A%5B%22%5D%7D&experiments=36&filters=&highlighted_experiments=19&highlighted_genes=ensg00000186868&highlighted_probes=&highlighted_tissues=&organism=homo_sapiens&organism_part=whole_body&organism_part_label=Whole%20body&organism_label=Homo%20sapiens&probe_type=none&sort_by=experiment&sort_order=asc&tissue_label=Whole%20body)

Expression Atlas: MAPT gene

ensg00000186868 (MAPT) *Homo sapiens* microtubule associated protein tau

Baseline expression

Differential expression

MAPT information

Filter your results



Display log₂-fold change

Download results

Kingdom

Animals

Species

Homo sapiens

Experiment type

Microarray 1-colour mRNA differential

RNA-seq mRNA differential

Experimental variables

Disease

Compound

Time

Phenotype

Log ₂ -fold change	Species	Gene name	Comparison	Experimental variables	Experiment name
	MAPT	'doxycycline; 0.5 microgram per milliliter; Day 4 post Neurogenin induction' vs 'control'	compound, treatment	Rapid neurogenesis through transcriptional activation in human stem cell (RNA-Seq)	
	MAPT	'11 day' vs '0 day'	time	RNA-Seq characterization of human H1-derived NPC differentiation timecourse	
	MAPT	'trichostatin A; 500 nanomolar' vs 'control'	compound	Transcription profiling by high throughput sequencing of human aortic endothelial cell treated with TSA	
		'suberoylanilide hydroxamic acid'		Vascular histone deacetylation by	

Expression Atlas: genes differentially expressed in an experiment

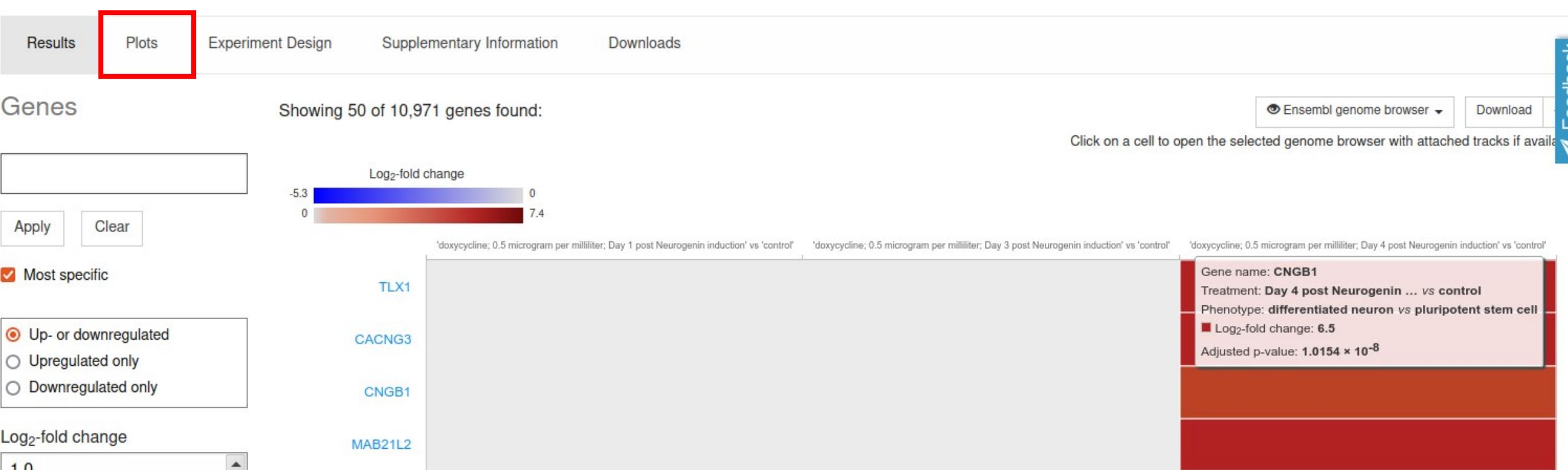
Rapid neurogenesis through transcriptional activation in human stem cell (RNA-Seq)

RNA-Seq mRNA differential

Organism: *Homo sapiens*

Publication:

- Busskamp V, Lewis NE, Guye P, Ng AH, Shipman SL et al. (2014) *Rapid neurogenesis through transcriptional activation in human stem cells.*



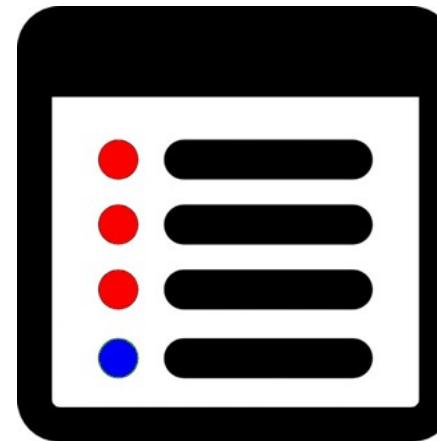
<https://www.ebi.ac.uk/gxa/experiments/E-GEO-60548/Results?specific=true&geneQuery=%255B%255D&filterFactors=%257B%257D&cutoff=%253D&sortOrder=desc&startIndex=1&size=50>

Expression Atlas: genes differentially expressed in an experiment

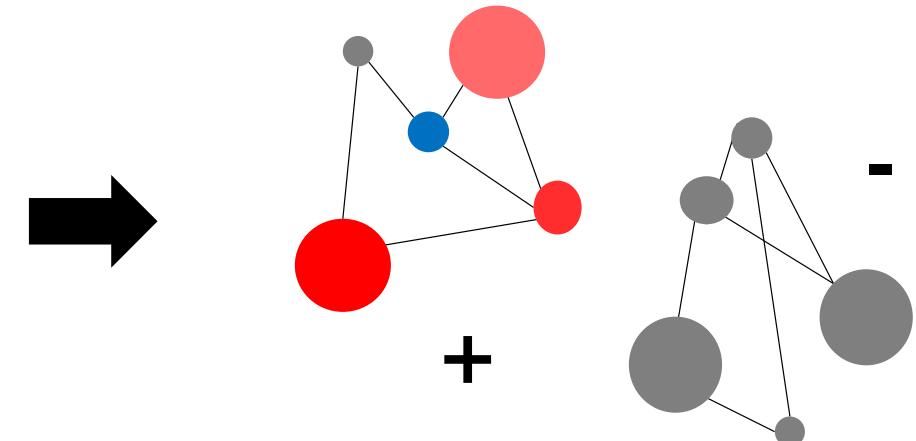
Differential expression



Experiment /
Condition



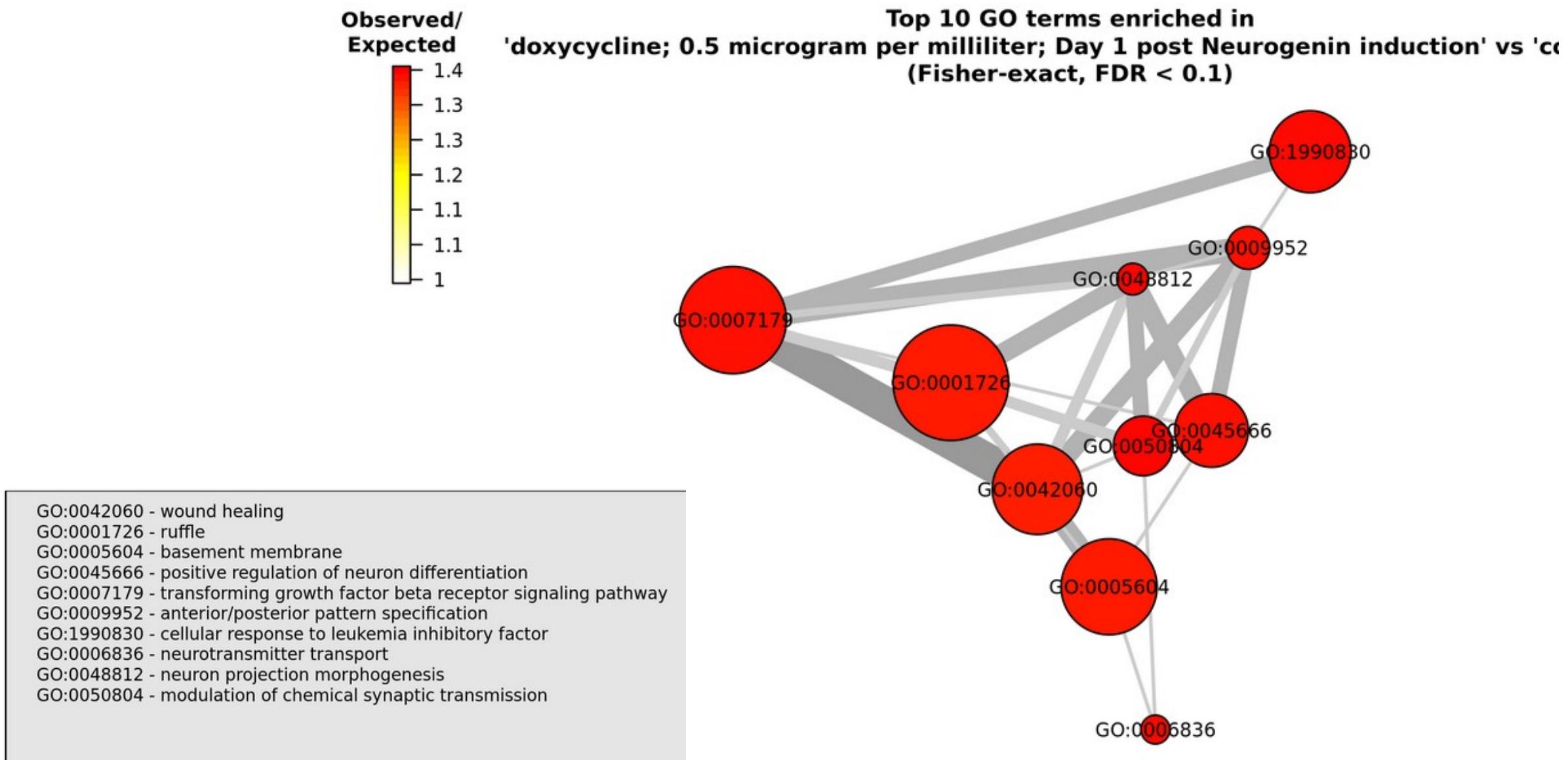
Differentially
expressed
genes



Overrepresented or
Underrepresented Molecular
Function, Biological Process,
or Cellular Component

<http://www.geneontology.org/page/go-enrichment-analysis>

Expression Atlas: genes differentially expressed in an experiment



Expression Atlas: Challenges

1. Incomplete and inconsistent meta-data
2. Studies carried out in the same “type” of samples but in different research teams: comparable?
3. How to quantify expression level within one data set? How to “normalize” expression levels across data sets?
4. Microarray and RNA-seq data on the same set of samples: comparable?
5. How to keep up with gene annotation and genome assembly updates?