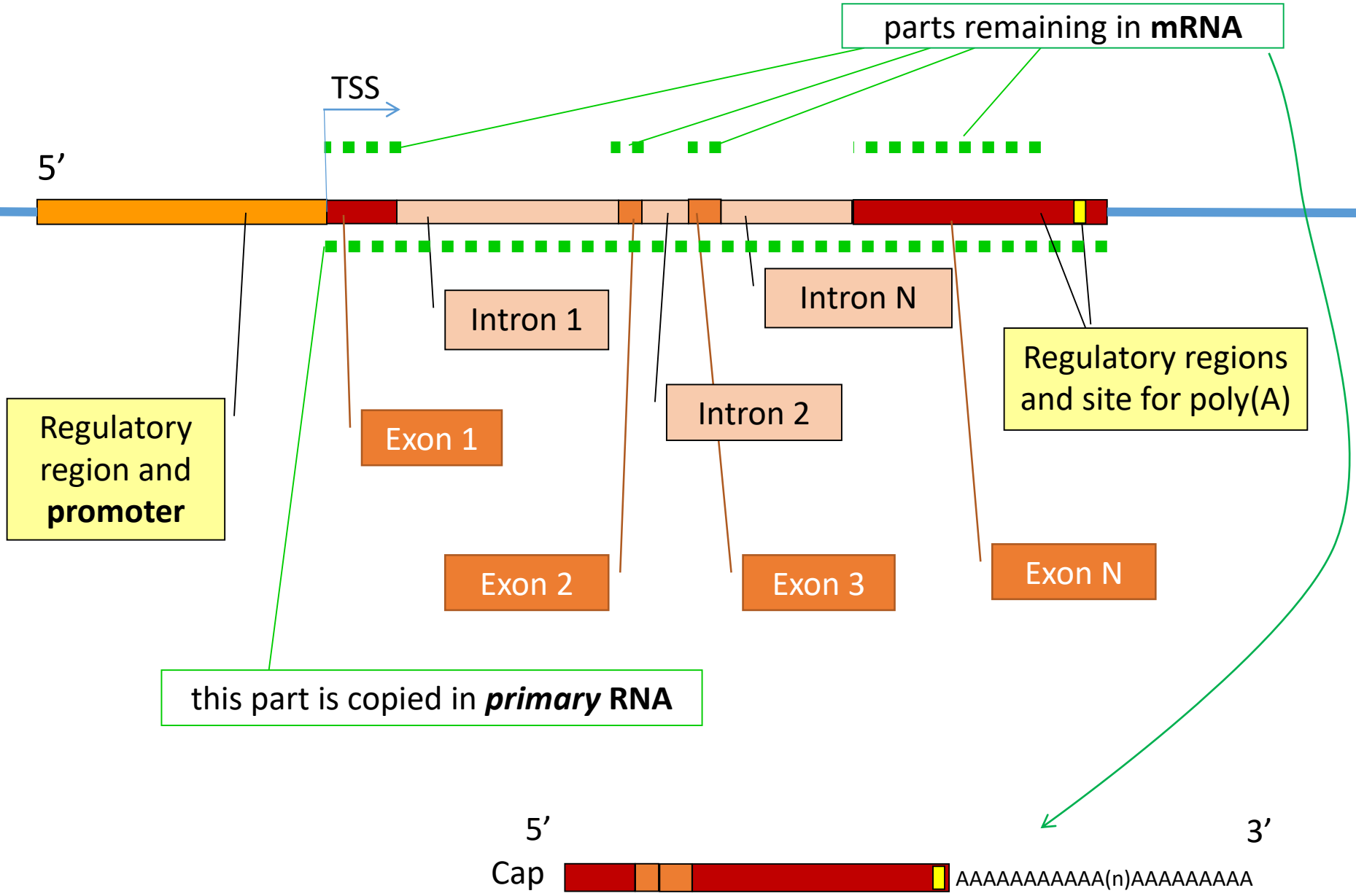
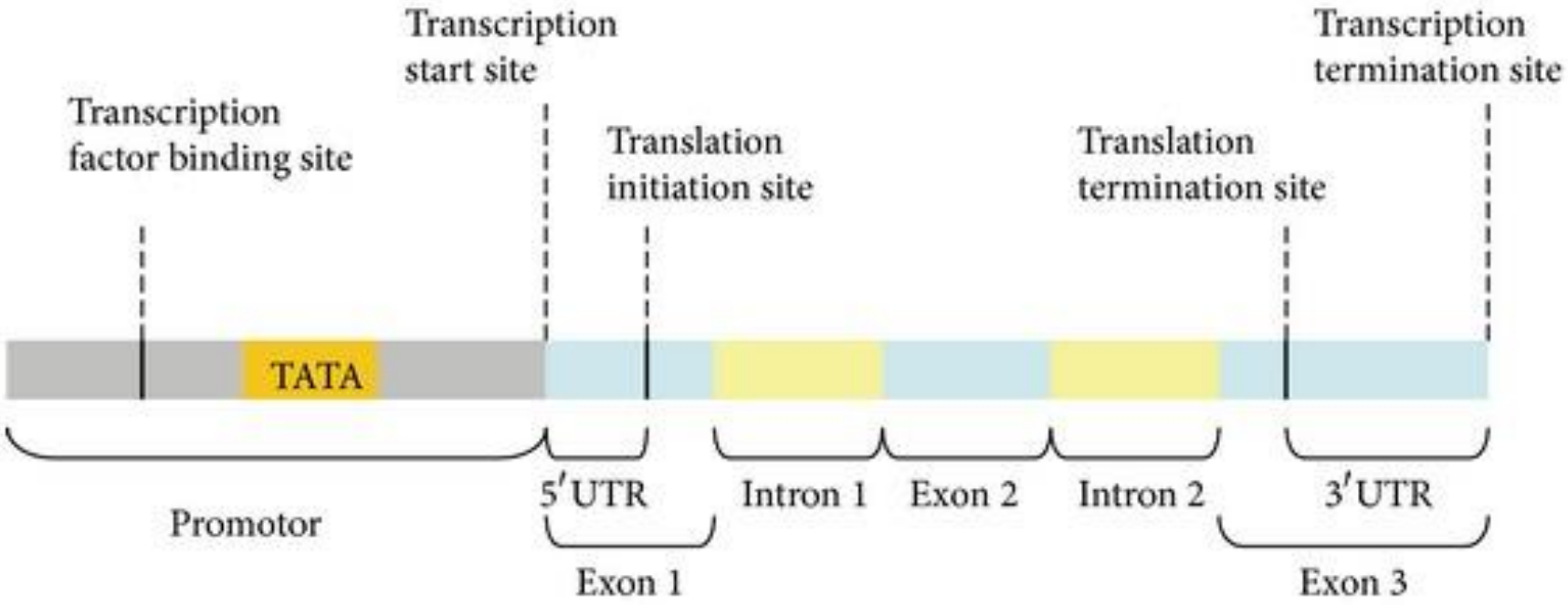
Eukaryotic protein coding genes



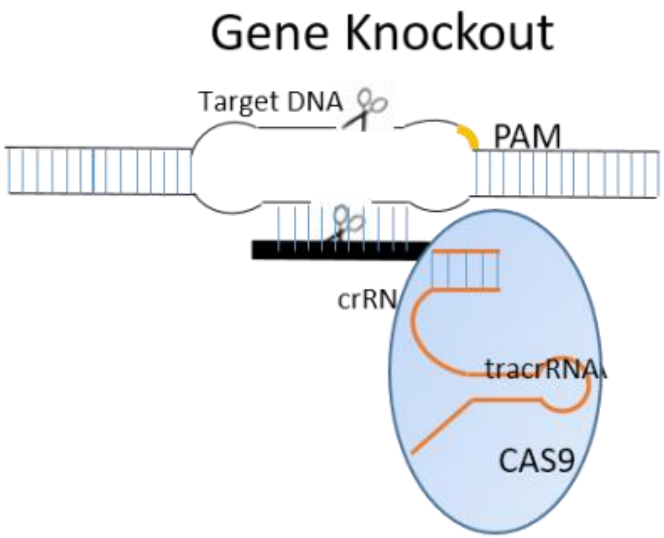
Identification of gene elements

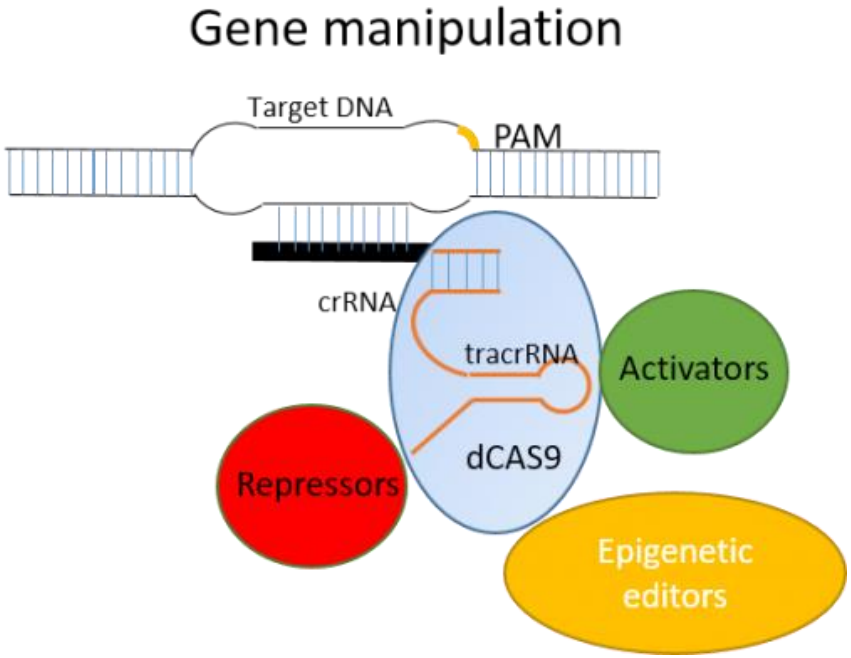


Why is important to identify TSS, ATG,

exons and introns for one specific gene?

Manipulation of genomic regions





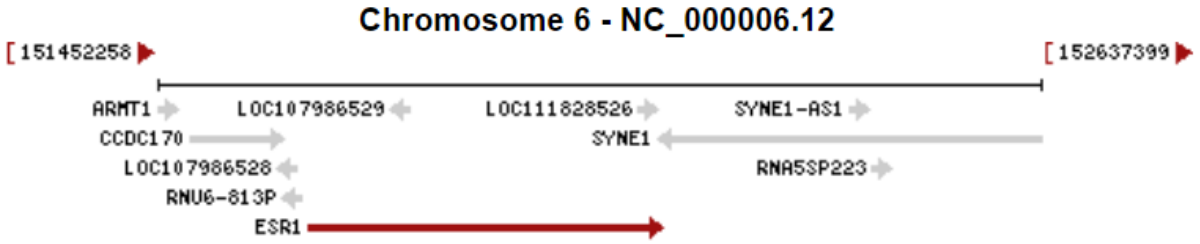
Where is my gene of interest?
Gene context could have a role in the gene regulation

See ESR1 in Genome Data Viewer

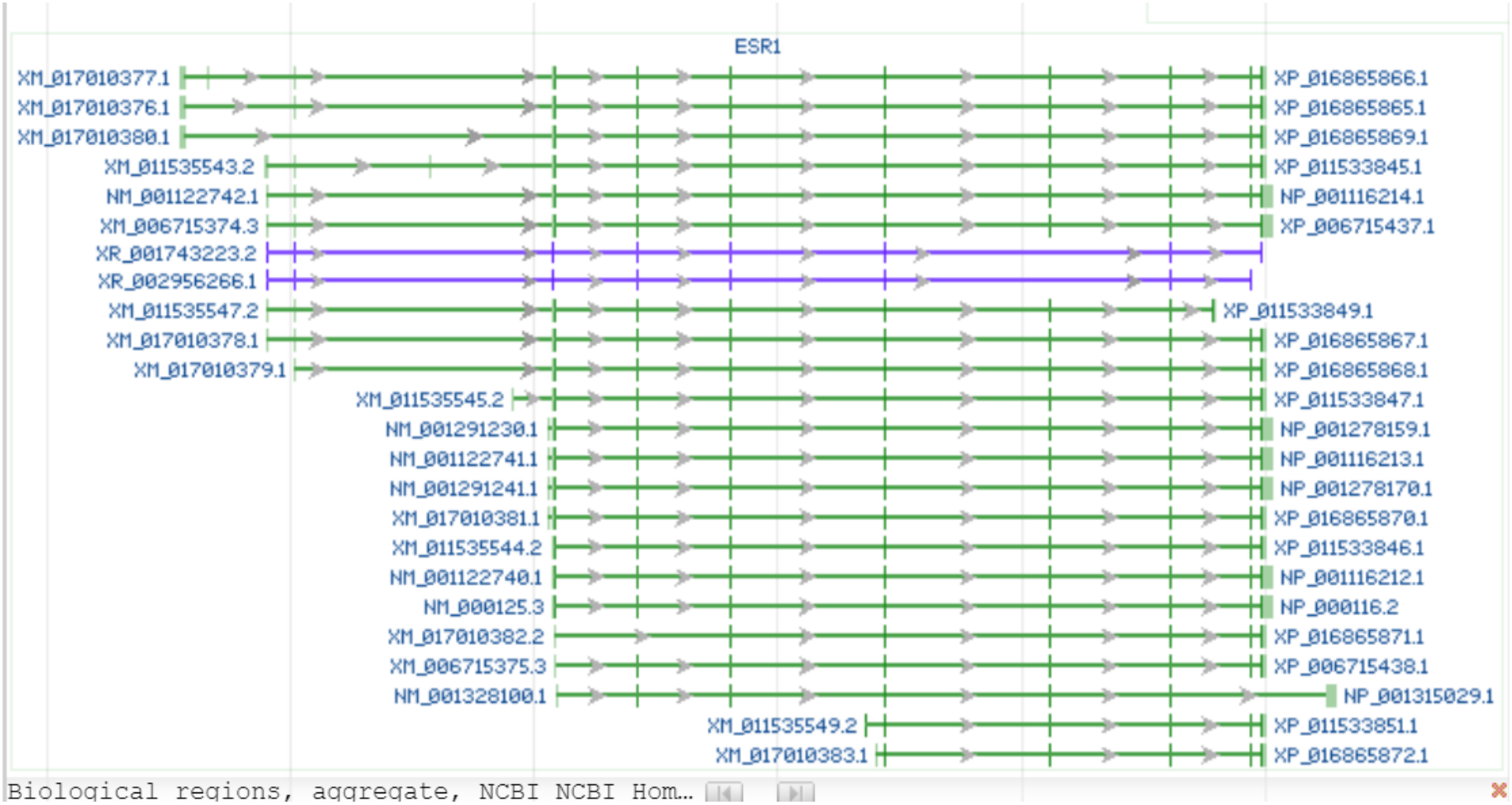
Location: 6q25.1-q25.2

Exon count: 23

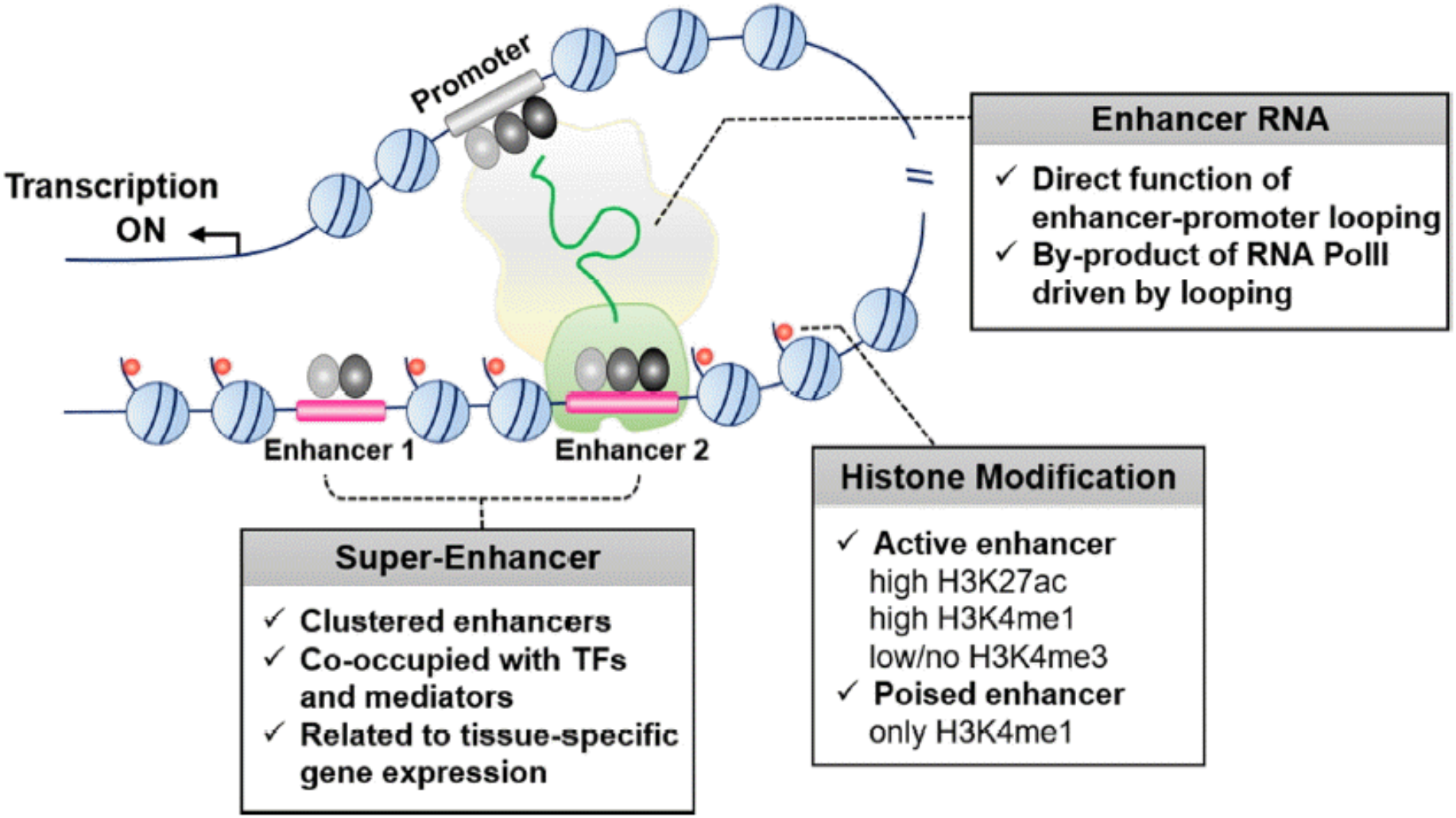
Annotation release	Status	Assembly	Chr	Location
109	current	GRCh38.p12 (GCF_000001405.38)	6	NC_000006.12 (151654148..152129604)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	6	NC_000006.11 (152011631..152424409)

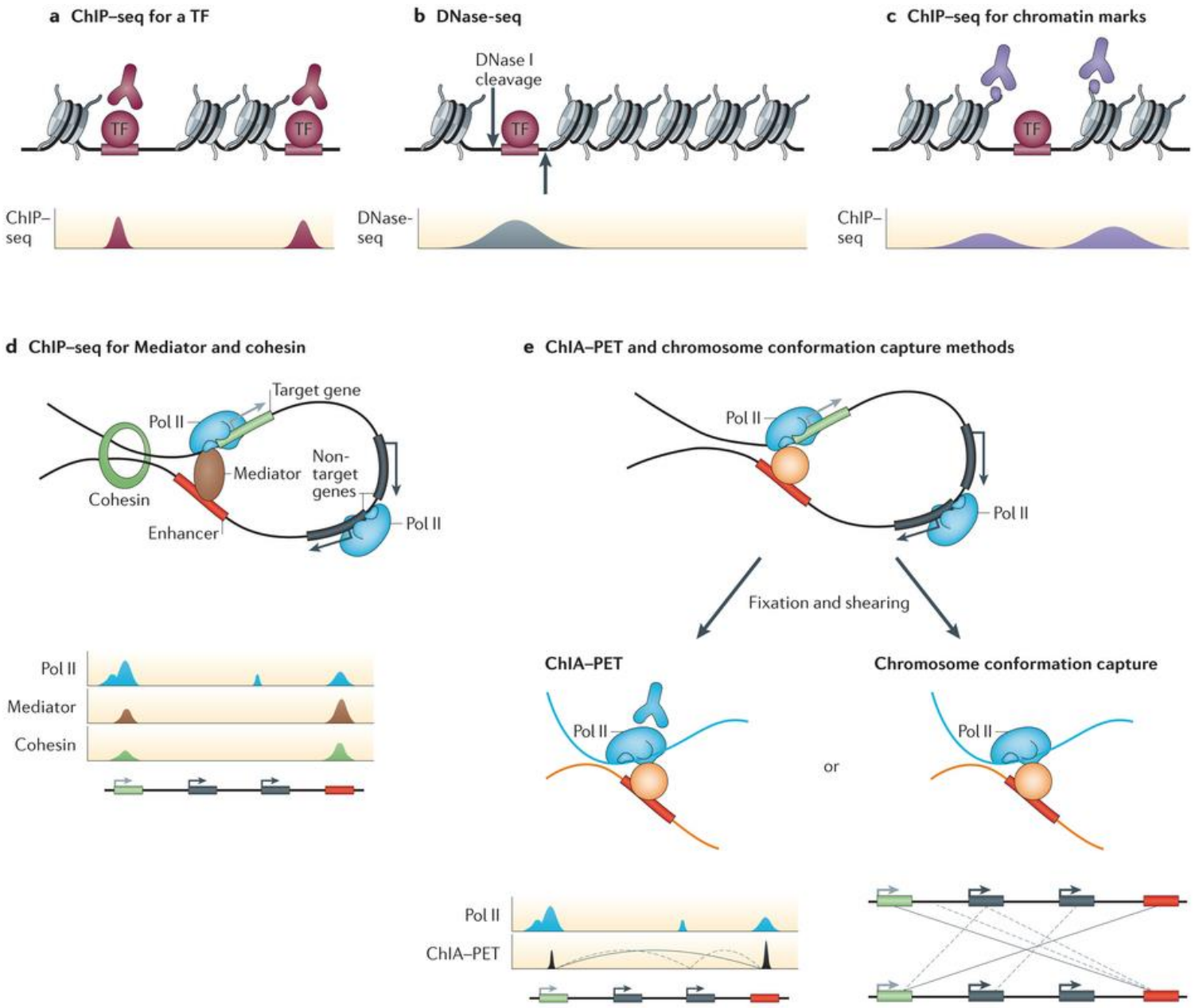


All trascript annotated to this locus
What is the transcript that is target of my research?

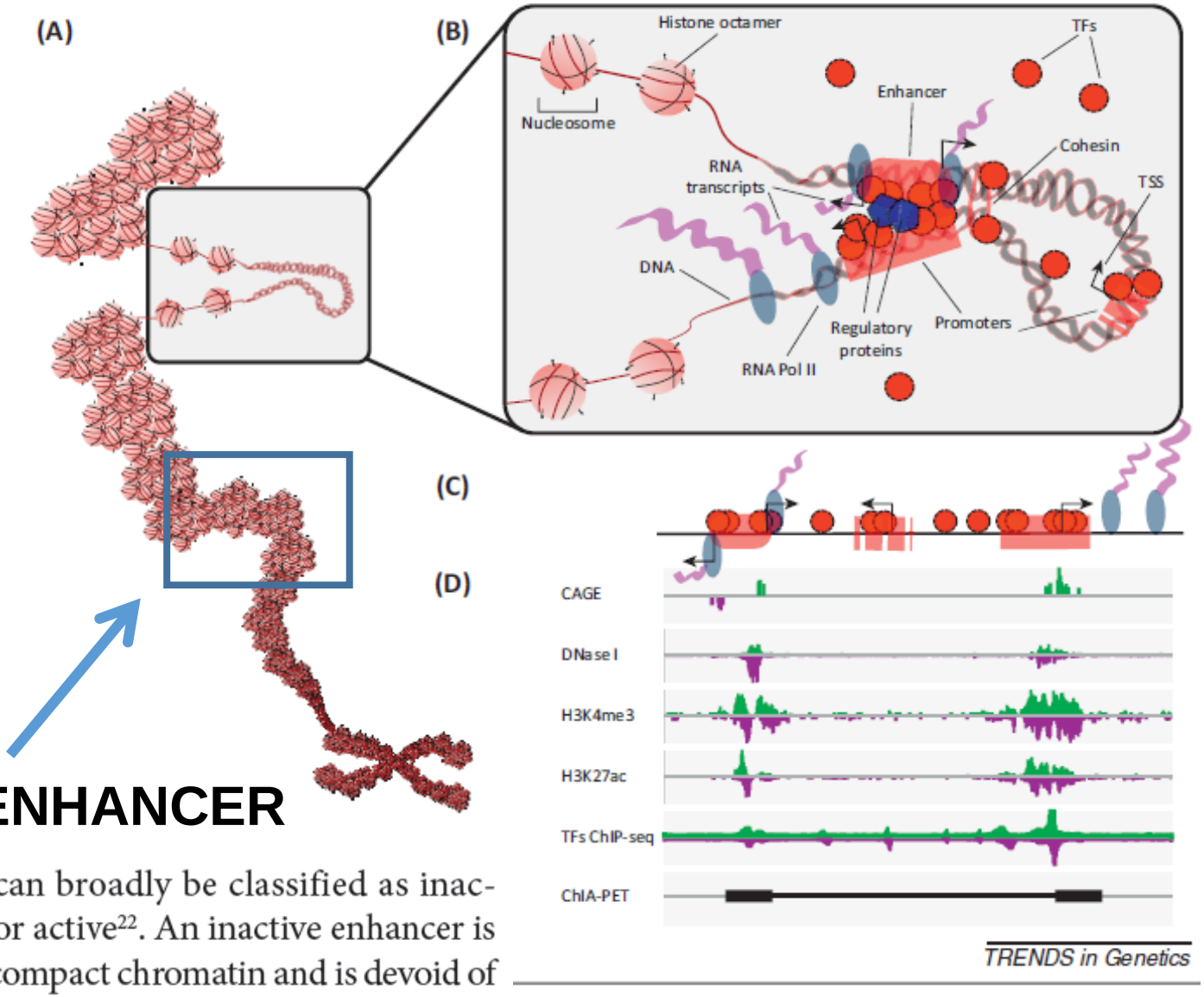


GENOMIC REGULATORY REGION ELEMENTS




a ChIP–seq for a TF
TF
TF
ChIP–seq
b DNase-seq
DNase I cleavage
TF
DNase-seq
c ChIP–seq for chromatin marks
TF
ChIP–seq
d ChIP–seq for Mediator and cohesin
Target gene
Pol II
Mediator
Non-target genes
Cohesin
Enhancer
Pol II
Pol II
Mediator
Cohesin
e ChIA–PET and chromosome conformation capture methods
Pol II
Pol II
Fixation and shearing
ChIA–PET
Chromosome conformation capture
Pol II
or
Pol II
Pol II
ChIA–PET

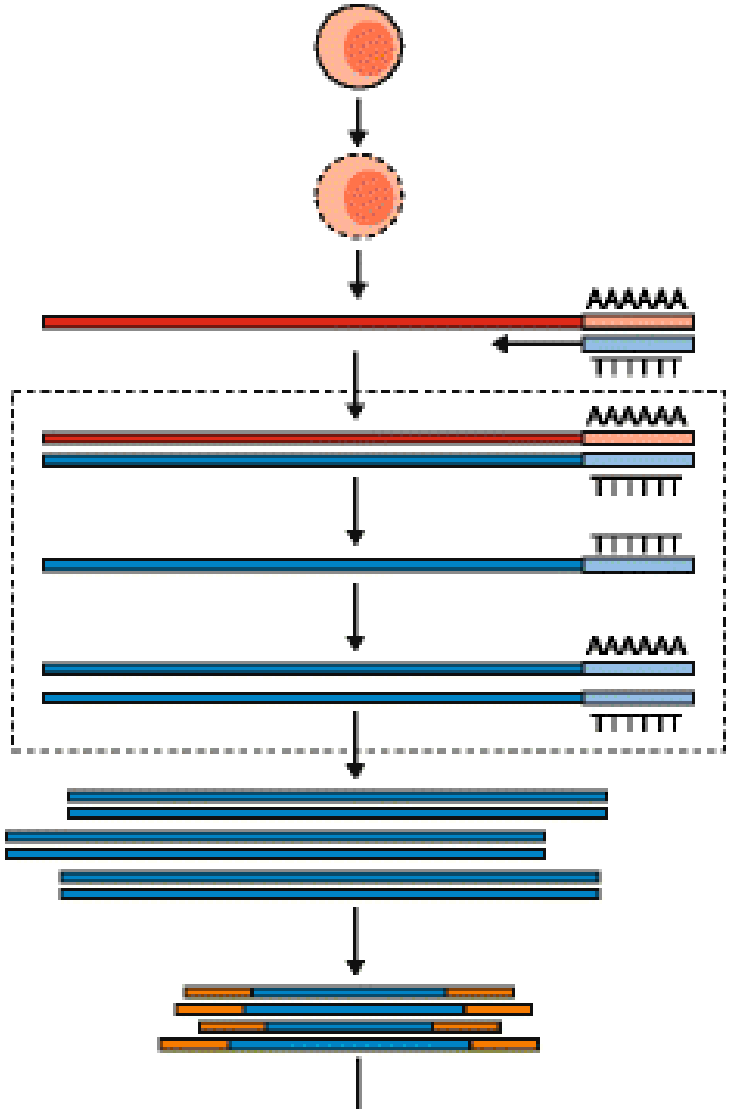
ACTIVE ENHANCER



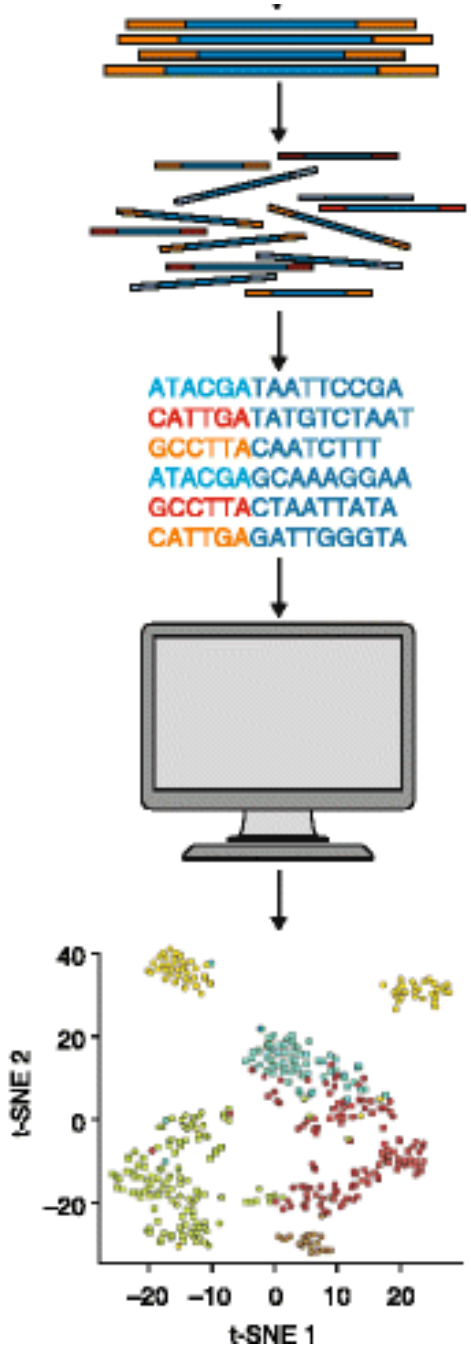
INACTIVE ENHANCER

Enhancer states can broadly be classified as inactive, primed, poised or active[22]. An inactive enhancer is essentially buried in compact chromatin and is devoid of transcription factor binding and histone modifications.

Single Cell Analysis description



1. Isolate single cells from a tissue sample (including micro-dissection and manipulation, flow cytometric cell-sorting, microfluidic platforms, and droplet-based methods)
2. Single cell lysis in a way that preserves cellular mRNA
3. mRNA molecule capture using poly[T] sequence primers that bind to mRNA poly[A] tails
4. Convert poly[T]-primed mRNA into cDNA using reverse transcription
5. cDNA amplification (usually by PCR or by *in vitro* transcription)
6. cDNA sequencing library preparation (insert 'index' nucleotide barcodes to identify each library)



6. cDNA sequencing library preparation (insert 'index' nucleotide barcodes to identify each library)

7. Pool cDNA sequencing libraries

Sequence libraries (via Next Generation Sequencing)

8. Use bioinformatic methods to perform quality control and to assess technical variability in the scRNA-seq data

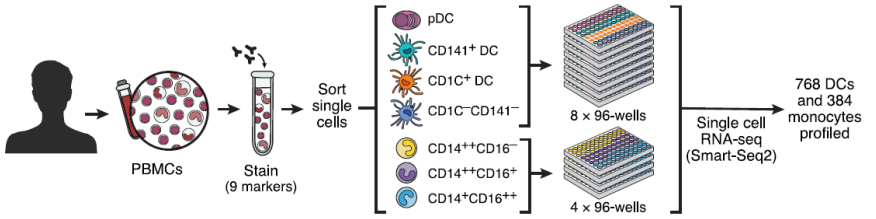
9. Use bioinformatic and/or computational methods to interpret robust data biologically

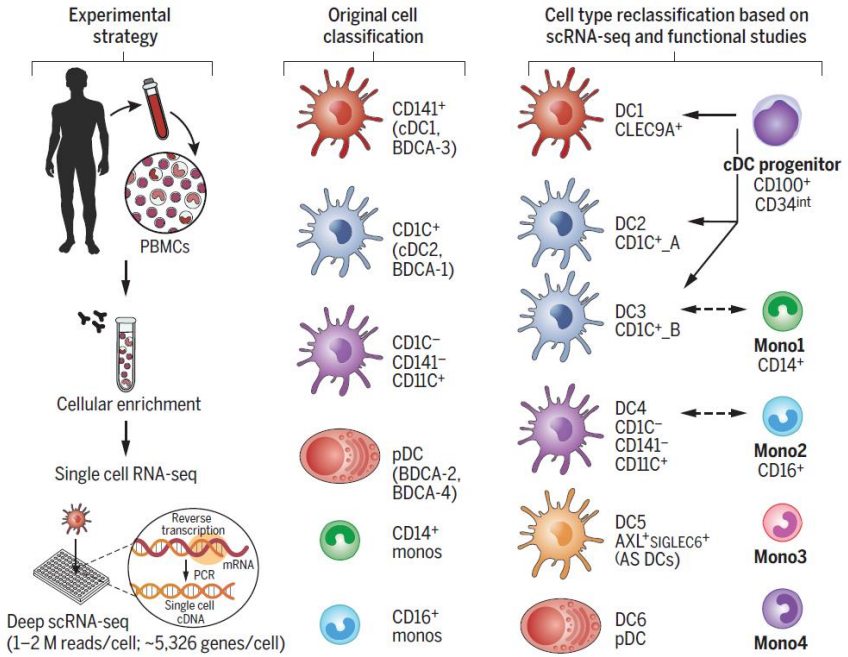
Single cell analysis to identify what genes are expressed in cell types

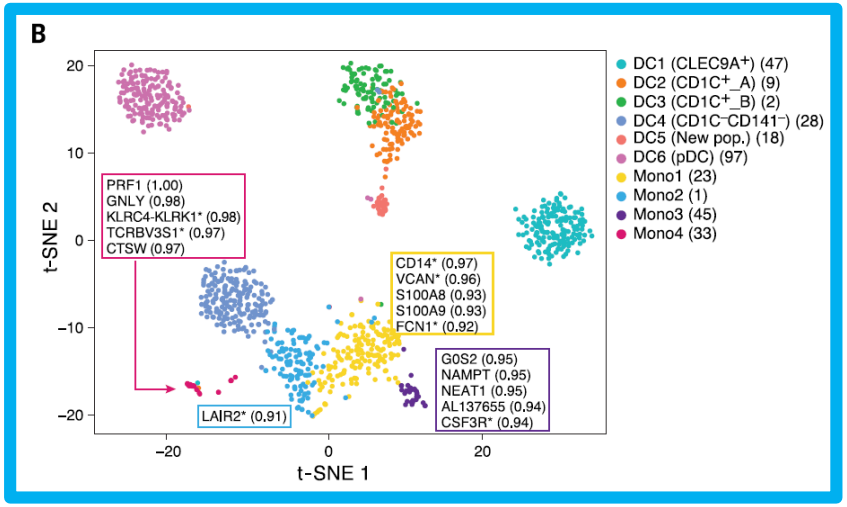
IMMUNOGENOMICS

Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors

Alexandra-Chloé Villani,*† Rahul Satija,* Gary Reynolds, Siranush Sarkizova, Karthik Shekhar, James Fletcher, Morgane Griesbeck, Andrew Butler, Shiwei Zheng, Suzan Lazo, Laura Jardine, David Dixon, Emily Stephenson, Emil Nilsson, Ida Grundberg, David McDonald, Andrew Filby, Weibo Li, Philip L. De Jager, Orit Rozenblatt-Rosen, Andrew A. Lane, Muzlifah Haniffa,† Aviv Regev,† Nir Hacohen†







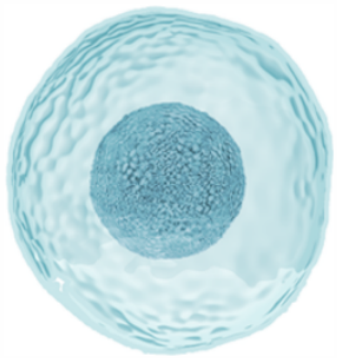
Villani *et al.*, *Science* **356**, eaah4573 (2017)

Single-cell transcriptome analysis of precursors of human CD4+ cytotoxic T lymphocytes

Single-cell RNA-Seq mRNA baseline

Number of cells: 1.411
Organism: *Homo sapiens*
Publication:

- Patil VS, Madrigal A, Schmiedel BJ, Clarke J, O'Rourke P et al. (2018) *Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis.*



Gene expression analysis in single cells across species and biological conditions

Single Cell Expression Atlas supports research in single cell transcriptomics. The Atlas annotates publicly available single cell RNA-Seq experiments with ontology identifiers and re-analyses them using standardised pipelines available through iRAP, our RNA-Seq analysis toolkit. The browser enables visualisation of clusters of cells, their annotations and supports searches for gene expression within and across studies.

Search

Gene ID or gene symbol

Species

Any

Examples: CFTR (gene symbol), ENSG00000115904 (Ensembl ID), 657 (Entrez ID), MGI:98354 (MGI ID), FBgn0004647 (FlyBase ID)

Search

Your favourite gene expression in Single Cell-Seq datasets

Gene ID or gene symbol

ESR1

Species

Homo sapiens

Search

Marker genes

Experiments with marker genes

Species

Homo sapiens

Inferred cell type

Select...

Organism part

ESR1 (ENSG00000091831) is expressed in:

Species	Marker genes	Title	Experimental variables	Number of assays
Homo sapiens	✖	Single cell RNA-seq of three human melanoma cell lines: Ma-Mel-123, Ma-Mel-108 and Ma-Mel-93	• single cell identifier	226
Homo sapiens	✖	Single-Cell RNAseq analysis of diffuse neoplastic infiltrating cells at the migrating front of human glioblastoma	• single cell identifier • biopsy site	3,588

Single Cell Expression Atlas

Single cell gene expression across species

Query bulk expression

Back to Expression Atlas

Home | Gene search | Browse experiments | Release notes | Help | Feedback

Single-cell transcriptome analysis of precursors of human CD4+ cytotoxic T lymphocytes

Single-cell RNA-Seq mRNA baseline

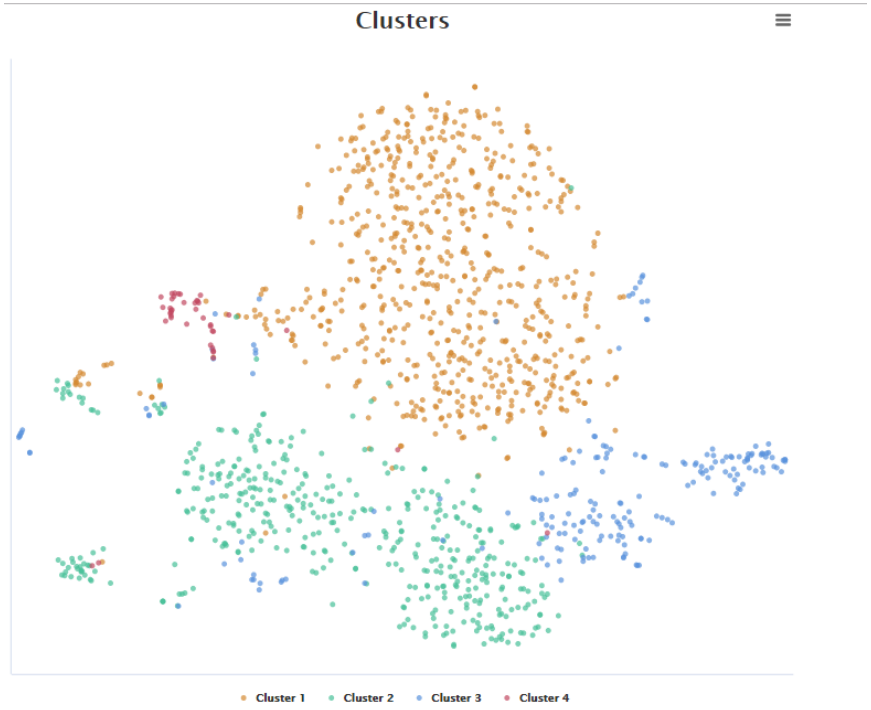
Number of cells: 1.411

Organism: *Homo sapiens*

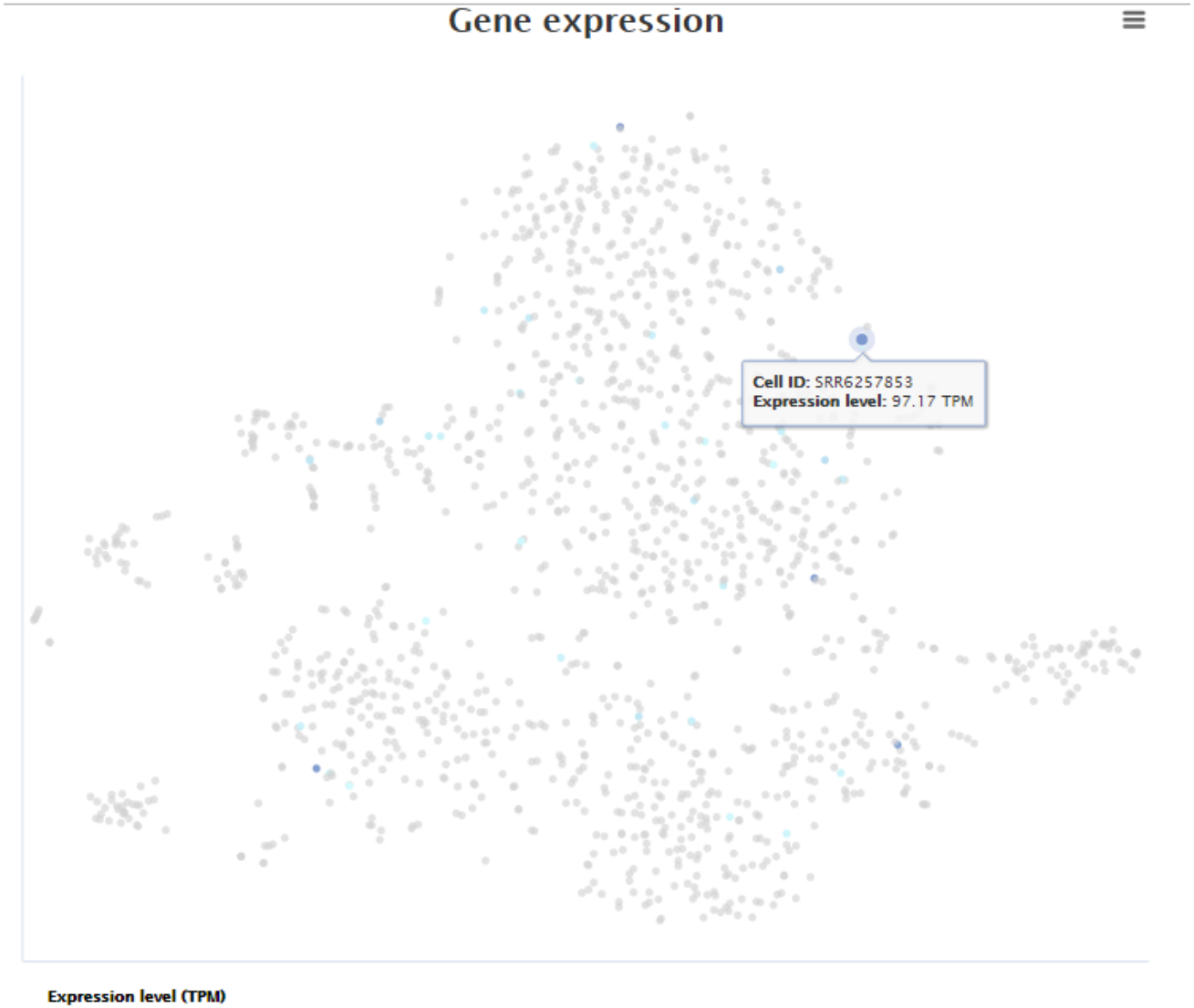
Publication:

- Patil VS, Madrigal A, Schmiedel BJ, Clarke J, O'Rourke P et al. (2018) *Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis.*

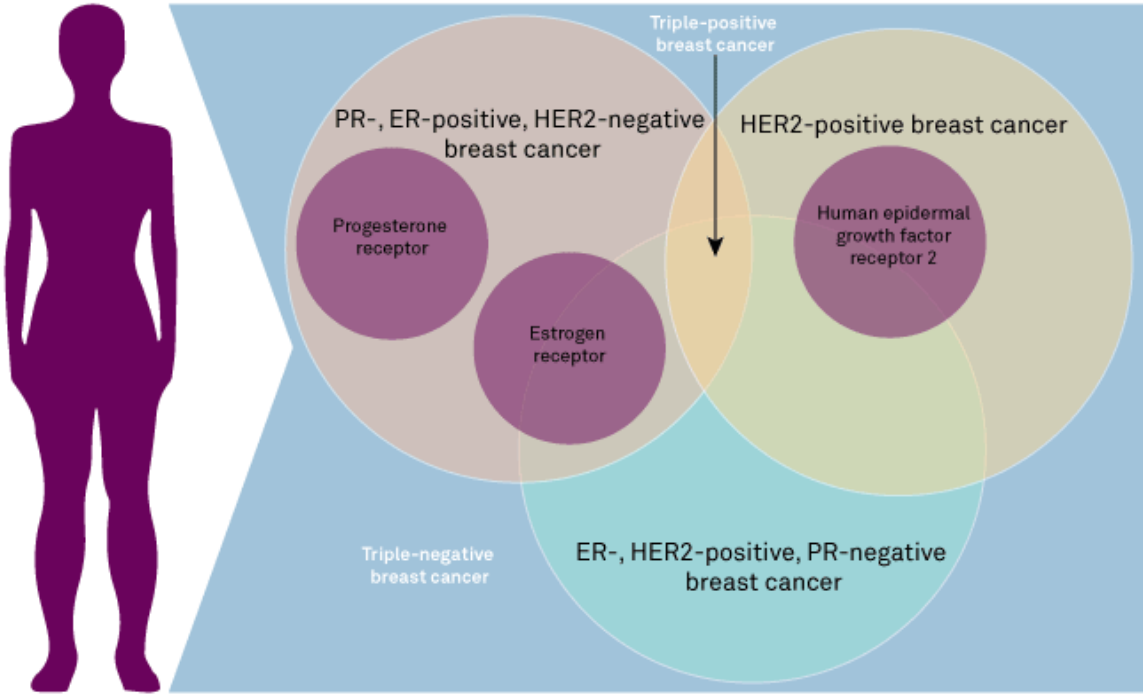
Cluster of cell types in Single Cell-Seq dataset



ESR1 expression in Single Cell-Seq dataset



FIVE MOST-COMMON
BREAST CANCER SUBTYPES



HER2-positive HER2-positive cancer that is ER- and PR-negative is one of the most common subtypes. HER2-positive cancer grows rapidly and is typically treated with targeted therapies such as Roche's Herceptin.

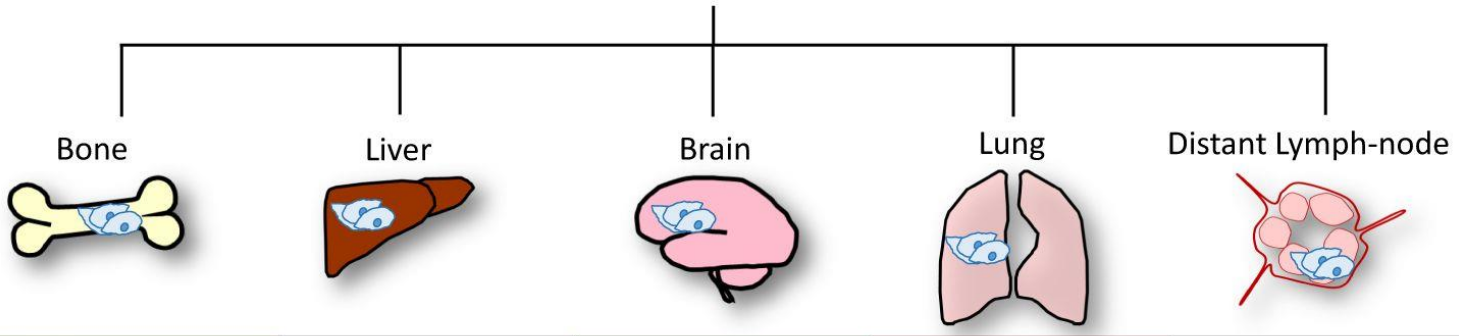
Luminal ER/HER2-positive, PR-negative, or Luminal A breast cancer; and PR-, ER-positive, HER2-negative, or Luminal B breast cancer, generally respond well to chemotherapy and certain hormone therapies.
In general, hormone receptor-positive breast cancers are slower-growing and have a better short-term outlook than other subtypes. Luminal A is the most common subtype in every race and age.

Triple-negative Between 10% and 20% of all breast cancer is triple-negative, showing no receptors for estrogen, progesterone or HER2. Triple-negative breast cancer is more common in younger women and those of African-American or Latin descent. It is generally treated with chemotherapy or radiation.

Triple positive Triple-positive breast cancer shows signs of all three receptors and is a relatively new distinct subtype, with early research suggesting that it be treated with anti-HER2 agents and chemotherapy.

Sources: American Cancer Society, Cancer Treatment Centers of America, National Center for Biotechnology Information
Credit : Faizan Hameed

Breast Cancer Distant Metastases



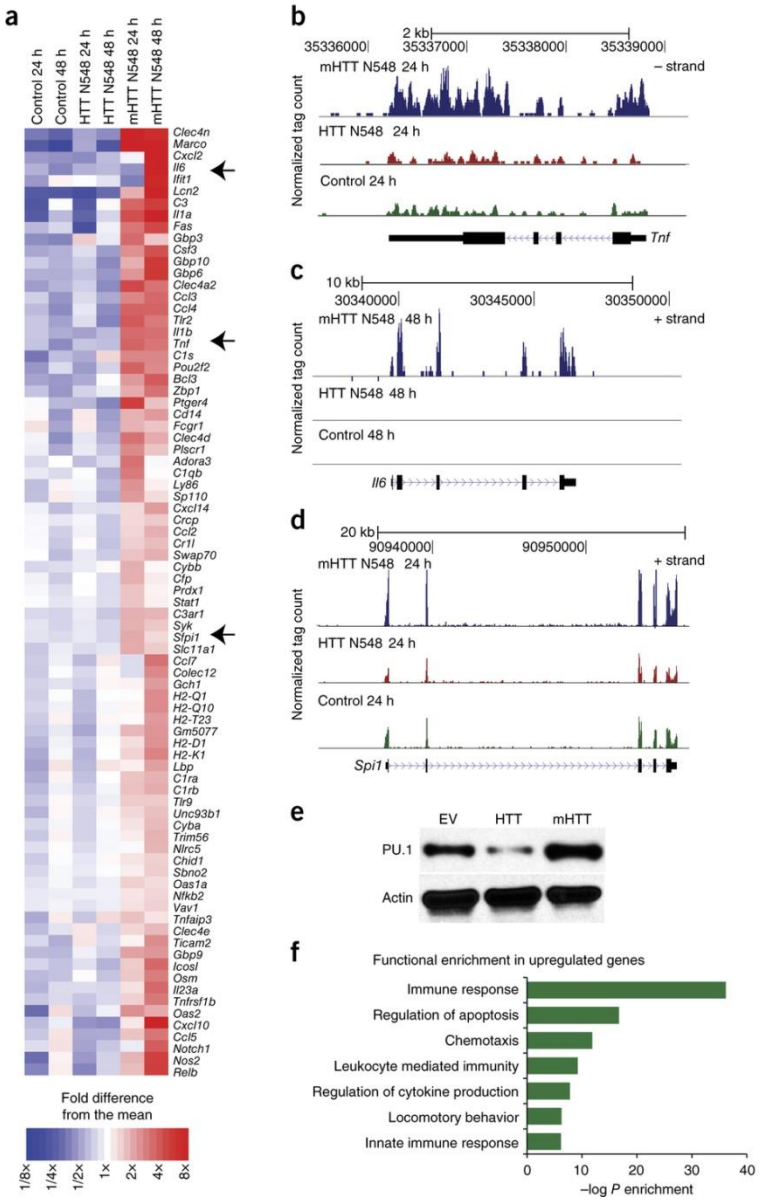
	Bone	Liver	Brain	Lung	Distant Lymph-node
Associated subtypes	Luminal-HER2	HER2-enriched ER-positive Luminal B Luminal-HER2	HER2-enriched Luminal-HER2 TN-nonbasal Basal-like	TN-nonbasal Basal-like Luminal B HER2+, HR-, p53-	Luminal type HER2-enriched
Molecular features	Growth factors: IGF1, PGE2, TGFβ, PDGF and FGF2 Interleukins: IL-11, IL-1, IL-6 PTHrP OPN Heparanase RANKL-RANK pathway Src-dependent pathway	Chemokines and receptors: CXCR4/CXCL12 Interleukins: IL-6 Integrin complexes: α2β1, α5β1 N-cadherin HIF-regulated genes: LOX, OPN, VEGF, TWIST β-catenin-independent WNT signaling Downregulation of ECM (stromal) genes	ST6GALNAC5 CSC markers: Nestin, CD133, and CD44 Growth factors: VEGF and HBEGF Chemokines and receptors: CXCR4 Cytokines: CK5 MMP-1 and MMP-9 IL-8 Ang-2 COX2 L1CAM	Growth factors and their receptors: TGFβ, EGFR, EREG, VEGF Matrix metalloproteinases: MMP-1 and MMP-2 COX2 LOX BMP inhibitors: GALNTs and Coco	Kallikreins: KLK10, KLK11, KLK12, and KLK13 Downregulation of BCR signal pathway

 Homo sapiens	• See cluster 1 for k = 4 • See cluster 2 for k = 7	Single cell RNA-seq of primary breast cancer cells and lymph node metastases from 11 patients representing the four subtypes of breast cancer: luminal A, luminal B, HER2 and triple negative breast cancer	• single cell identifier • histology • sampling site	540

ESR1 expression in specific cell subtypes in Single Cell-Seq dataset

RNA-Seq analysis reveals that mHTT N terminus expression triggers pro-inflammatory gene expression in BV2 microglia

Heat map: Gene modulated in different conditions



UCSC browser for specific gene locus

Gene ontology

The **Gene Ontology project**
provides controlled vocabularies of defined terms
representing **gene product properties**.

These cover three domains:

Cellular Component: the parts of a cell or its extracellular environment;

Molecular Function, the elemental activities of a gene product at the molecular level, such as binding or catalysis

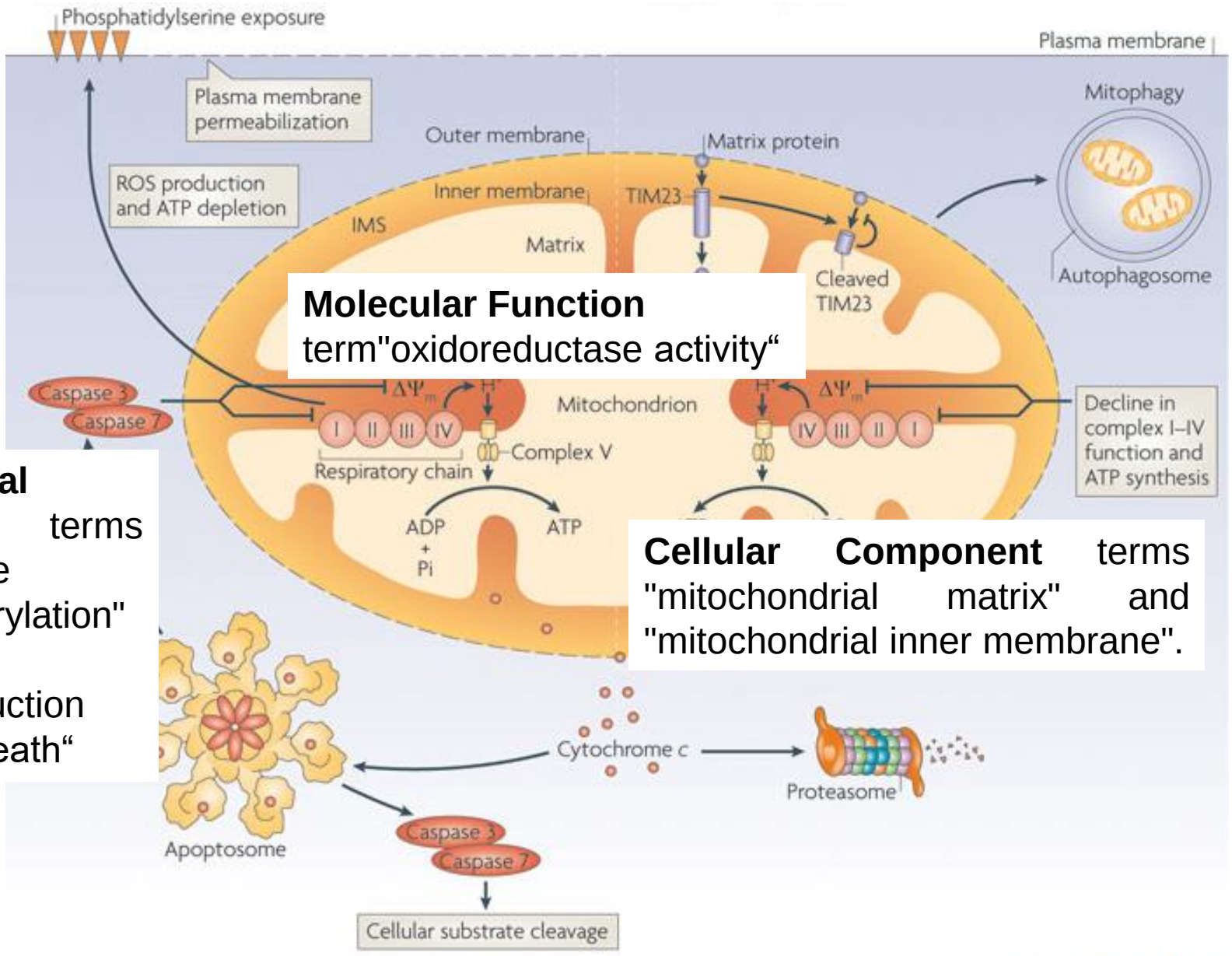
Biological Process, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms.

In an example of GO annotation,

the gene product **"cytochrome c"**
Molecular Function term "oxidoreductase activity“

Biological Process terms "oxidative phosphorylation" and "induction of cell death“

Cellular Component terms "mitochondrial matrix" and "mitochondrial inner membrane".



Molecular Function
term"oxidoreductase activity“

Biological Process terms "oxidative phosphorylation"

and "induction of cell death“

Cellular Component terms "mitochondrial matrix" and "mitochondrial inner membrane".