

UNIVERSITÀ DEGLI STUDI
DEL SANNIO Benevento

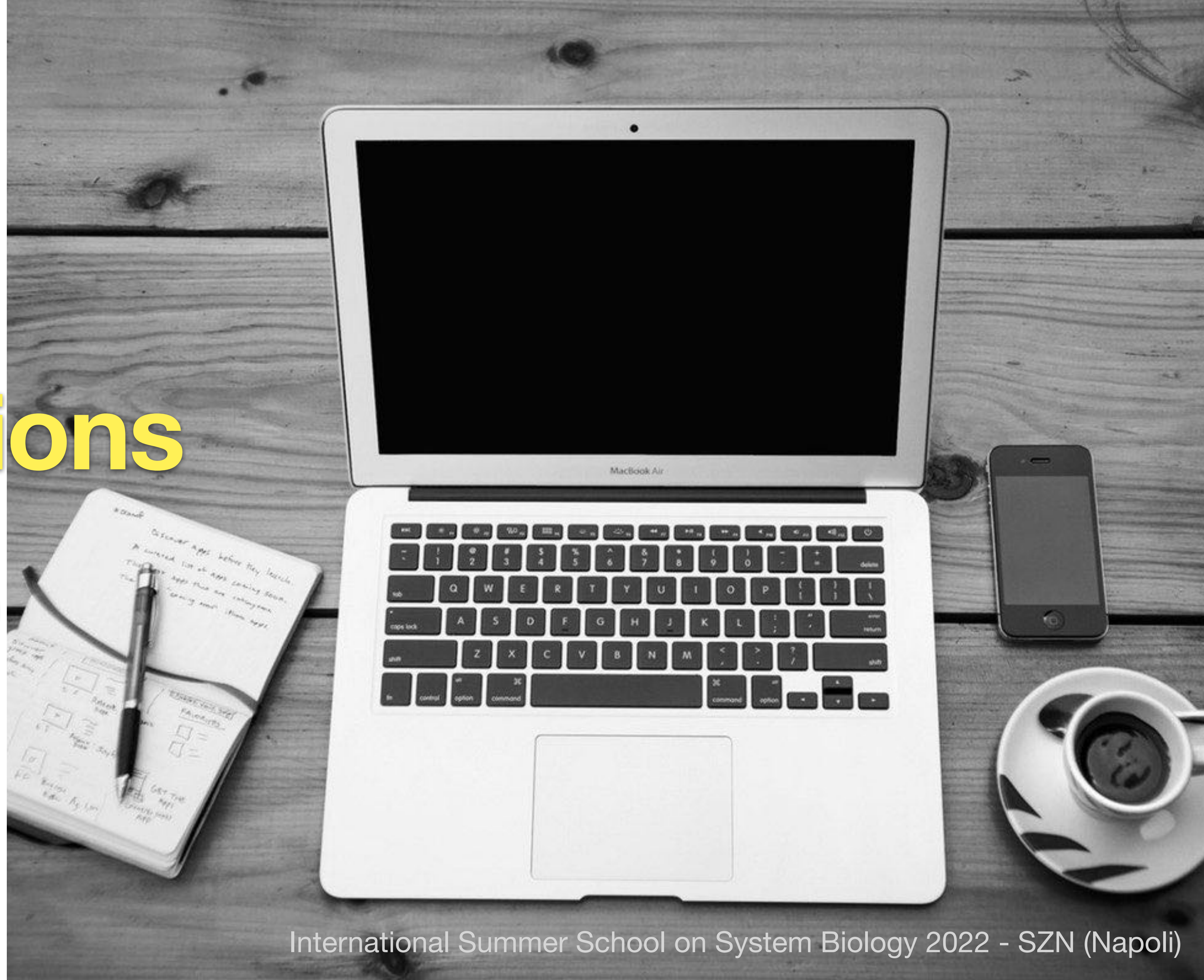
DST

DIPARTIMENTO DI
SCIENZE E TECNOLOGIE

Lab sessions

Applications

- Differential expression
- Promoter analysis
- Master Regulator



Research involving in-silico analysis

Classical workflows

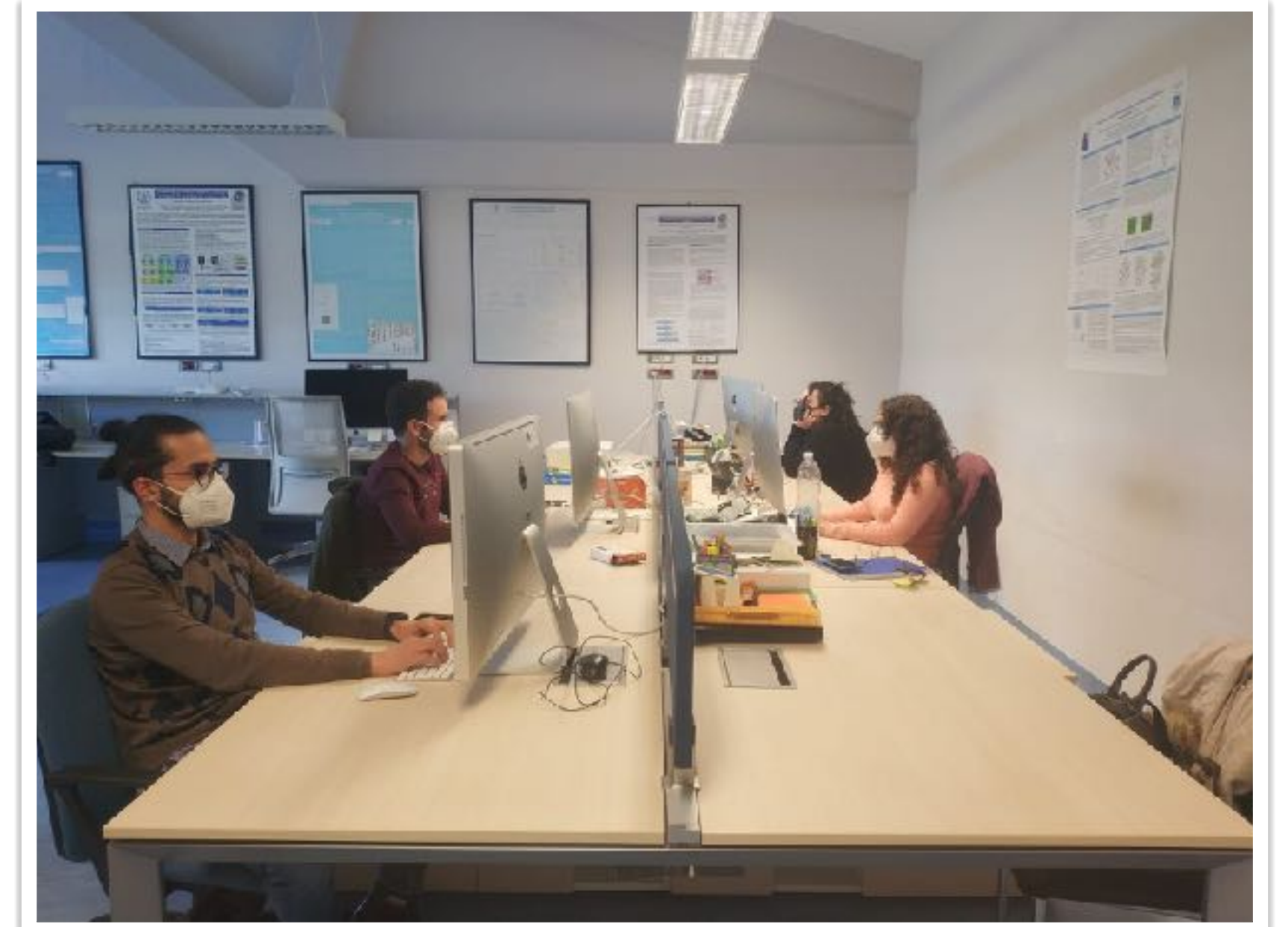
Biological Hypothesis supported by in-silico analyses



Research involving in-silico analysis

Classical workflows

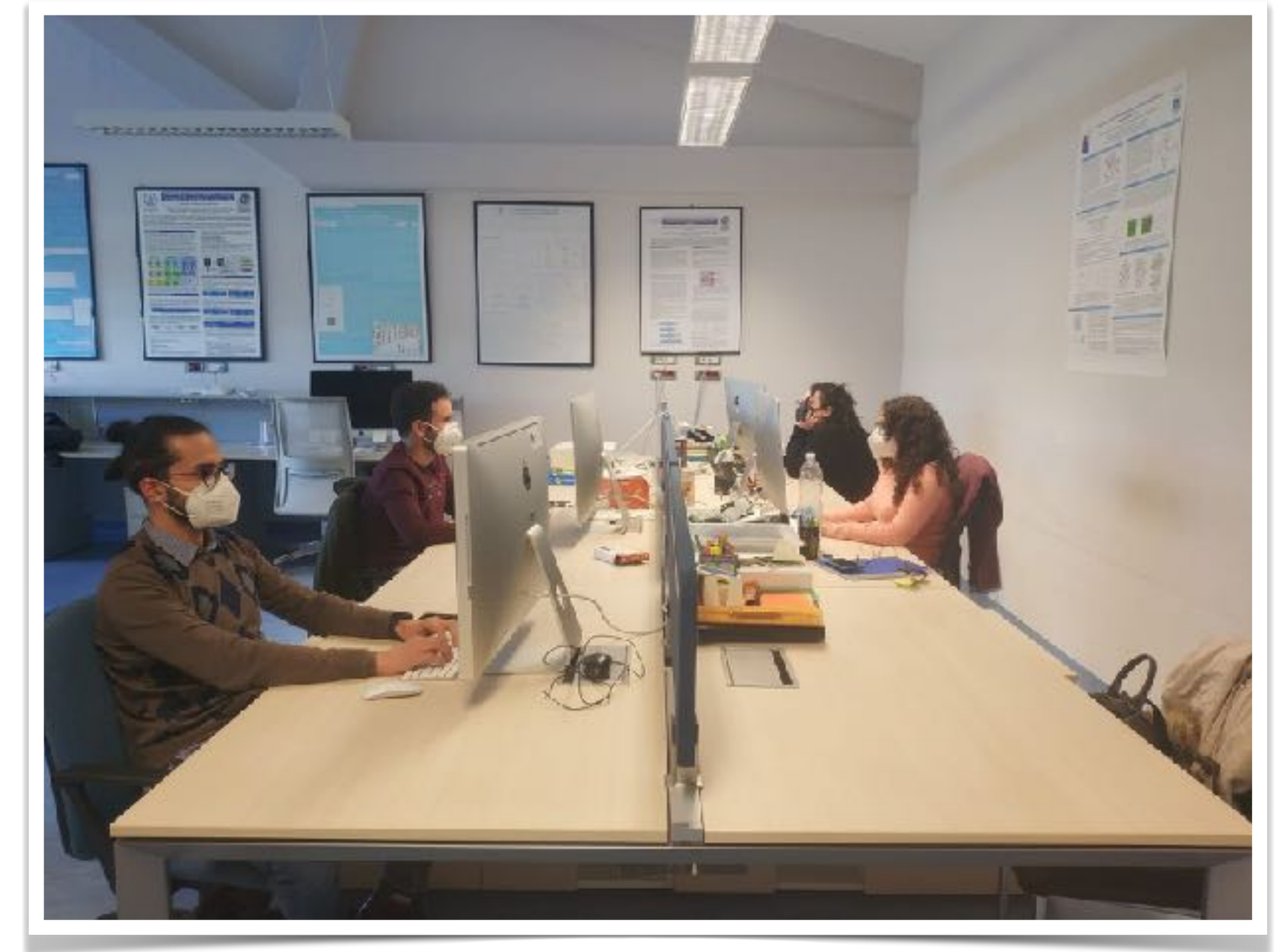
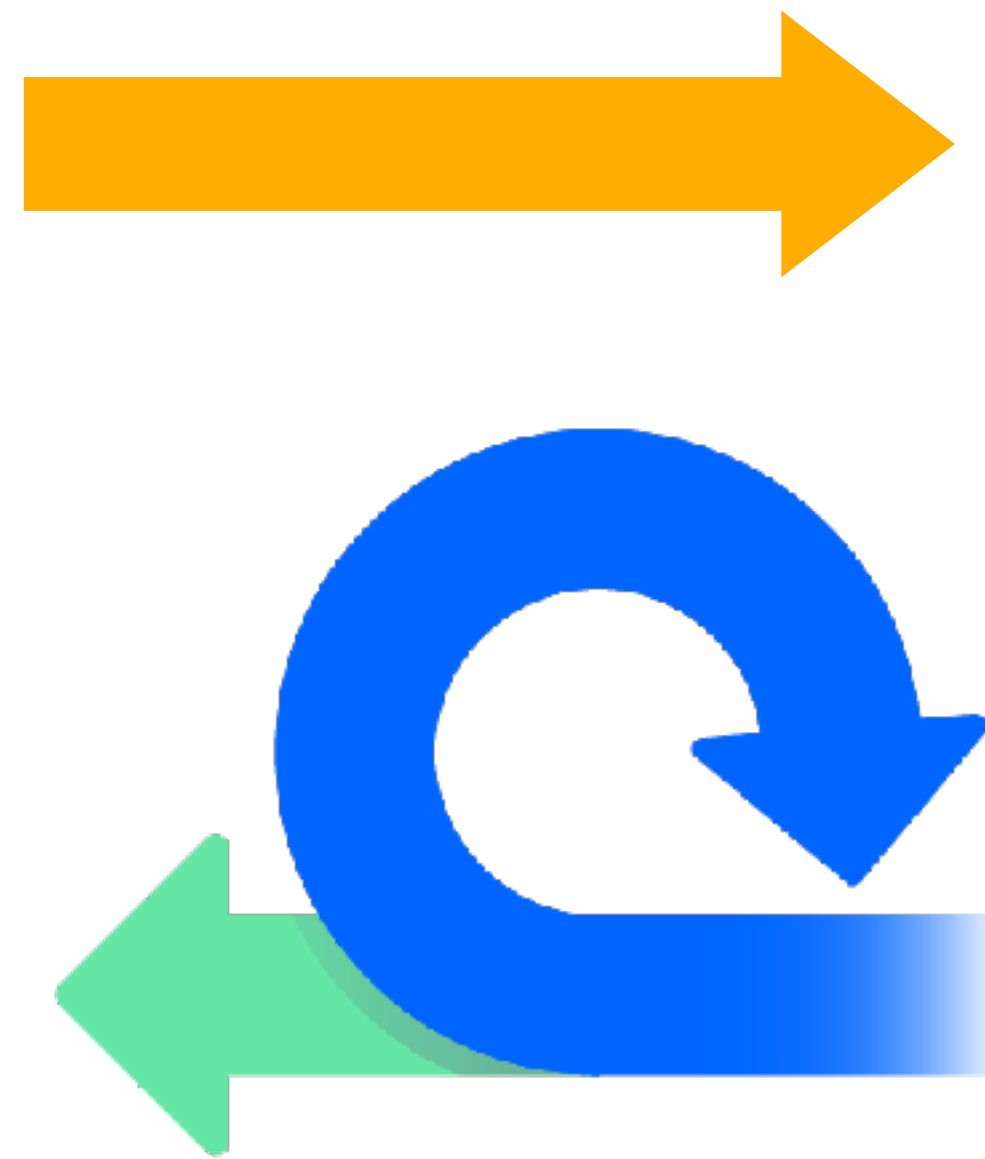
Biological Hypothesis supported by in-silico analyses



Research involving in-silico analysis

Classical workflows

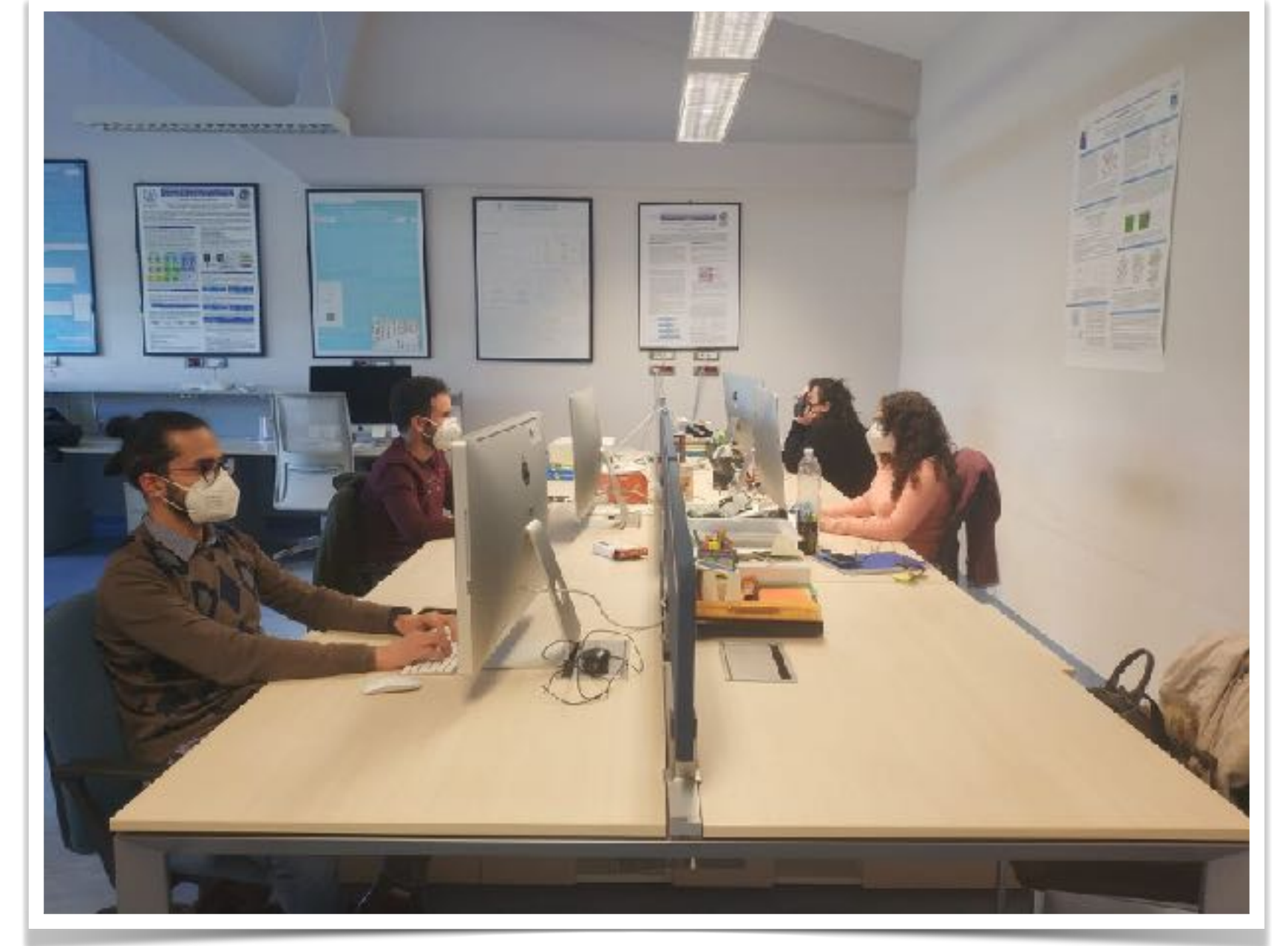
Biological Hypothesis supported by in-silico analyses



Research involving in-silico analysis

Classical workflows

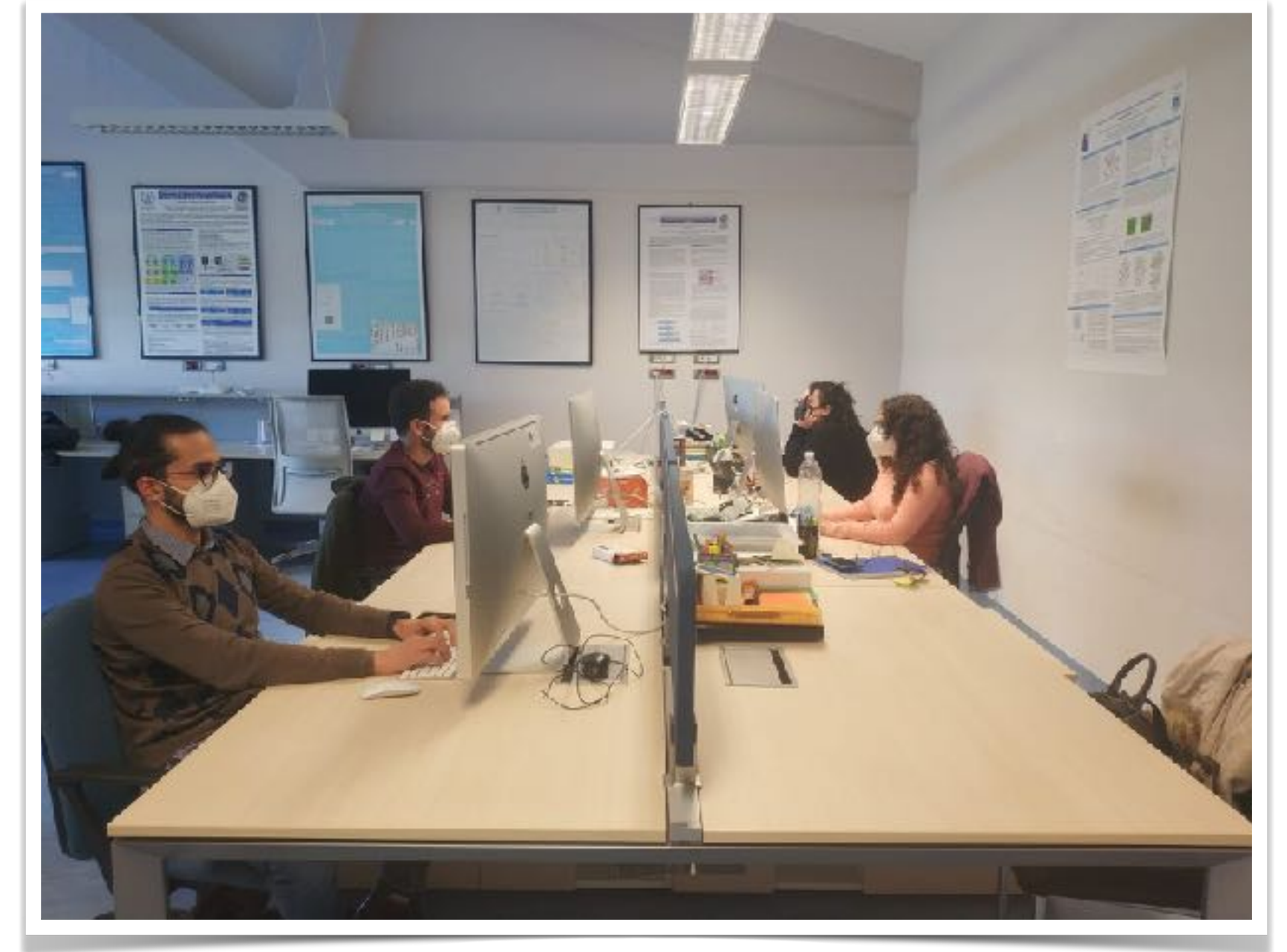
In-silico Hypothesis tested in wet Lab



Research involving in-silico analysis

Classical workflows

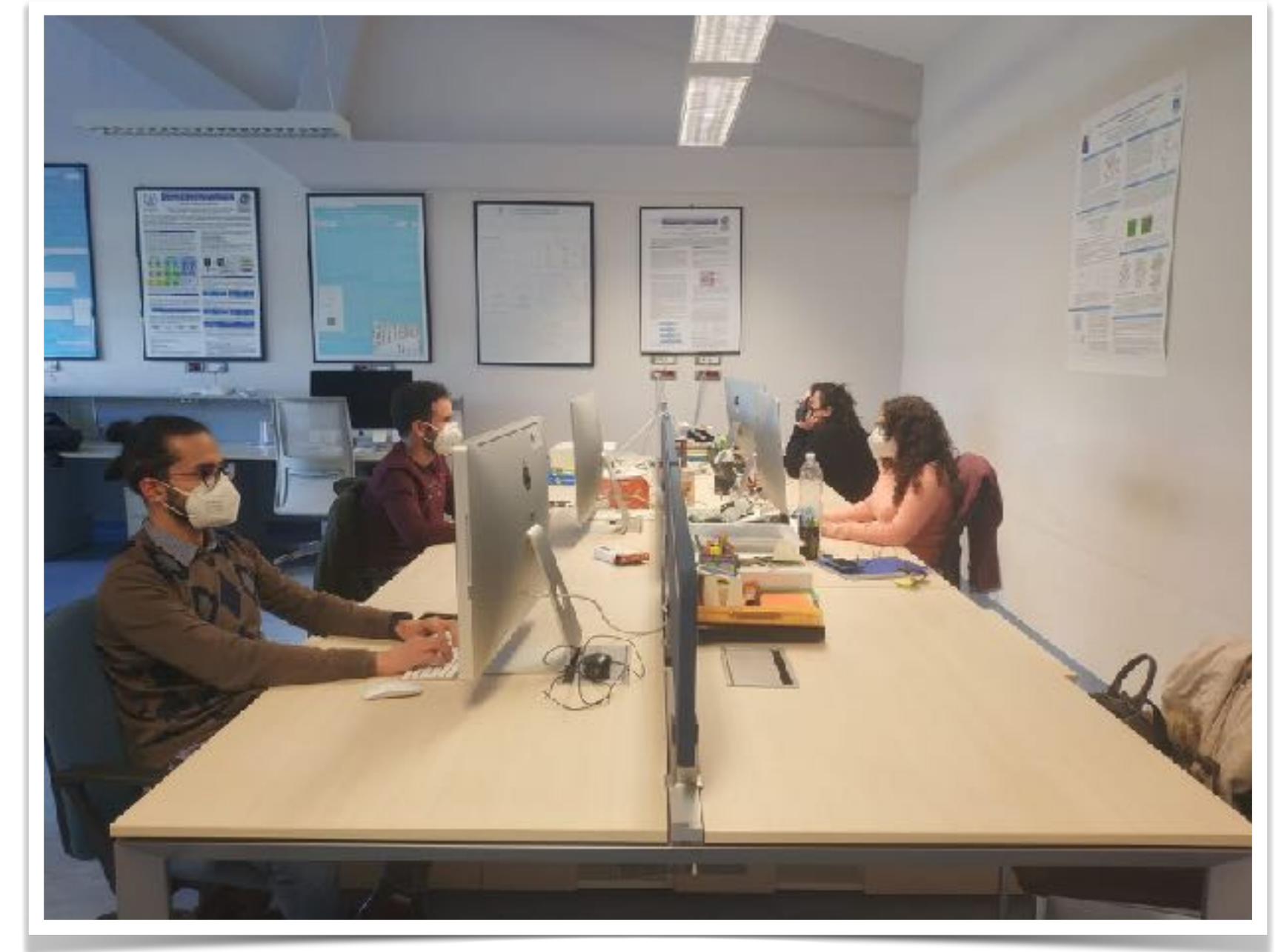
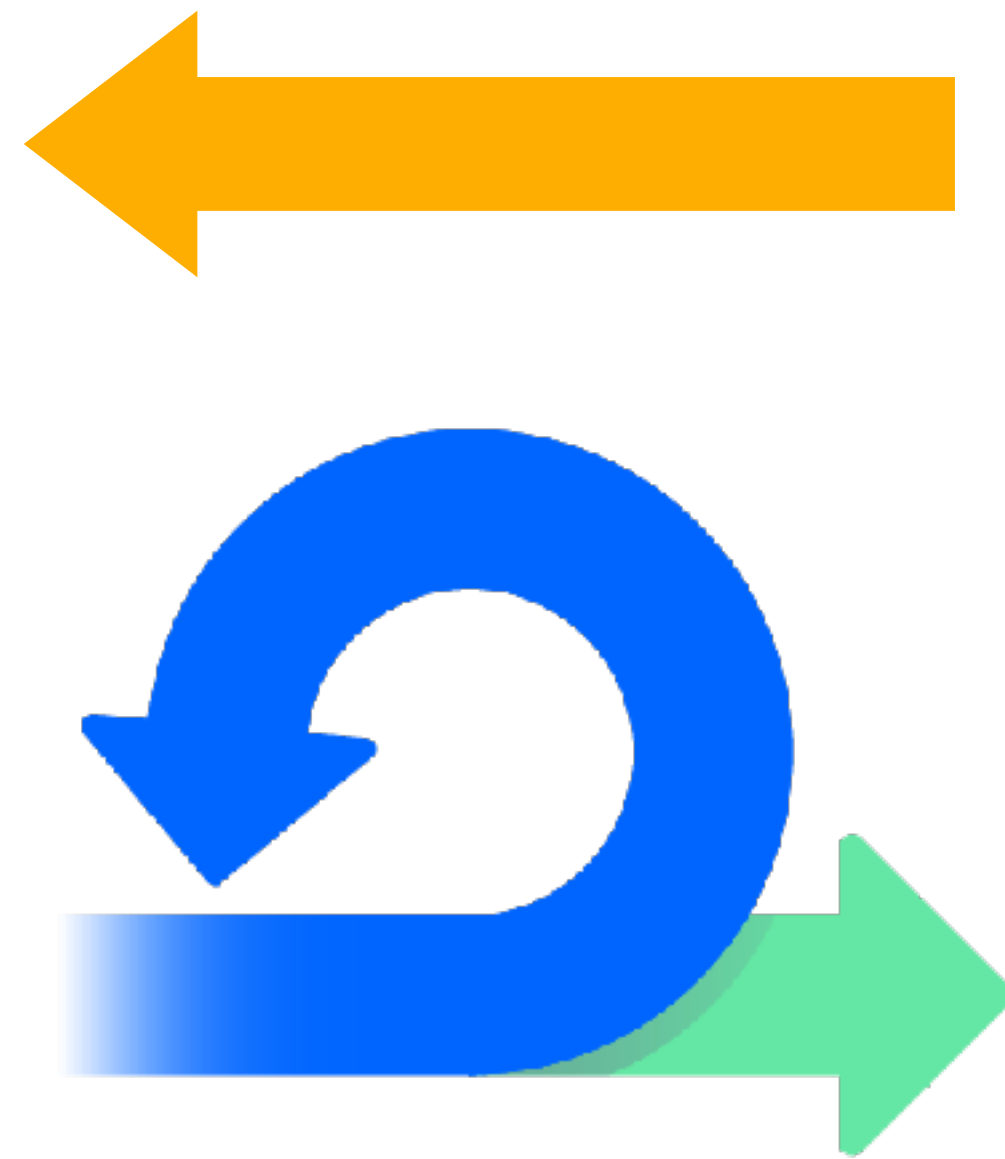
In-silico Hypothesis tested in wet Lab



Research involving in-silico analysis

Classical workflows

In-silico Hypothesis tested in wet Lab



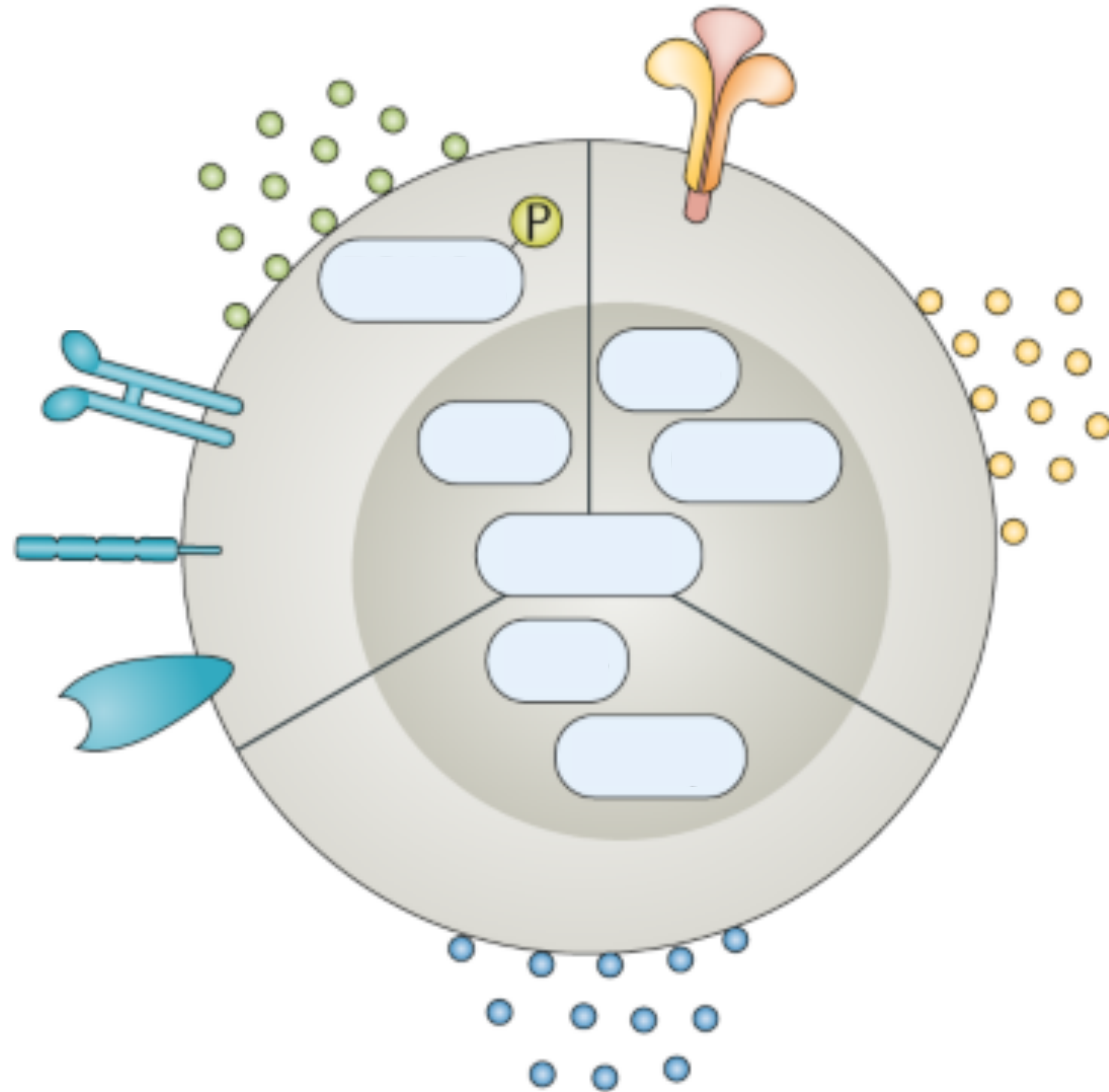
Research involving in-silico analysis

Integrating wet and dry labs



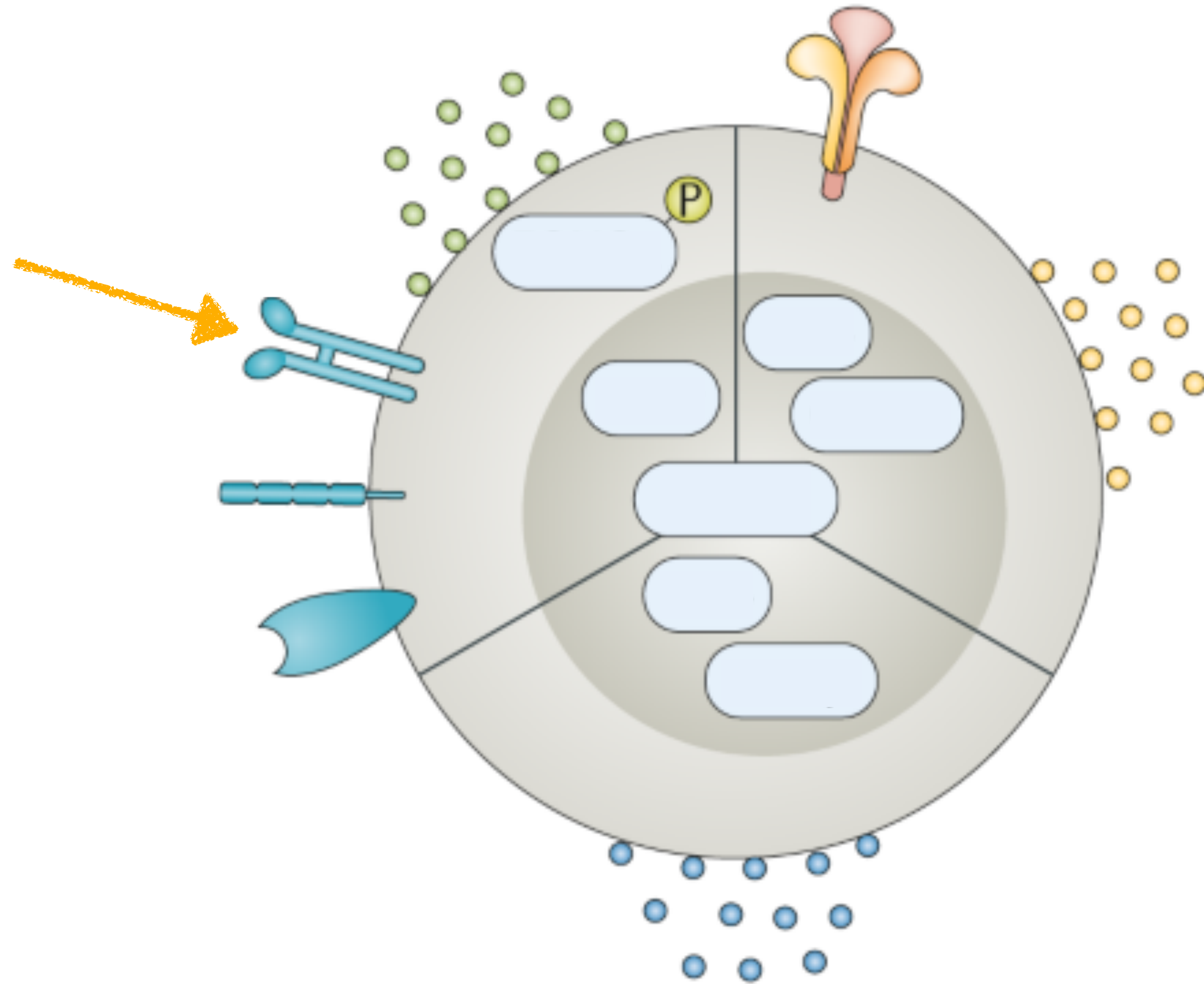
Dissect unknown molecular mechanisms

Typical workflow



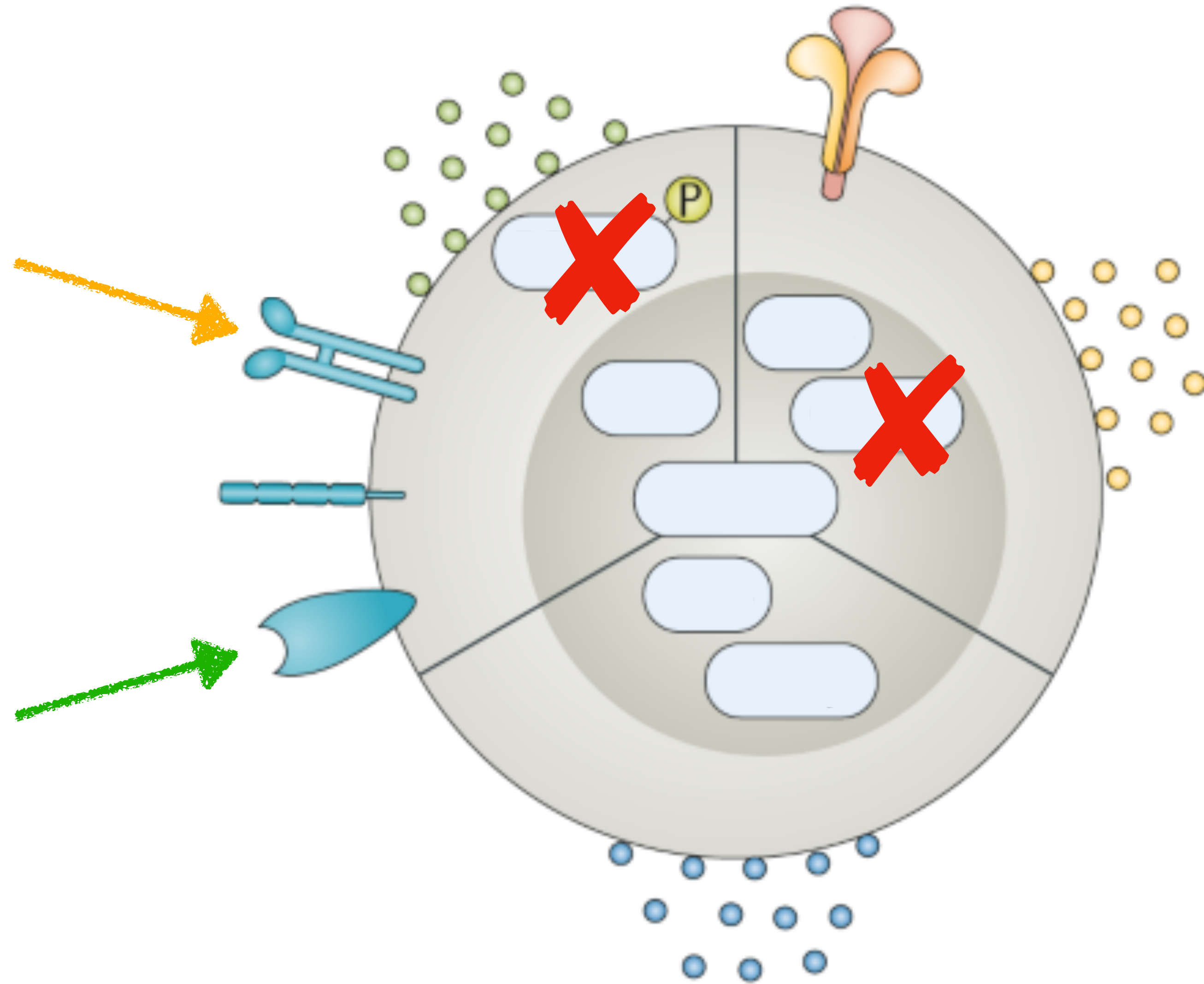
Dissect unknown molecular mechanisms

Typical workflow



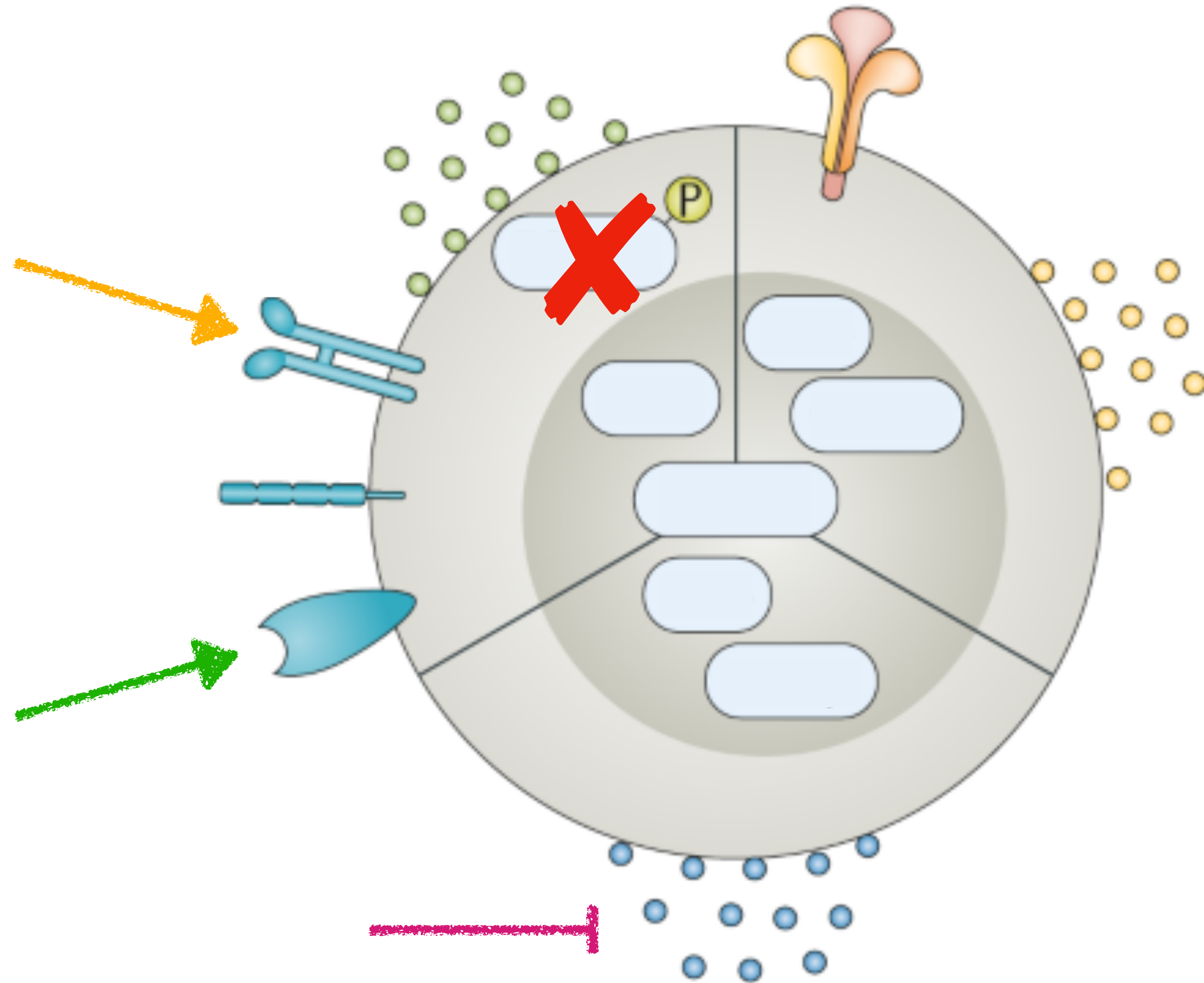
Dissect unknown molecular mechanisms

Typical workflow



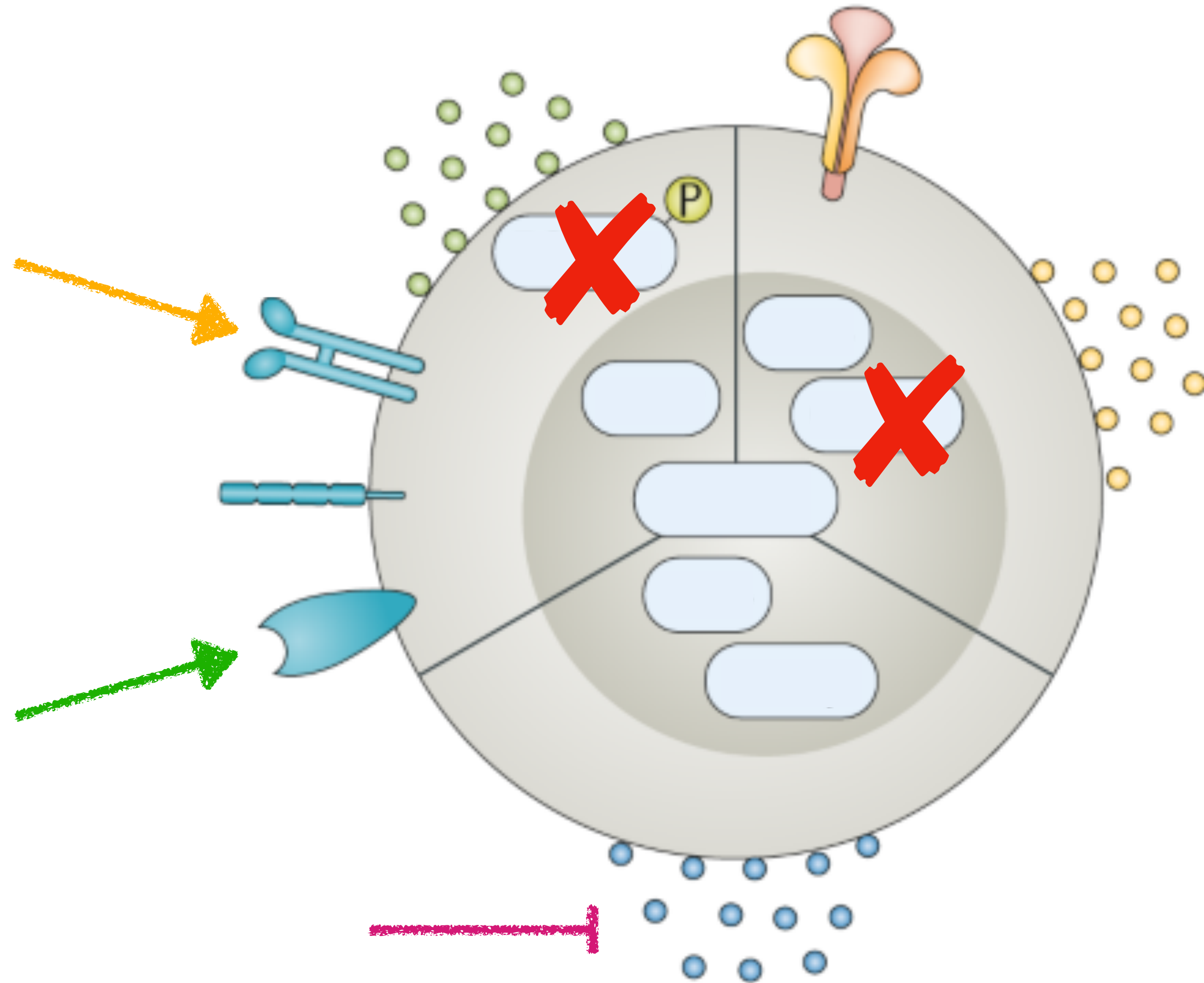
Dissect unknown molecular mechanisms

Typical workflow



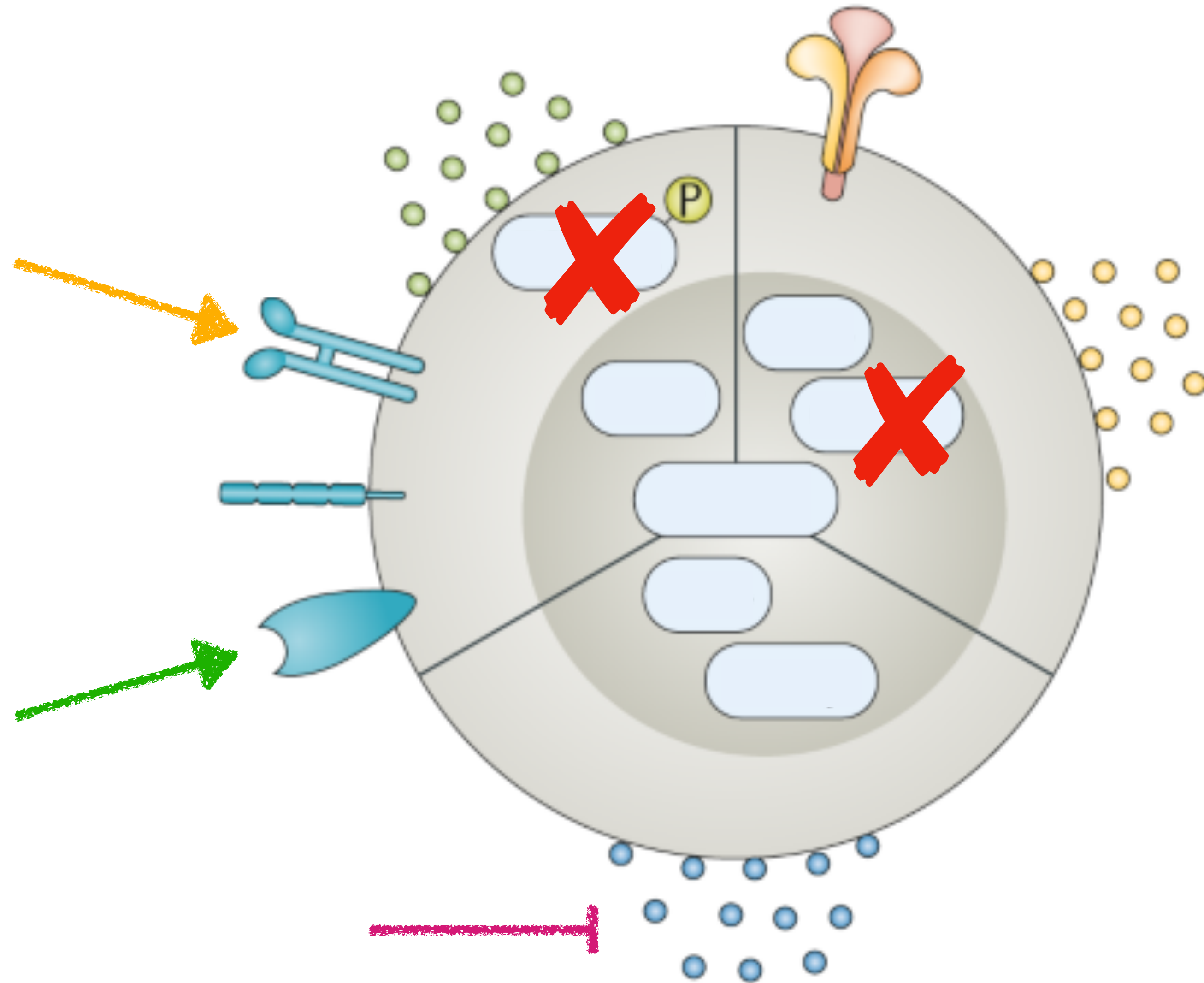
Dissect unknown molecular mechanisms

Typical workflow



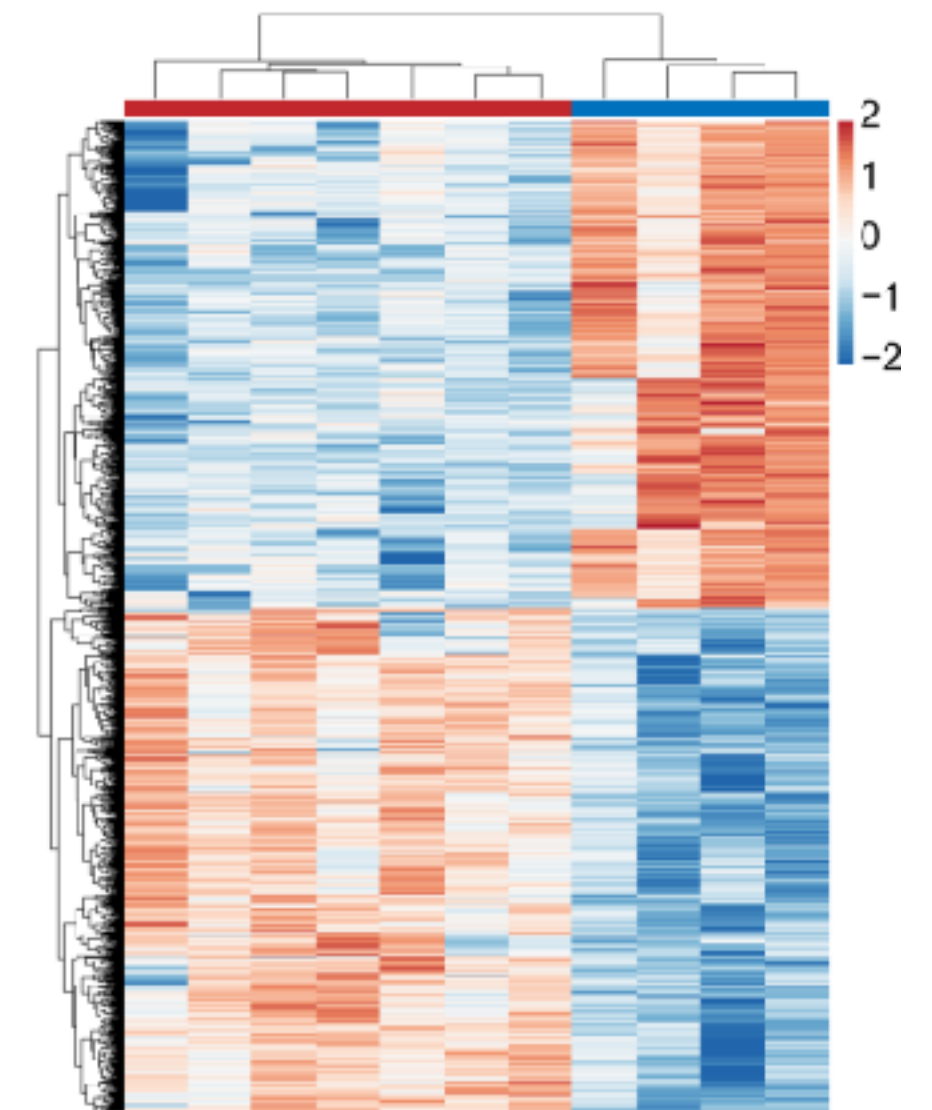
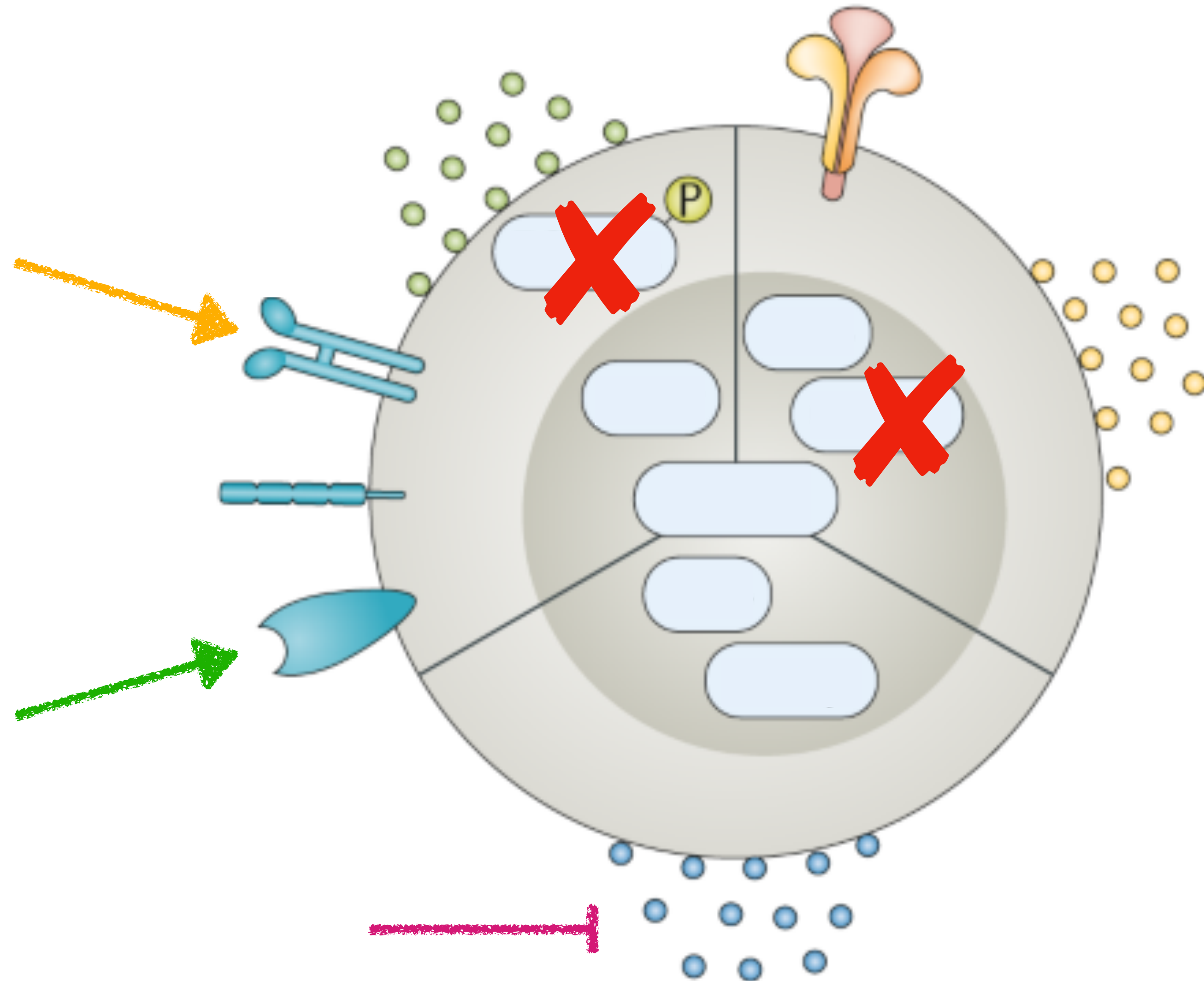
Dissect unknown molecular mechanisms

Typical workflow



Dissect unknown molecular mechanisms

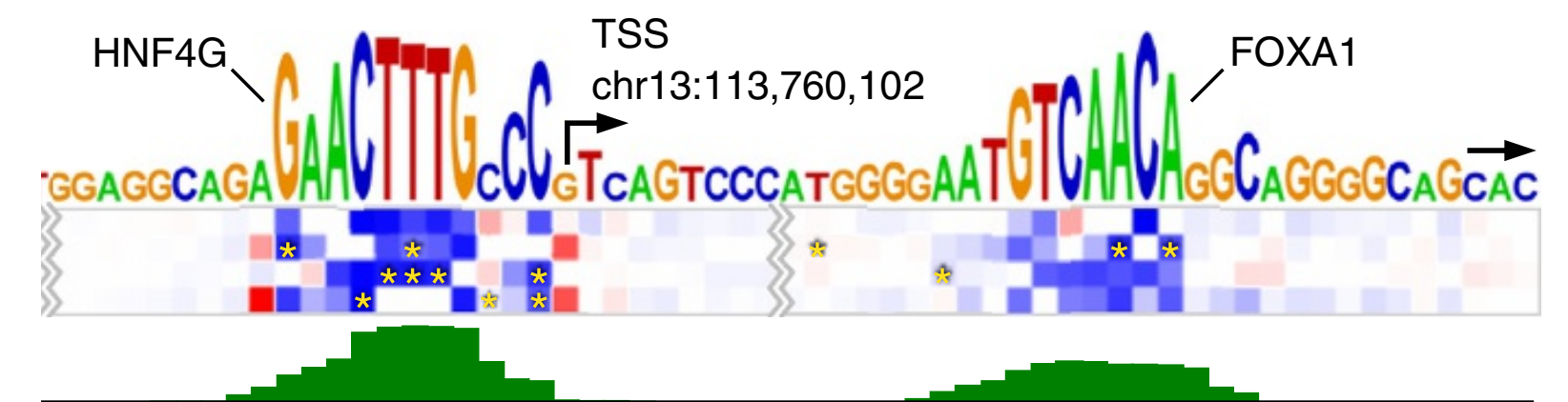
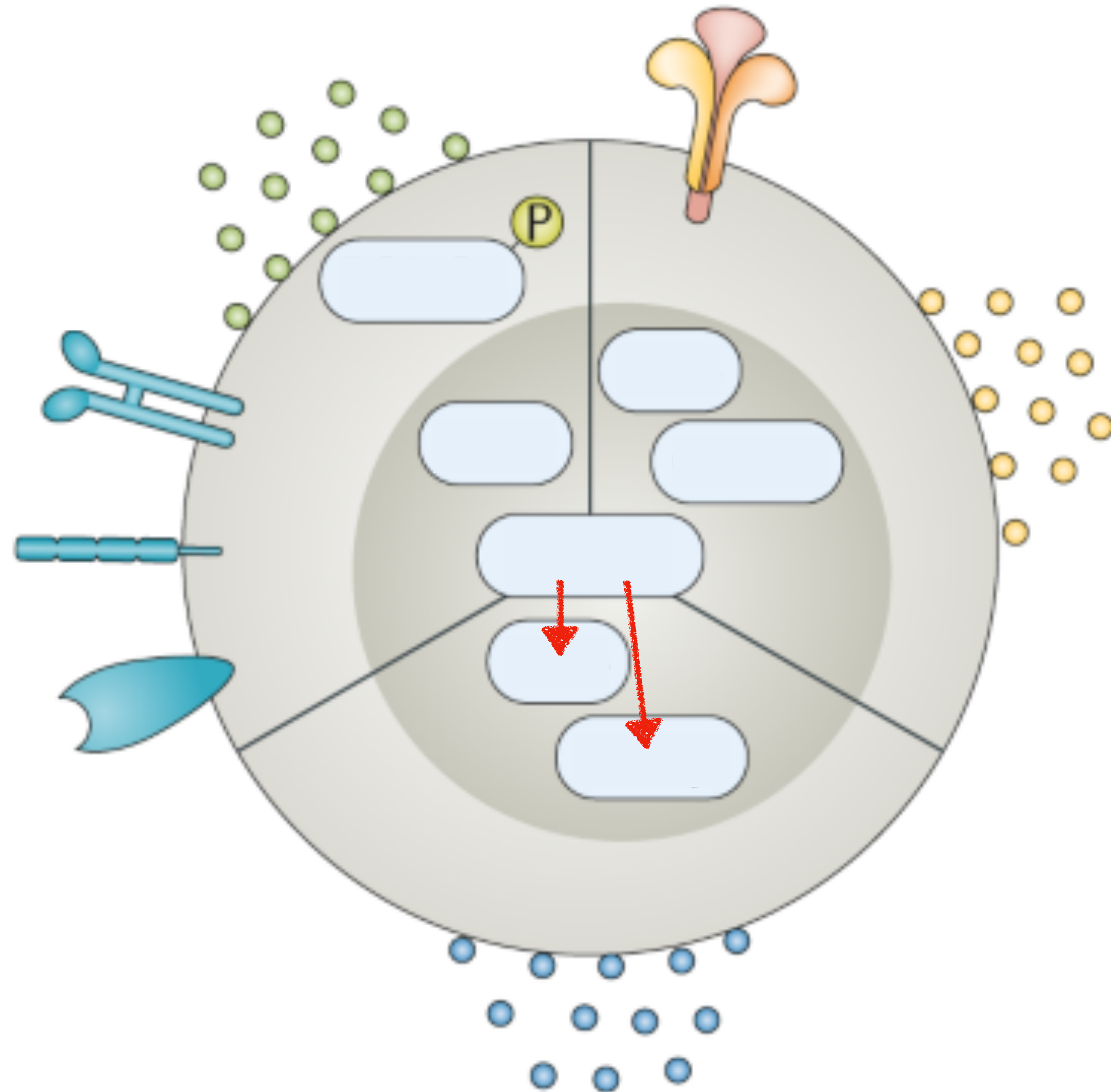
Typical workflow



Differential expression Analysis

Dissect unknown molecular mechanisms

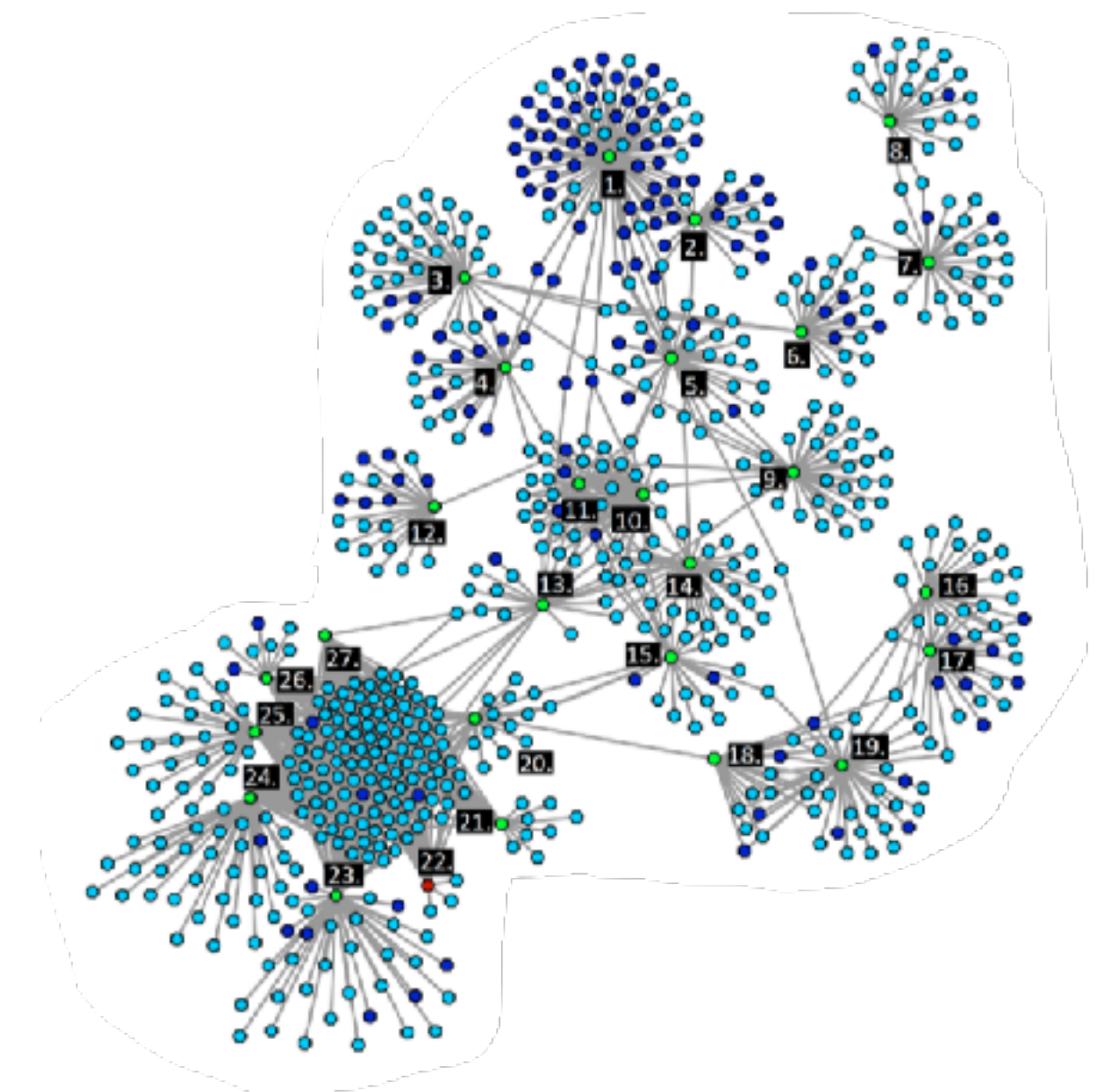
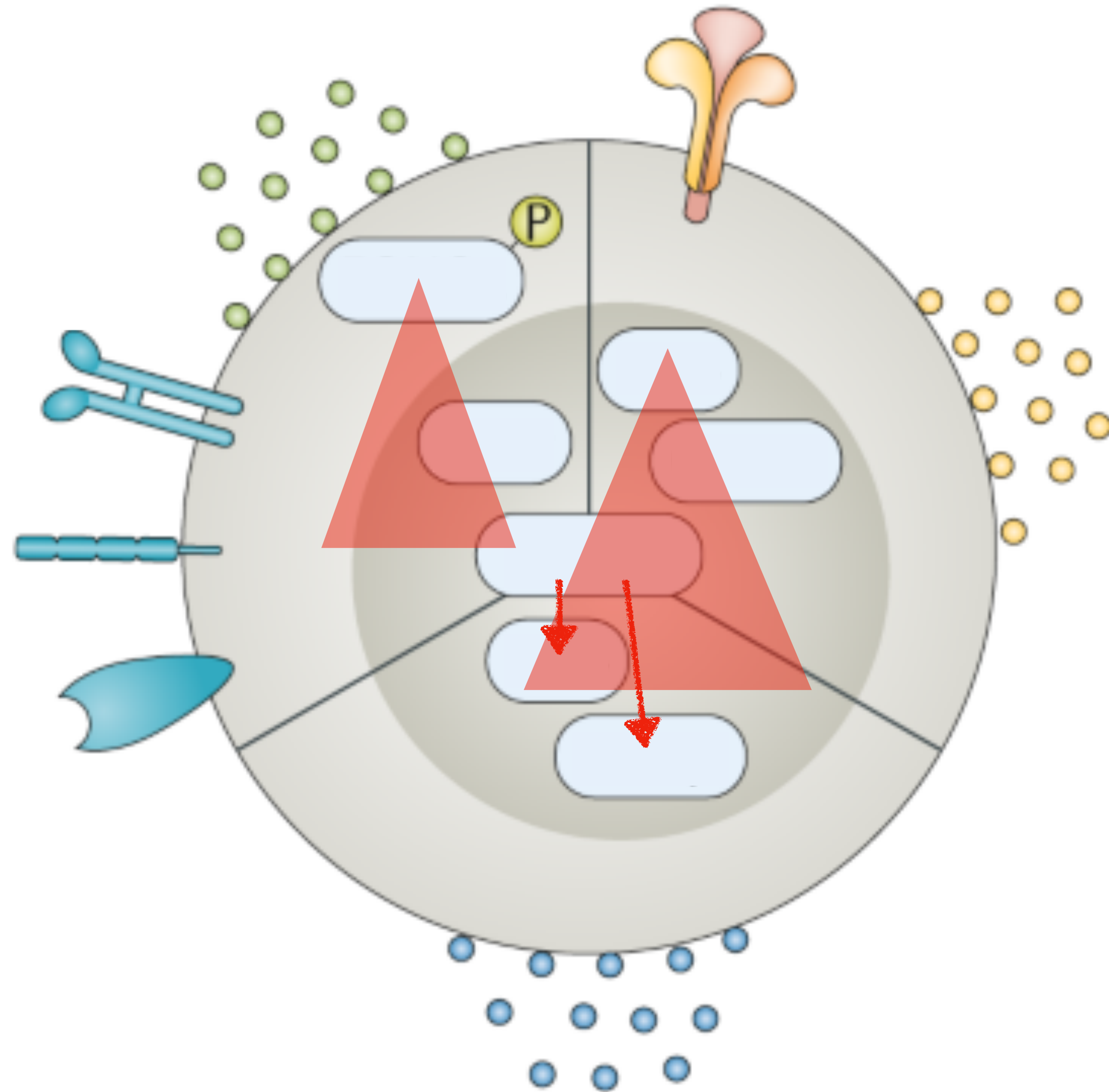
Typical workflow



Promoter Analysis

Dissect unknown molecular mechanisms

Typical workflow



Master Regulator Analysis

Dissect unknown molecular mechanisms

Master regulator analysis

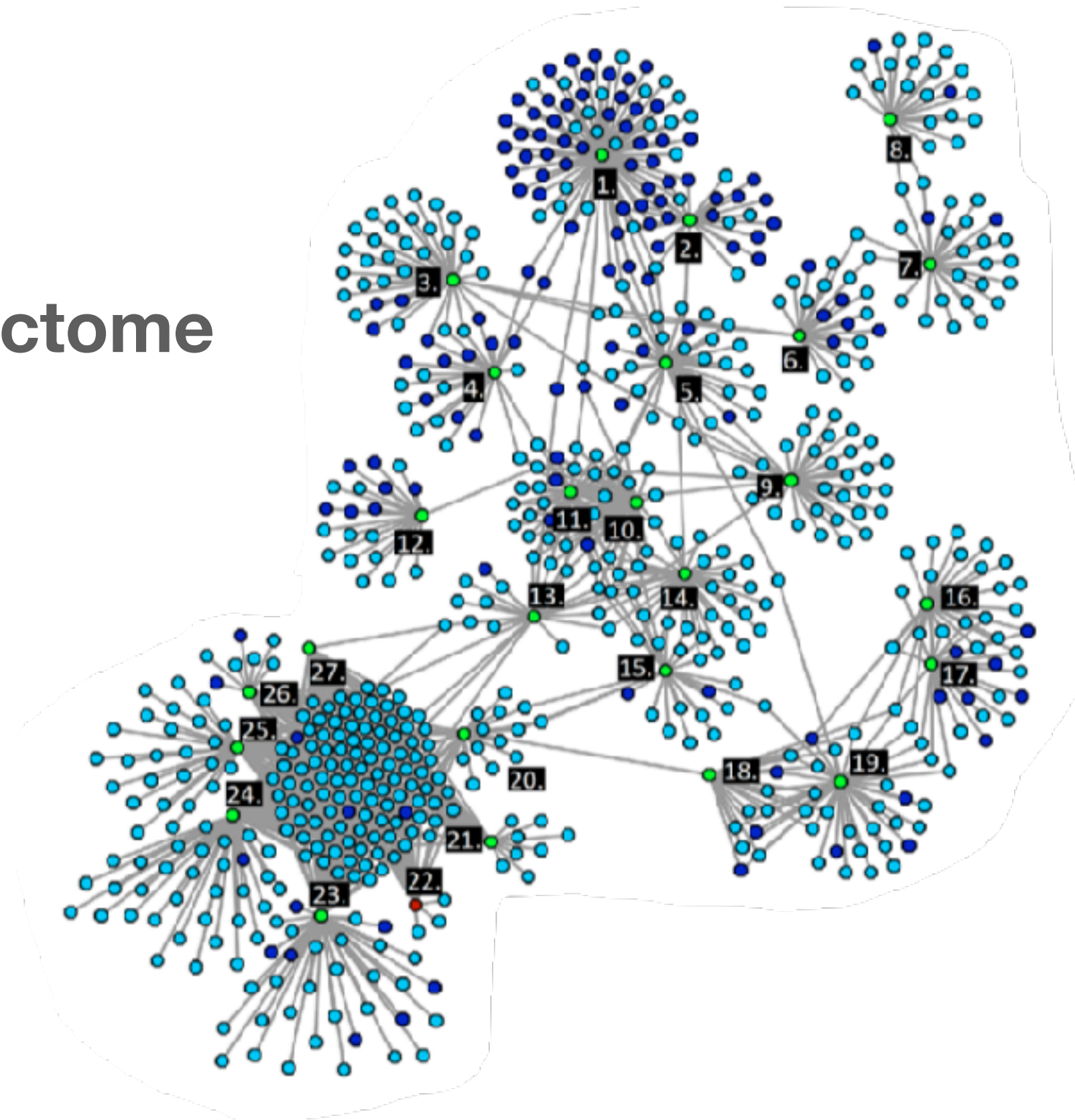
Dissect unknown molecular mechanisms

Master regulator analysis

1

Build a gene regulatory network active in the context of interest (e.g. using public data)

Interactome



Dissect unknown molecular mechanisms

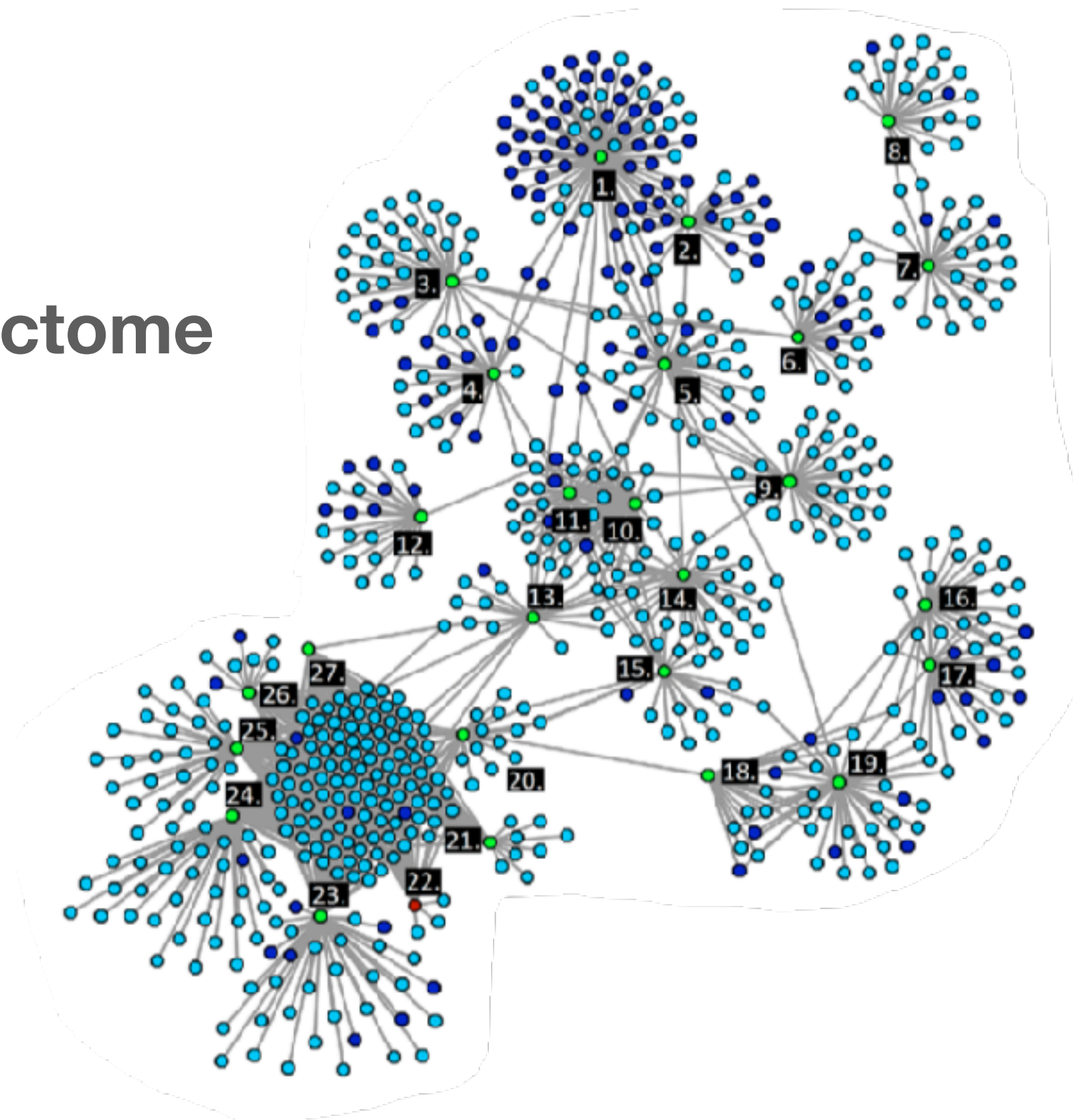
Master regulator analysis

1

Build a gene regulatory network active in the context of interest (e.g. using public data)

Or adopt a known static network (e.g. from STRING)

Interactome



Dissect unknown molecular mechanisms

Master regulator analysis

1

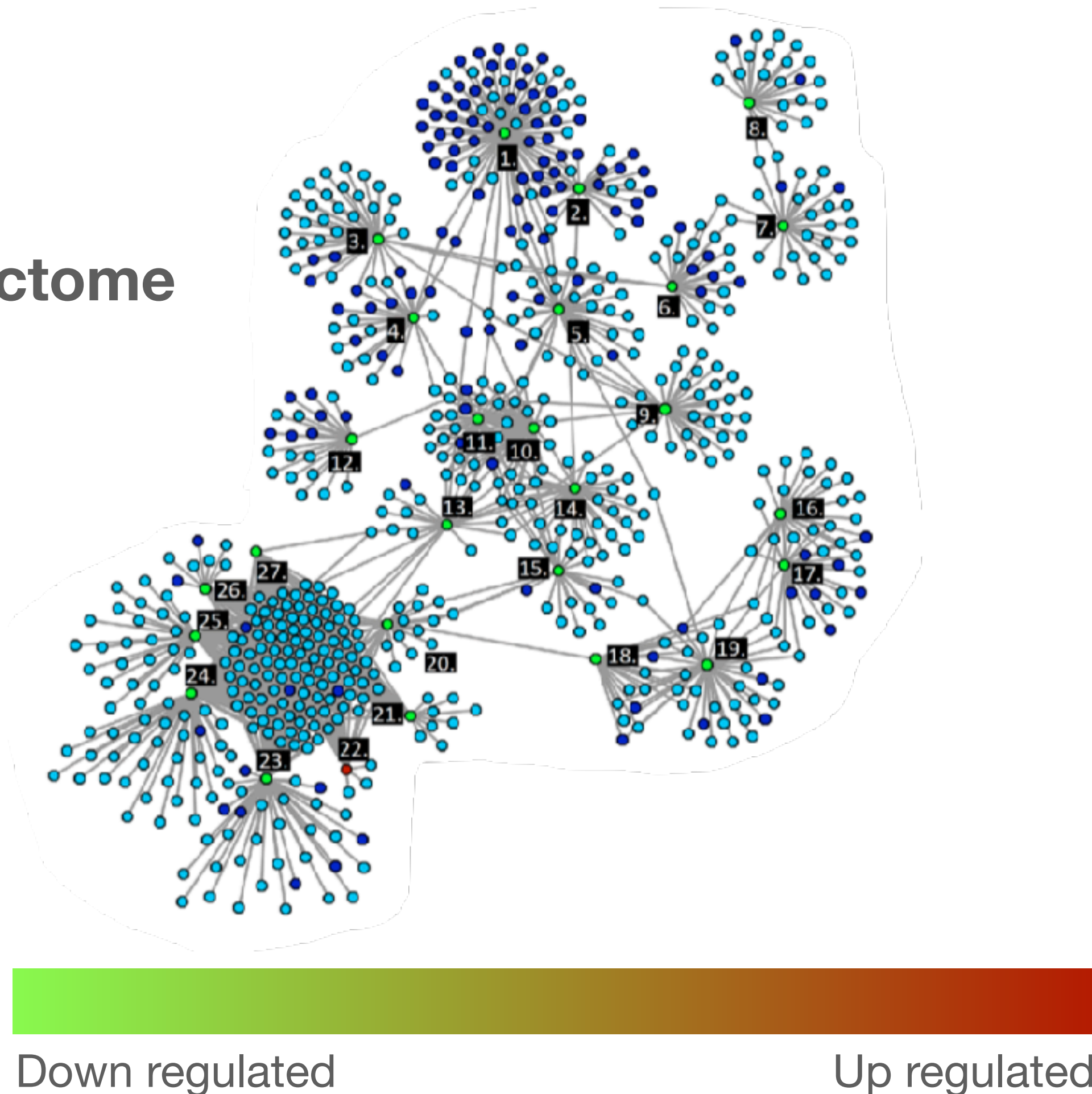
Build a gene regulatory network active in the context of interest (e.g. using public data)

Or adopt a known static network (e.g. from STRING)

2

Perform a transcriptomic analysis that activate the phenomenon of interest (e.g. differential expr. analysis)

Interactome



Dissect unknown molecular mechanisms

Master regulator analysis

1

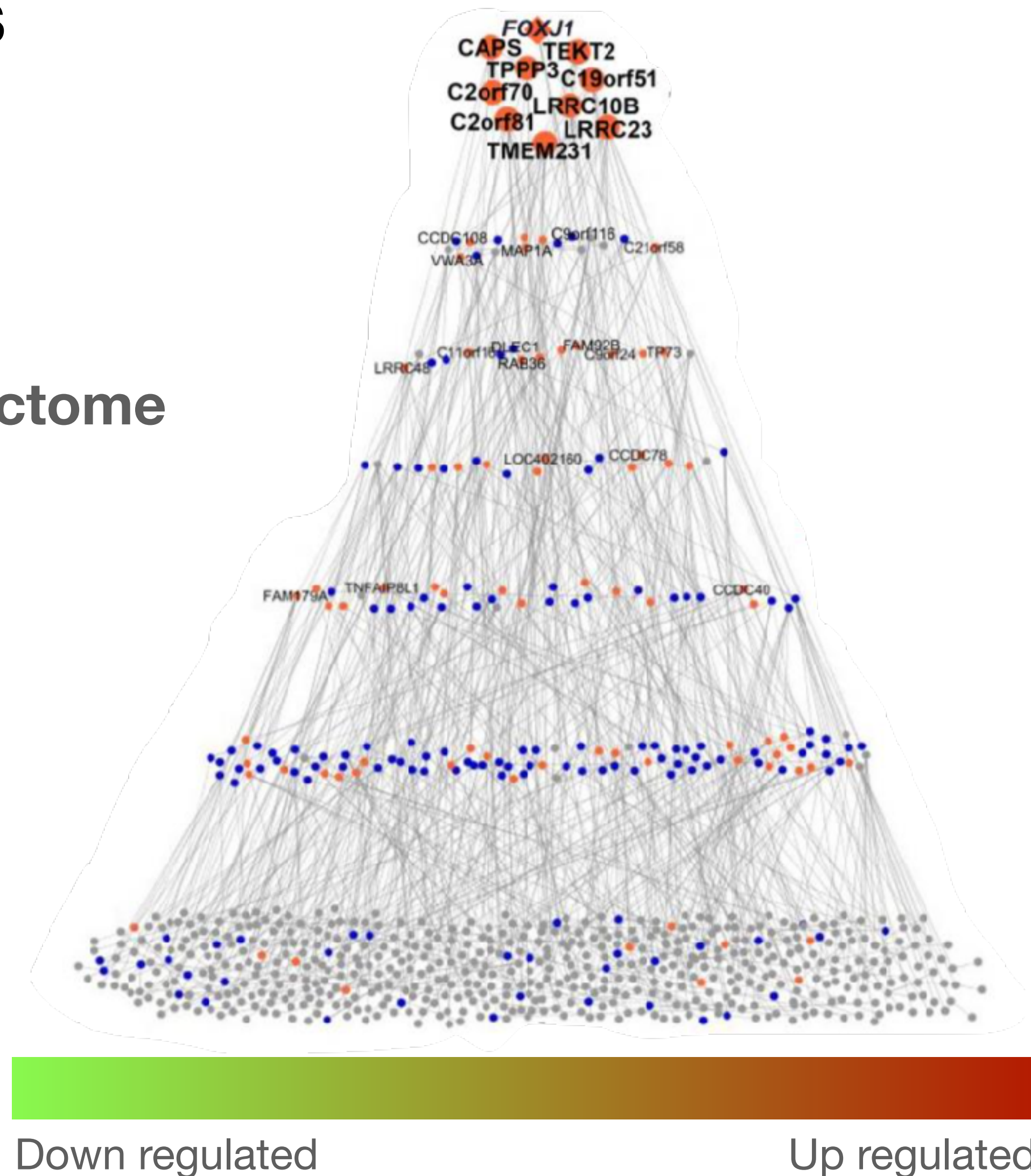
Build a gene regulatory network active in the context of interest (e.g. using public data)

Or adopt a known static network (e.g. from STRING)

2

Perform a transcriptomic analysis that activate the phenomenon of interest (e.g. differential expr. analysis)

Interactome



Dissect unknown molecular mechanisms

Master regulator analysis

1

Build a gene regulatory network active in the context of interest (e.g. using public data)

Or adopt a known static network (e.g. from STRING)

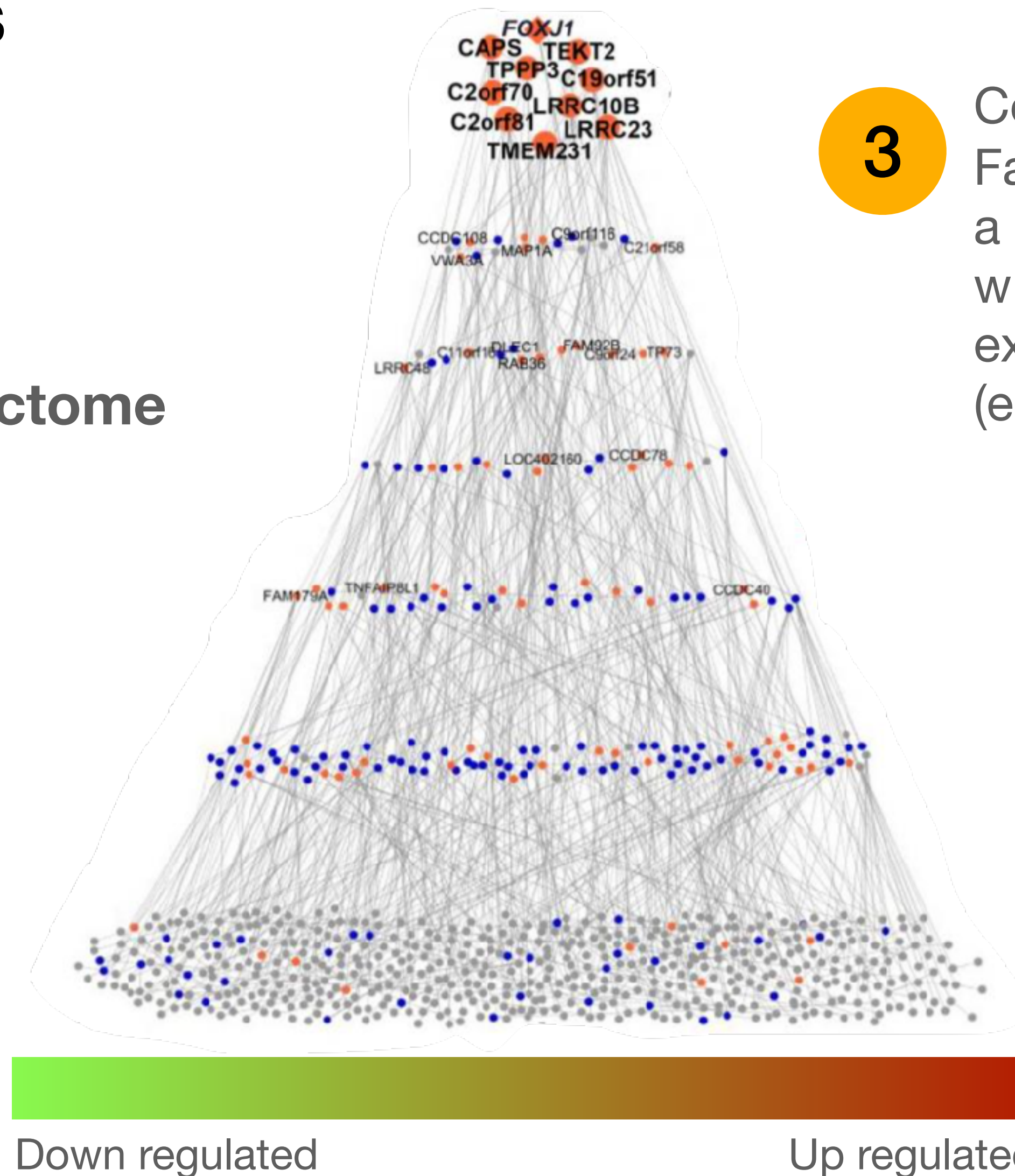
2

Perform a transcriptomic analysis that activate the phenomenon of interest (e.g. differential expr. analysis)

3

Collect all Transcription Factor **regulons** and perform a ranked enrichment analysis with the set of differential expressed genes (e.g. Wilcoxon ranked test)

Interactome



Dissect unknown molecular mechanisms

Master regulator analysis

1

Build a gene regulatory network active in the context of interest (e.g. using public data)

Or adopt a known static network (e.g. from STRING)

2

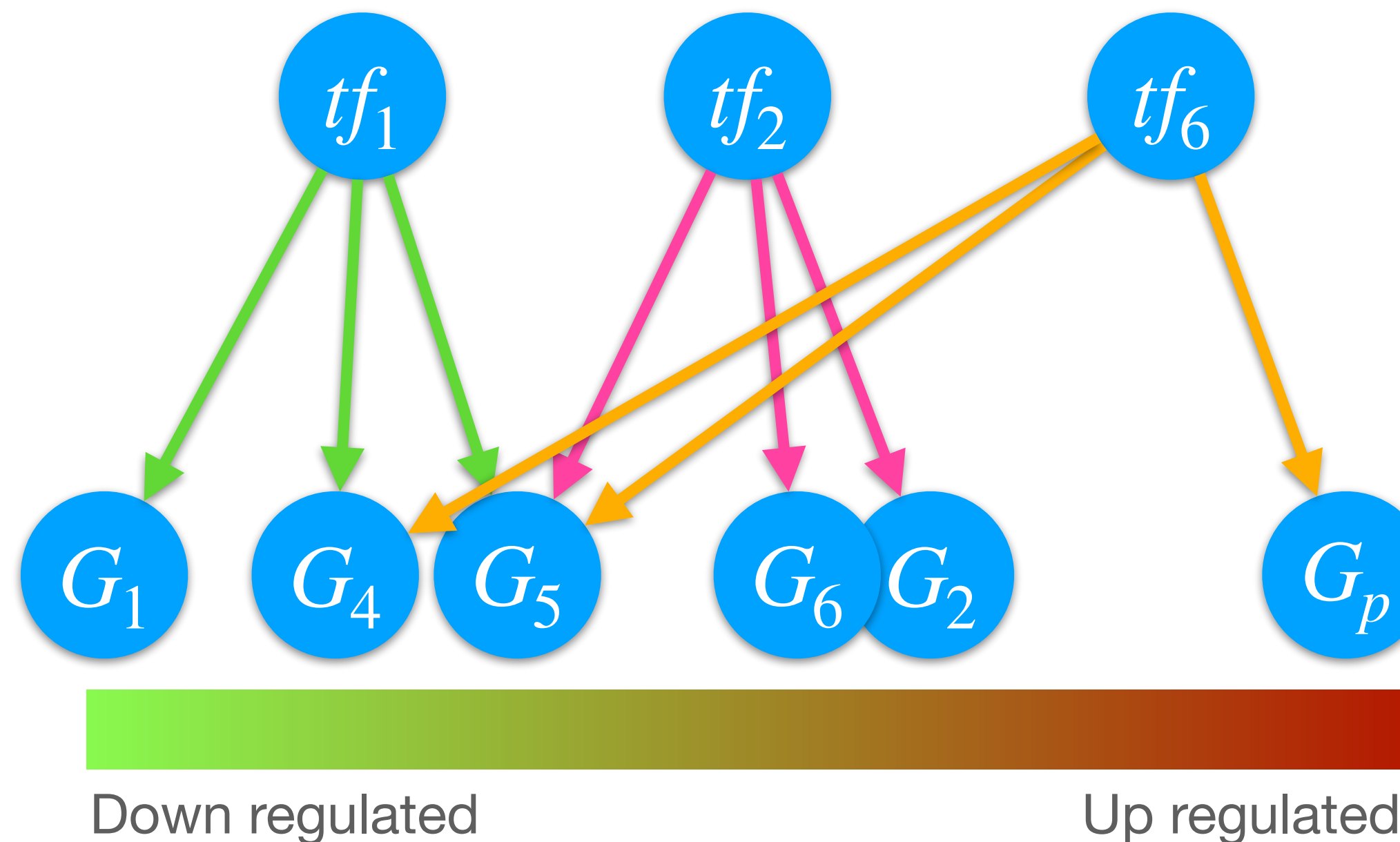
Perform a transcriptomic analysis that activate the phenomenon of interest (e.g. differential expr. analysis)

3

Collect all Transcription Factor **regulons** and perform a ranked enrichment analysis with the set of differential expressed genes (e.g. Wilcoxon ranked test)

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

Interactome



Dissect unknown molecular mechanisms

Master regulator analysis

1

Build a gene regulatory network active in the context of interest (e.g. using public data)

Or adopt a known static network (e.g. from STRING)

2

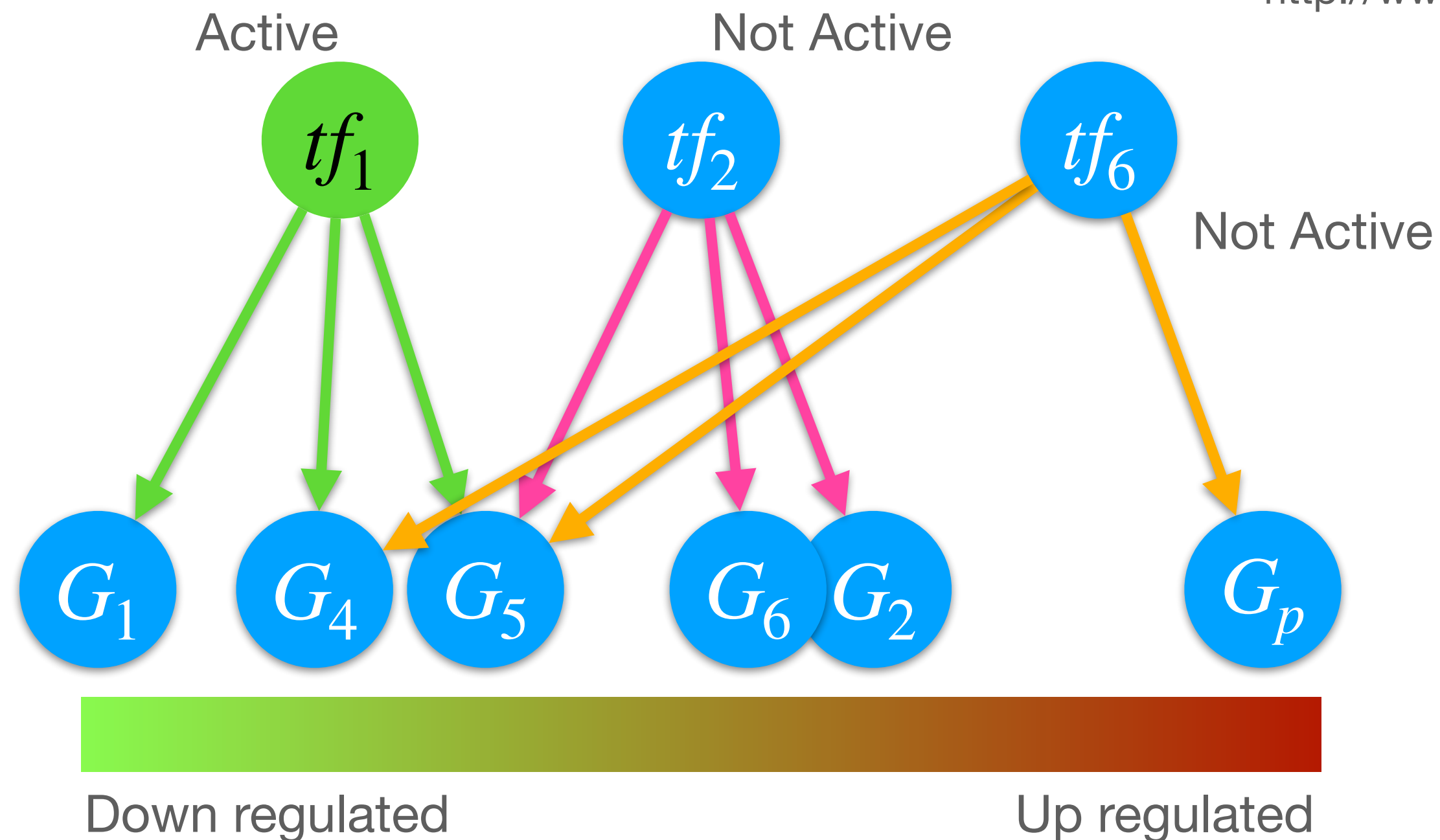
Perform a transcriptomic analysis that activate the phenomenon of interest (e.g. differential expr. analysis)

3

Collect all Transcription Factor **regulons** and perform a ranked enrichment analysis with the set of differential expressed genes (e.g. Wilcoxon ranked test)

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

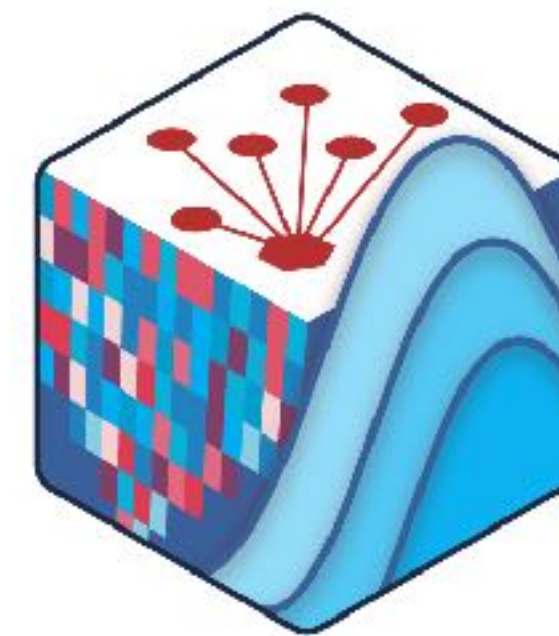
Interactome



Reverse engineering of gene regulatory nets

SCENIC

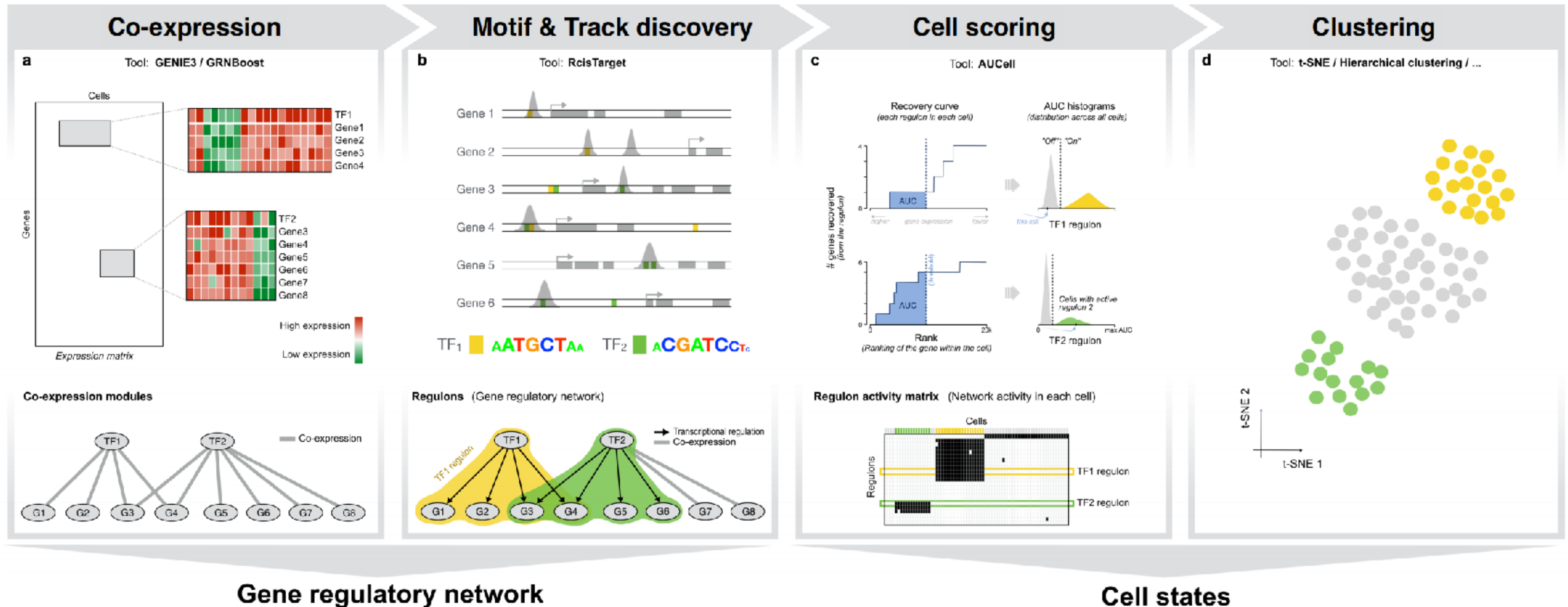
SCENIC is a suite of tools to study and decipher gene regulation (with GENIE3) to infer transcription factors, gene regulatory networks, and cell types from single-cell RNA-seq data or the combination of single-cell RNA-seq and single-cell ATAC-seq data.



SCENIC+
Single-cell enhancer-gene
regulatory networks

Reverse engineering of gene regulatory nets

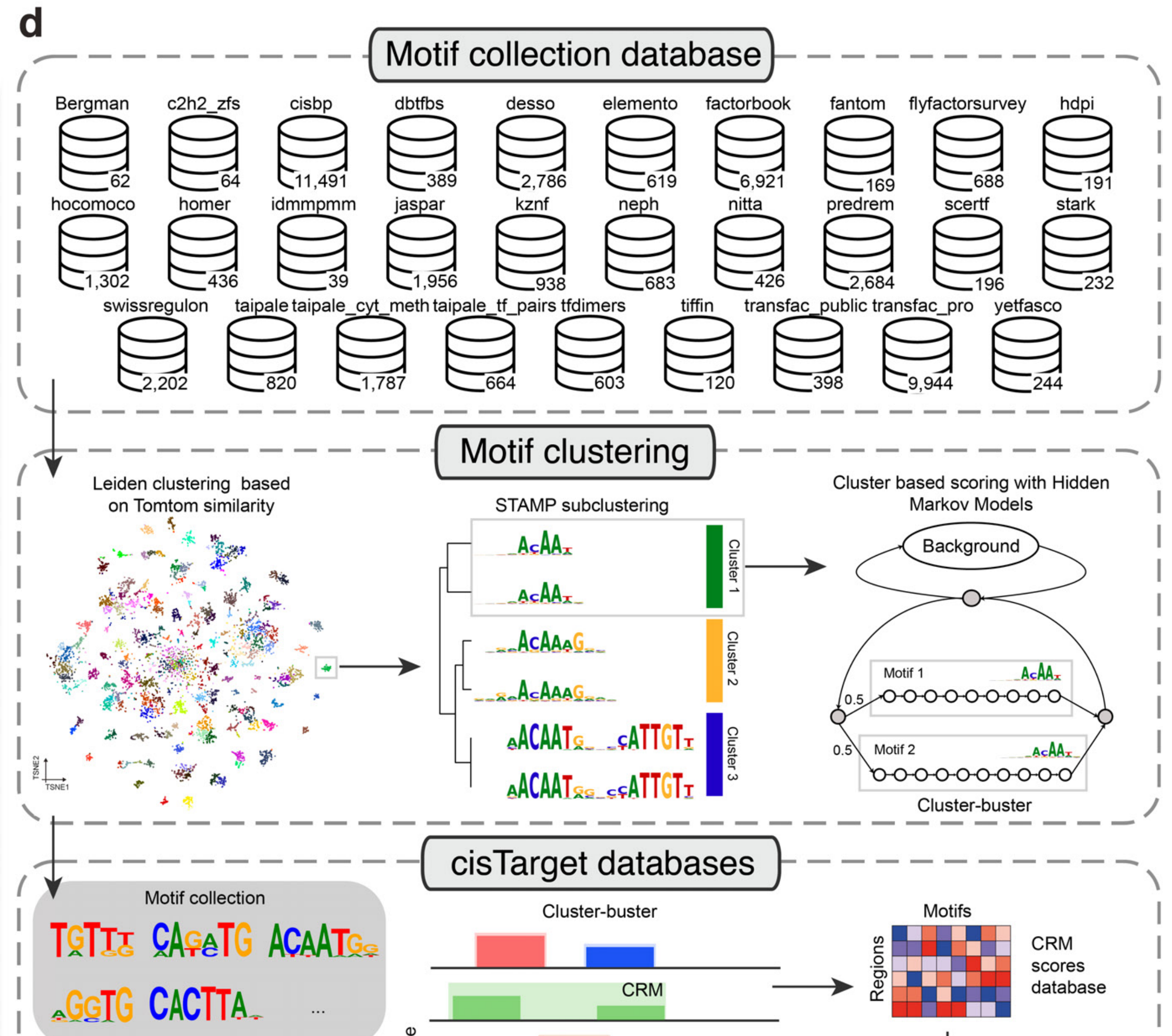
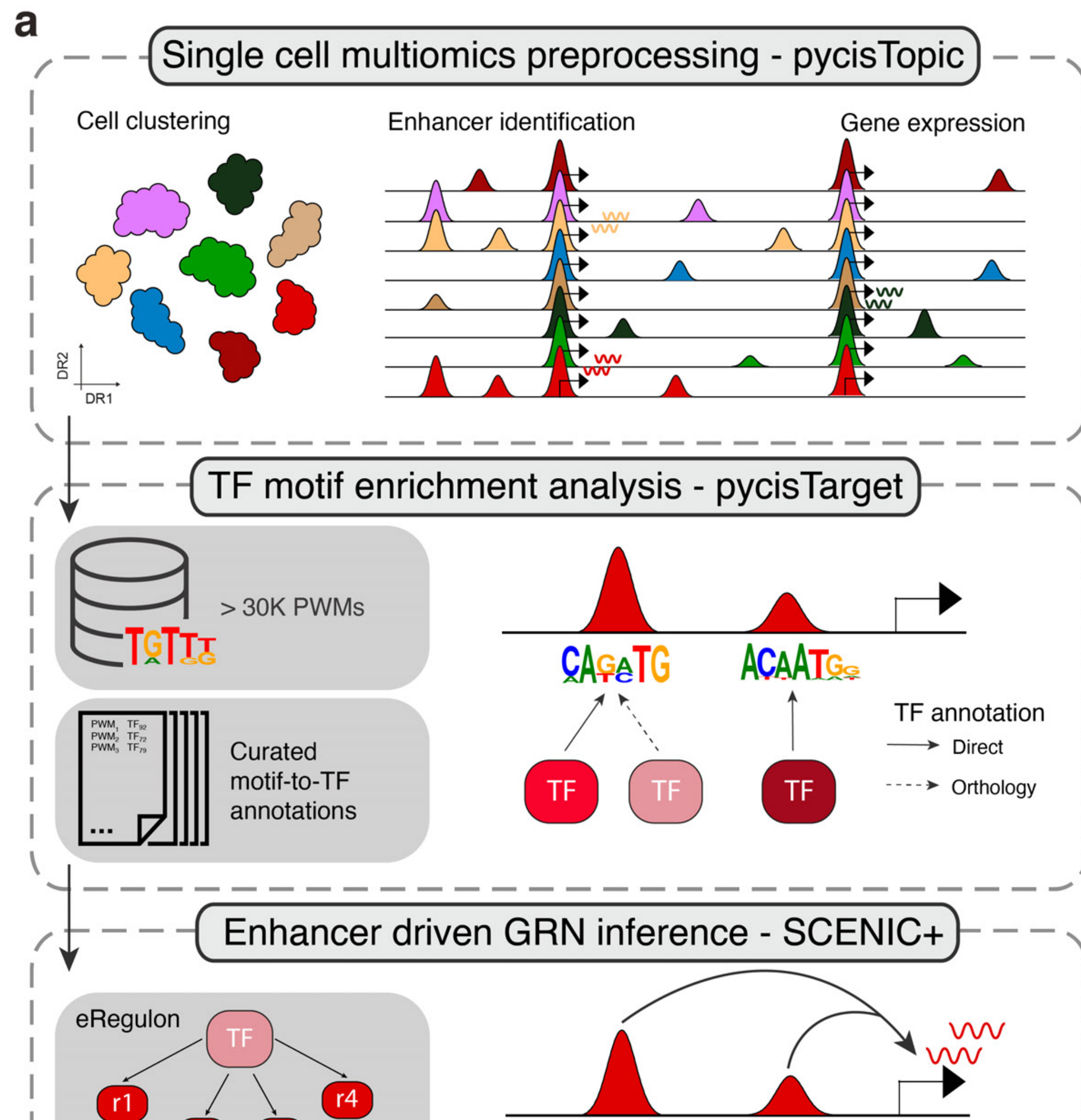
SCENIC



Reverse engineering of gene regulatory nets

SCENIC

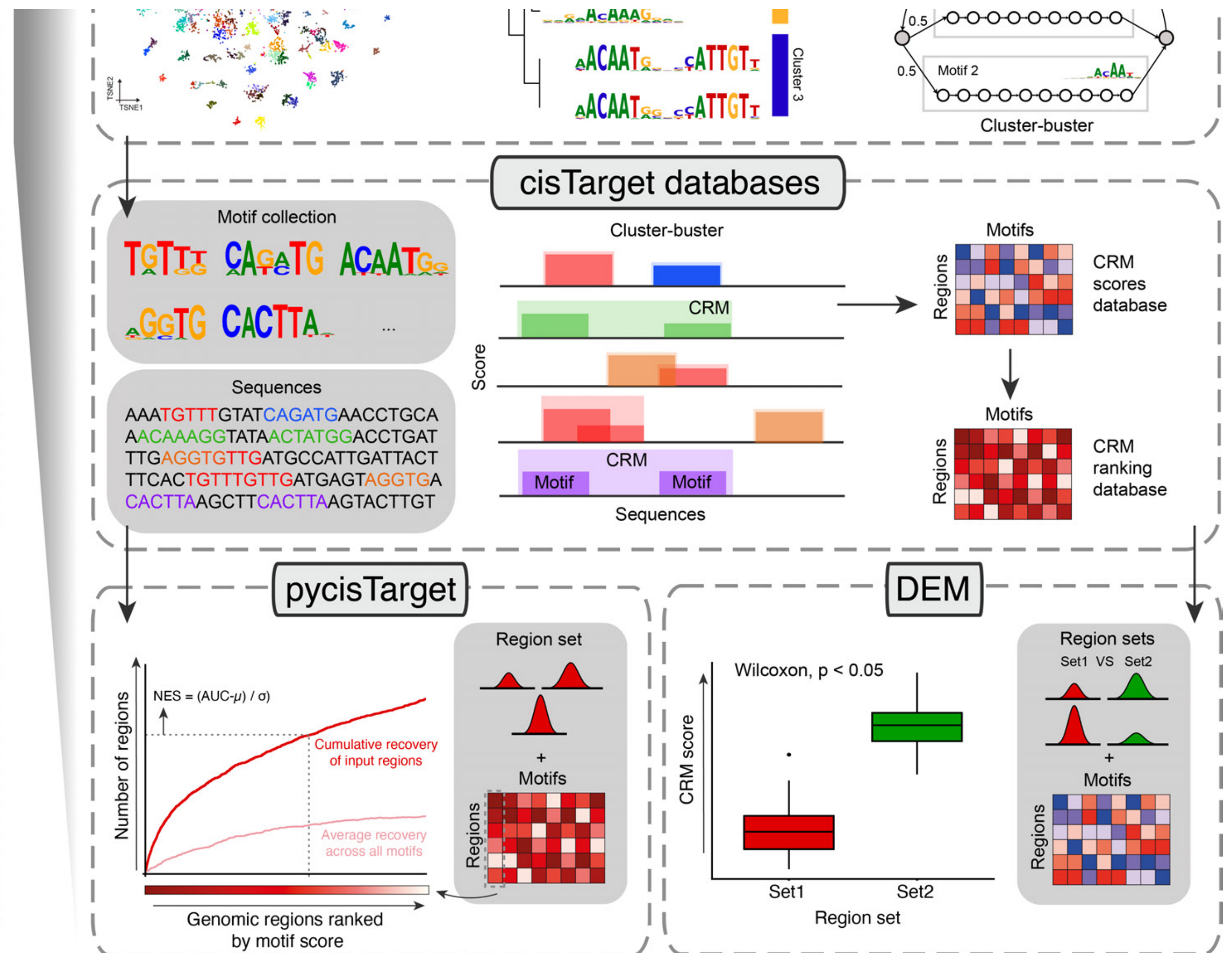
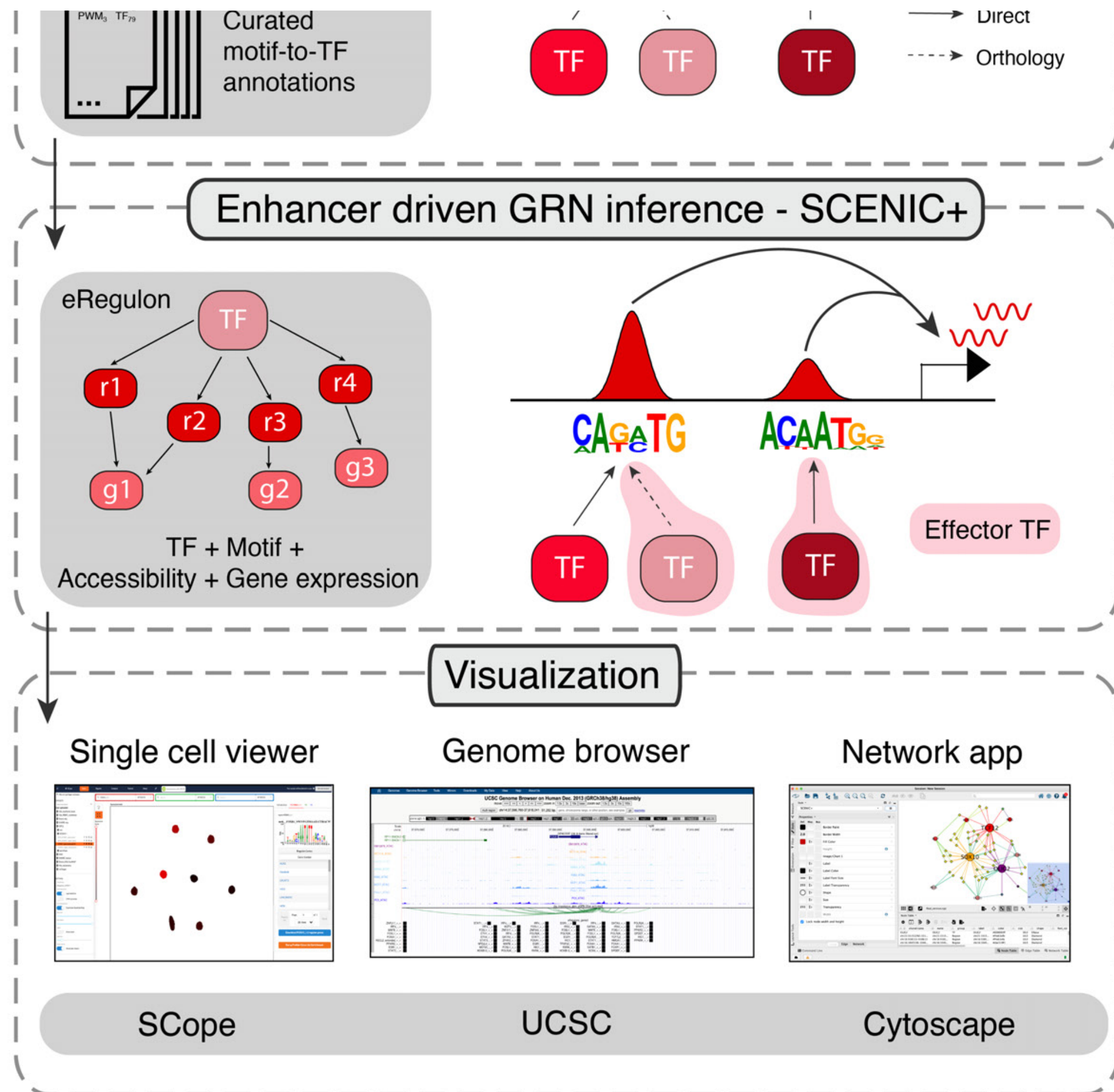
SCENIC+ Bravo González-Blas et al., preprint bioRxiv, 2021



Reverse engineering of gene regulatory nets

SCENIC

SCENIC+ Bravo González-Blas et al., preprint bioRxiv, 2021



Functional genomics

Zscan regulation in Embryonic Stem Cells

OPEN ACCESS Freely available online



Identification of a Novel Gene Signature of ES Cells Self-Renewal Fluctuation through System-Wide Analysis

**Luigi Cerulo^{1,2,3}, Daniela Tagliaferri^{1,2,3}, Pina Marotta¹, Pietro Zoppoli¹, Filomena Russo¹,
Claudia Mazio¹, Mario DeFelice^{1,3}, Michele Ceccarelli^{1,2*}, Geppino Falco^{1,2*}**

¹ Department of Stem Cell and Development, Istituto di Ricerche Genetiche Gaetano Salvatore Biogem scrl, Ariano Irpino, Italy, ² Department of Science, Università degli Studi del Sannio, Benevento, Italy, ³ Department of Medicina Molecolare e Biotecnologie mediche, Università di Napoli Federico II, Naples, Italy

Functional genomics

Zscan regulation in Embryonic Stem Cells

OPEN ACCESS Freely available online



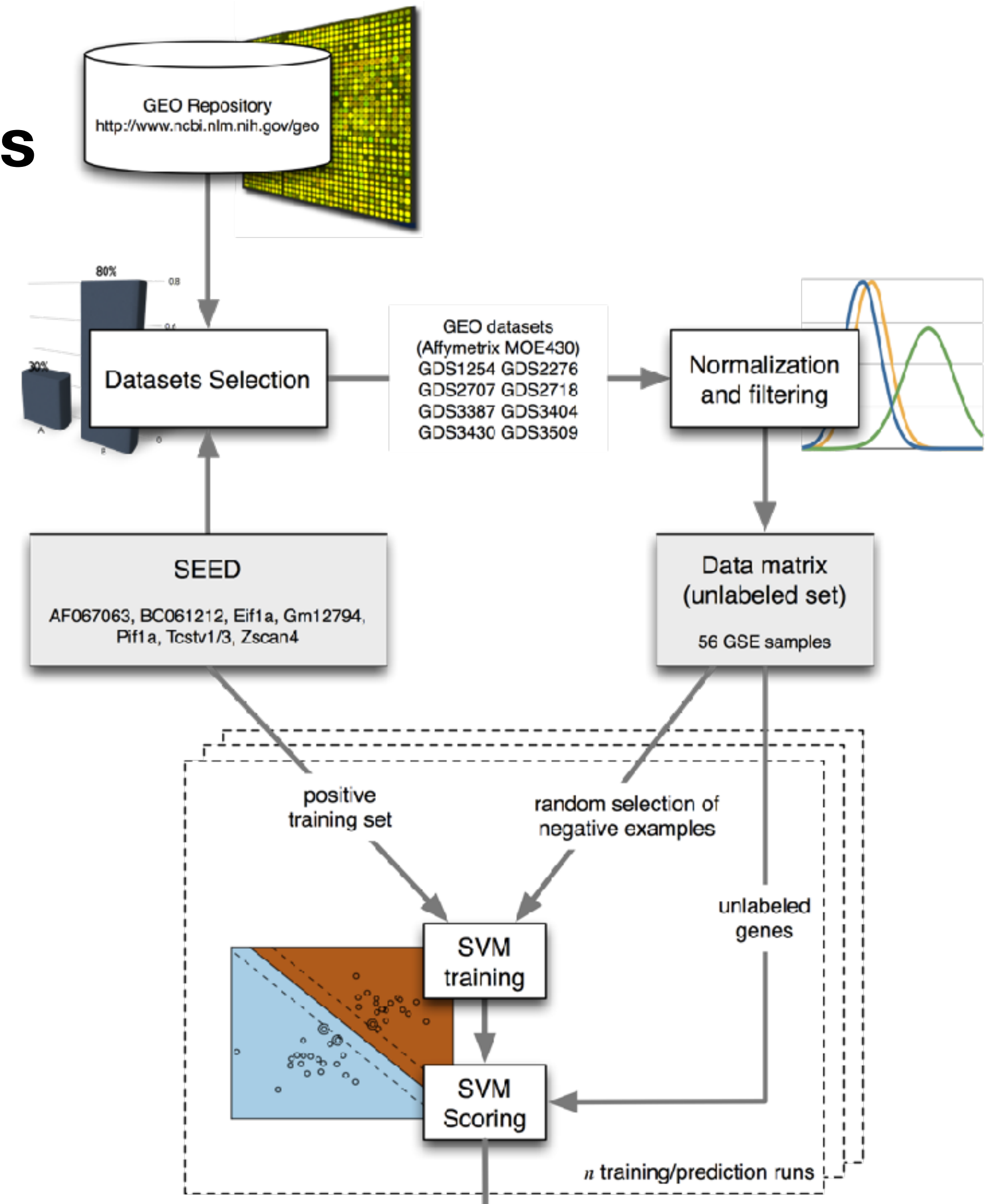
Identification of a Novel Gene Signature of ES Cells Self-Renewal Fluctuation through System-Wide Analysis

Luigi Cerulo^{1,2,3}, Daniela Tagliaferri^{1,2,3}, Pina Marotta¹, Pietro Zoppoli^{1,2}, Filomena Russo¹,
Claudia Mazio¹, Mario DeFelice^{1,3}, Michele Ceccarelli^{1,2*}, Geppino Falco^{1,2*}

¹ Department of Stem Cell and Development, Istituto di Ricerche Genetiche Gaetano Salvatore Biogem scrl, Ariano Irpino, Italy, ² Department of Science, Università degli Studi del Sannio, Benevento, Italy, ³ Department of Medicina Molecolare e Biotecnologie mediche, Università di Napoli Federico II, Naples, Italy

Data preparation

MGS Hypothesis



Functional genomics

Zscan regulation in Embryonic Stem Cells

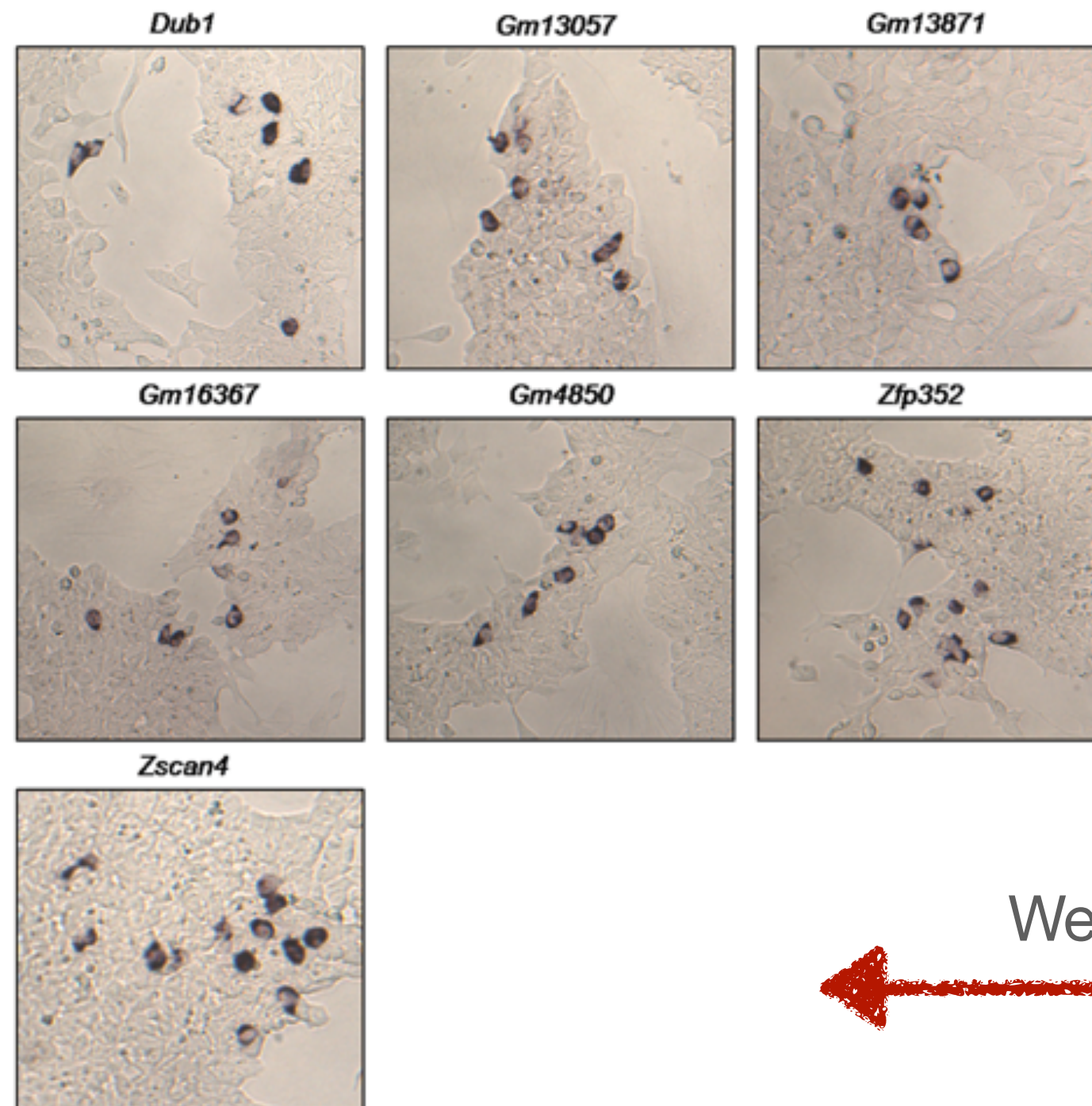
OPEN ACCESS Freely available online

PLOS ONE

Identification of a Novel Gene Signature of ES Cells Self-Renewal Fluctuation through System-Wide Analysis

Luigi Cerulo^{1,2,3}, Daniela Tagliaferri^{1,2,3}, Pina Marotta¹, Pietro Zoppoli^{1,2}, Filomena Russo¹,
Claudia Mazio¹, Mario DeFelice^{1,3}, Michele Ceccarelli^{1,2*}, Geppino Falco^{1,2*}

¹ Department of Stem Cell and Development, Istituto di Ricerche Genetiche Gaetano Salvatore Biogem scrl, Ariano Irpino, Italy, ² Department of Science, Università degli Studi del Sannio, Benevento, Italy, ³ Department of Medicina Molecolare e Biotecnologie mediche, Università di Napoli Federico II, Naples, Italy

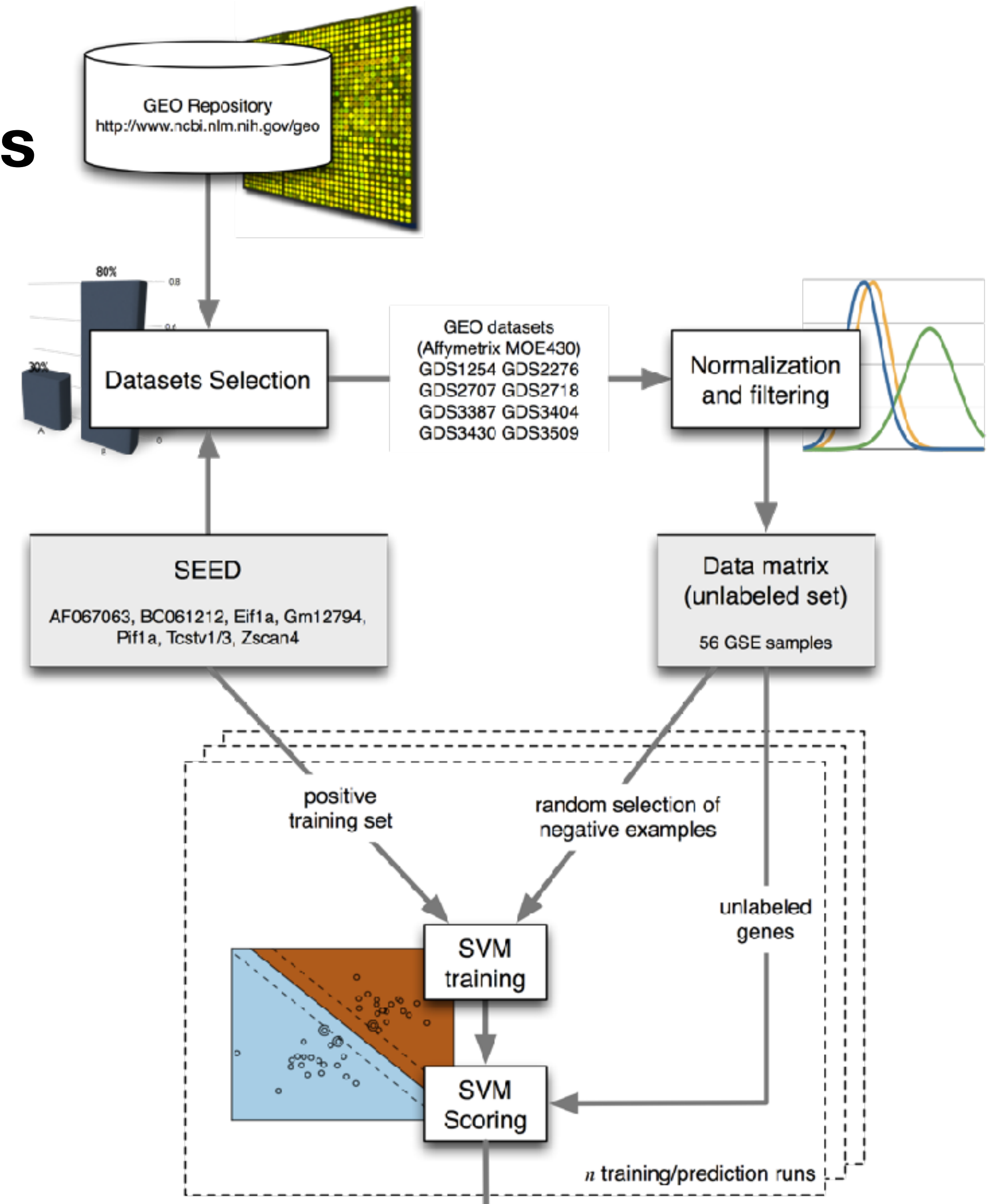


Wet Lab validation



Data preparation

MGS Hypothesis



Functional genomics

Zscan regulation in Embryonic Stem Cells

 **frontiers**
in Cell and Developmental Biology

Retinoic Acid Induces Embryonic Stem Cells (ESCs) Transition to 2 Cell-Like State Through a Coordinated Expression of *Dux* and *Duxbl1*

OPEN ACCESS

Daniela Tagliaferri^{1†}, Pellegrino Mazzone^{1†}, Teresa M. R. Noviello^{1,2†}, Martina Addeo^{1,3}, Tiziana Angrisano³, Luigi Del Vecchio^{4,5}, Feliciano Visconte⁵, Vitalba Ruggieri⁶, Sabino Russi⁶, Antonella Caivano⁶, Irene Cantone⁴, Mario De Felice^{4,7}, Michele Ceccarelli^{1,2}, Luigi Cerulo^{1,2} and Geppino Falco^{1,2,6,7*}

Edited by:
William L. Stanford,

Functional genomics

Zscan regulation in Embryonic Stem Cells

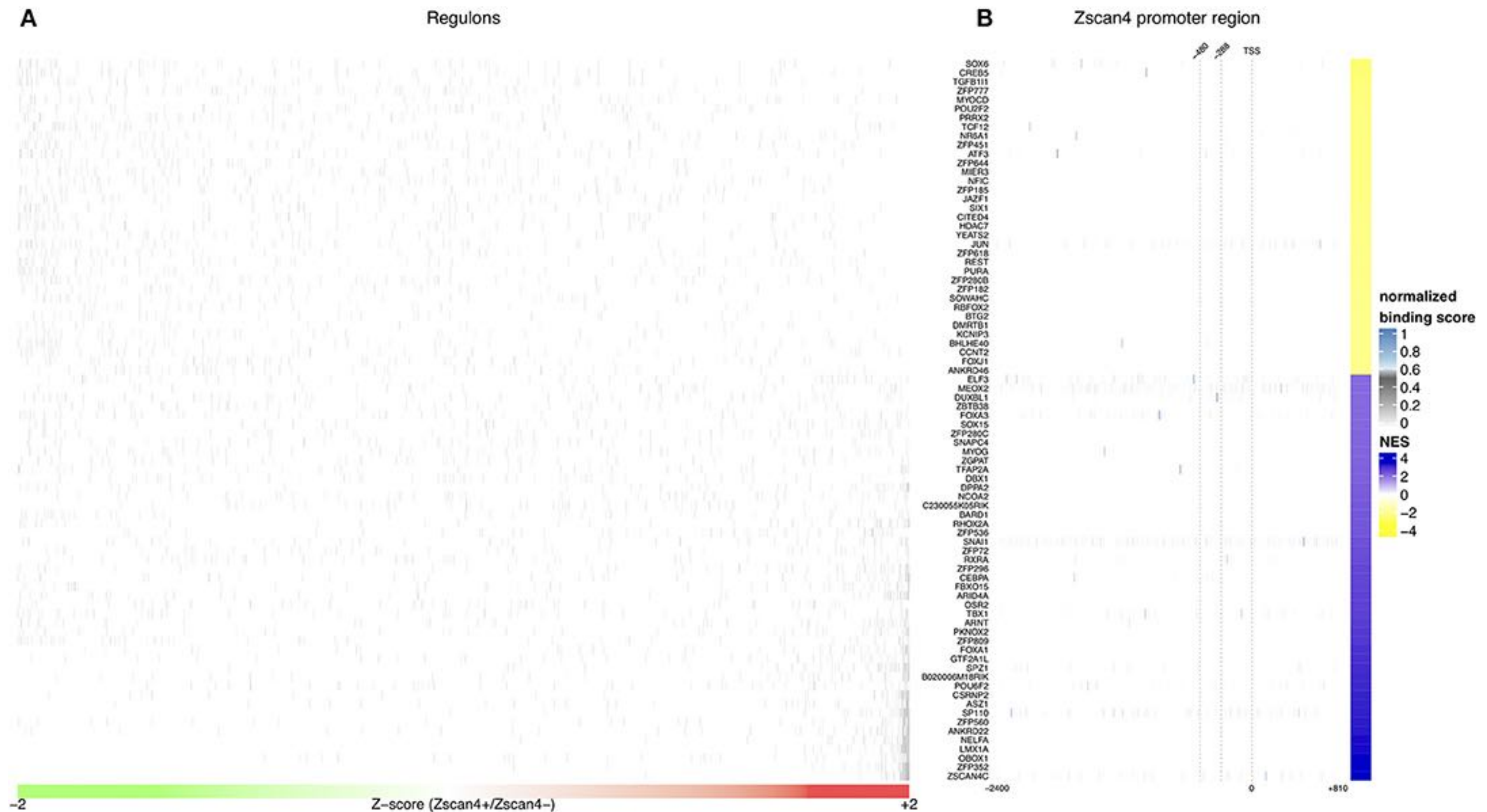
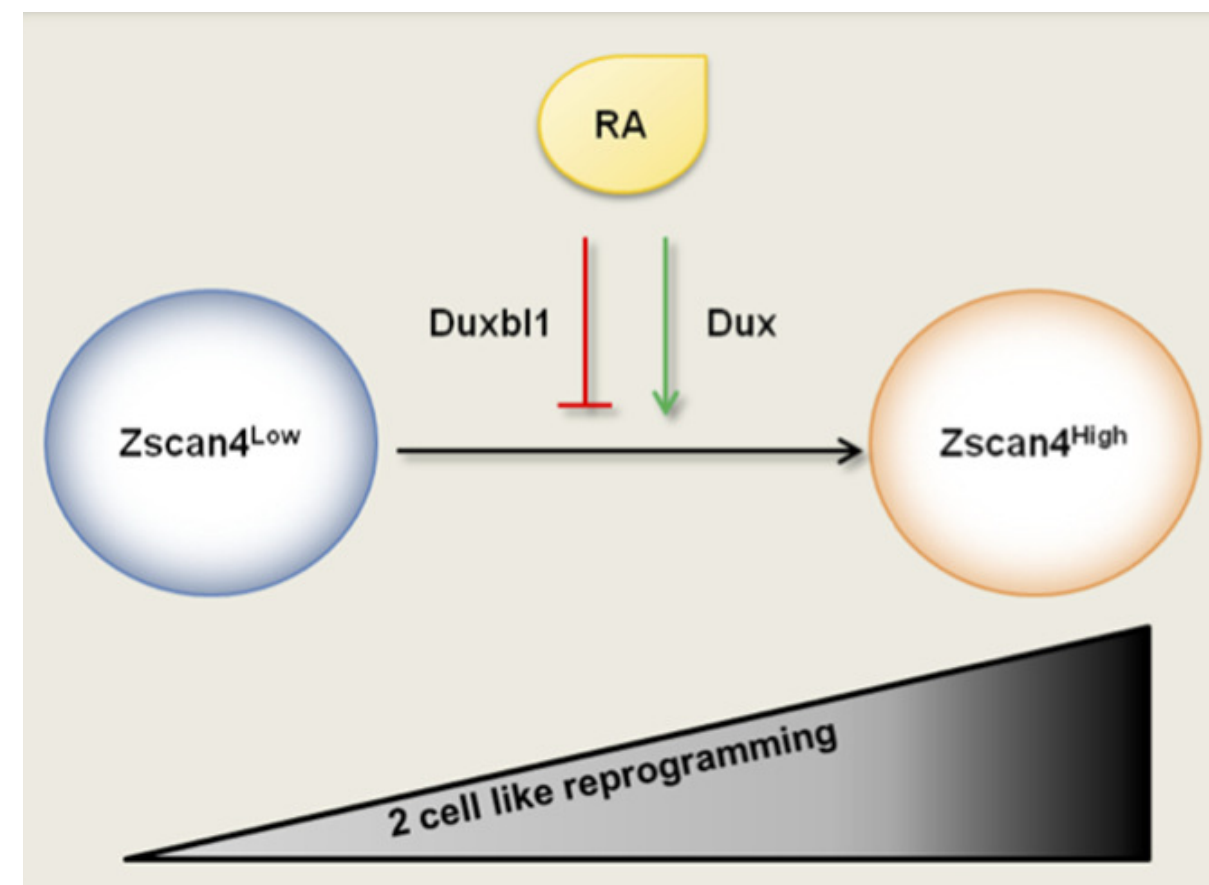
frontiers
in Cell and Developmental Biology

Retinoic Acid Induces Embryonic Stem Cells (ESCs) Transition to 2 Cell-Like State Through a Coordinated Expression of *Dux* and *Duxbl1*

OPEN ACCESS

Daniela Tagliaferri^{1†}, Pellegrino Mazzone^{1†}, Teresa M. R. Noviello^{1,2†}, Martina Addeo^{1,3}, Tiziana Angrisano³, Luigi Del Vecchio^{4,5}, Feliciano Visconte⁵, Vitalba Ruggieri⁶, Sabino Russi⁶, Antonella Caivano⁶, Irene Cantone⁴, Mario De Felice^{4,7}, Michele Ceccarelli^{1,2}, Luigi Cerulo^{1,2} and Geppino Falco^{1,2,6,7*}

Edited by:
William L. Stanford,



Applications

De novo Motif discovery

A motif is an approximate sequence pattern with no gaps that occurs repeatedly in a group of related sequences.

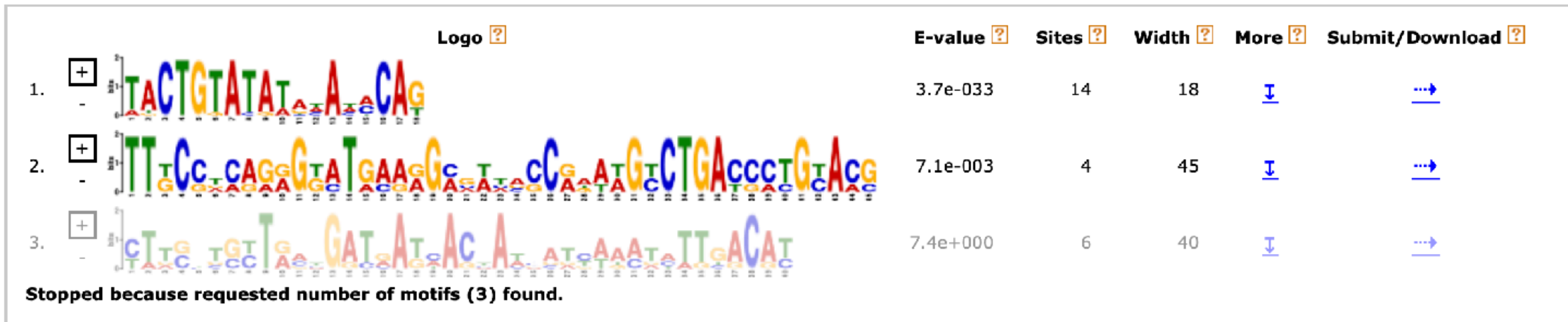
A motifs can be modelled as position-dependent letter-probability matrices that describe the probability of each possible letter at each position in the pattern.

The input is a group of sequences and the outputs are the most significant motifs occurring in all sequences.

Applications

MEME suite

DISCOVERED MOTIFS



MOTIF LOCATIONS



Applications

Motif scanning

Motif scanners scan sequences (e.g. promoters) to match known motifs represented as position-dependent letter-probability matrices that describe the probability of each possible letter at each position in the pattern.

The input is a group of sequences and a group of known motifs. The outputs are the most significant motifs found in all sequences.

Applications

FIMO motif scanner

DATABASE AND MOTIFS

DATABASE mm9_tss_500bp_sampled_1000.fna
Database contains 1000 sequences, 500000 residues

MOTIFS some_vertbrates.meme (DNA)

MOTIF	WIDTH	BEST POSSIBLE MATCH
MA0108.1	15	GTATAAAAGGCGGGG
UP00093_1	16	TCGACCCCGCCCCTAT
UP00020_1	16	ACGATGACGTCATCGA
UP00002_2	15	CAAAGGCGTGGCCAG
MA0060.1	16	CTCAGCCAATCAGCGC

Random model letter frequencies (--nrdb--):
A 0.275 C 0.225 G 0.225 T 0.275

Motif ID	Alt ID	Sequence Name	Strand	Start	End	p-value	q-value	Matched Sequence
MA0060.1	NFYA	chr2	-	60221163	60221178	3.36e-09	0.00195	CTCGGCCAATCAGAGC
MA0060.1	NFYA	chr3	-	54858838	54858853	4.77e-09	0.00195	ATCAGCCAATCAGCGG
UP00093_1	Klf7_primary	chr2	+	172266018	172266033	1.06e-08	0.00496	GCGACCCCGCCCCTTT
UP00093_1	Klf7_primary	chr11	+	65981465	65981480	1.31e-08	0.00496	TTGACCCCGCCCCTCA
UP00020_1	Atf1_primary	chr3	+	65470135	65470150	2.63e-08	0.0222	GCTGTGACGTCACCGC
MA0060.1	NFYA	chrX	+	51925276	51925291	3.55e-08	0.00584	TTCAGCCAATCAGCGC
MA0060.1	NFYA	chrX	+	52016291	52016306	3.55e-08	0.00584	TTCAGCCAATCAGCGC
UP00093_1	Klf7_primary	chr1	+	93302476	93302491	3.87e-08	0.00898	TCGGCCCGCCCCTCC

Applications

Exercise 1 - motif scanning

We would like to test whether Mitochondrial genes in Mouse have TBP binding sequences in their promoter regions (500 bp flank-coding upstream region)

1. With Jaspar (<https://jaspar.uio.no>) search for TBP PWM matrix and download it in MEME format
2. With Biomart (<https://www.ensembl.org/biomart/martview>) download promoter sequences of all genes located in Mouse Mitochondrial chromosome
3. With FIMO (<https://meme-suite.org/meme/tools/fimo>) search TBP binding regions using previous downloaded data

Applications

Exercise 1 - motif scanning

1. With Jaspar (<https://jaspar.uio.no>) search for TBP PWM matrix and download it in MEME format

The screenshot shows the JASPAR 2022 website interface. The top navigation bar includes the JASPAR 2022 logo, a hamburger menu, a shopping cart icon with a '0' badge, and a link to the JASPAR Blog. A left sidebar contains navigation links: Home, About, Search, Browse JASPAR CORE, Unvalidated Profiles, Browse Collections, Tools, and RESTful API. The main content area features a search bar with 'TBP' entered. Below the search bar, there are examples: 'SPI1, P17676, CHIP-seq, Homo sapiens'. A tooltip points to the search bar with the text: 'You can search by TF name or ID, species, taxon, UniProt ID or any other keyword.' To the right of the search bar is a 'Search' button and a link to 'Advanced Options'. Below the search bar, there is a section titled 'Browse JASPAR CORE for 6 different taxonomic groups' with four image-based buttons: 'Fungi' (with a mushroom image), 'Insecta' (with a fly image), and two others (with a red and green image). On the right side of the page, there is a large banner for '2022 The high-quality transcription factor binding profile database'.

Applications

Exercise 1 - motif scanning

1. With Jaspar (<https://jaspar.uio.no>) search for TBP PWM matrix and download it in MEME format

The screenshot shows the JASPAR 2022 website interface. The top navigation bar includes the JASPAR 2022 logo, a menu icon, a shopping cart icon with '0' items, and a link to the JASPAR Blog. Below the navigation bar, there are search examples: 'SPI1, P17676, ChIP-seq, Homo sapiens' and a link to 'Advanced Options'. The main content area displays '9 profile(s) found' and a 'Display 10 profiles' dropdown. A search filter box is present. The results are shown in a table with columns for selection, ID, Name, Species, Class, Family, and Logo. The first two rows are for TBP profiles (MA0108.1 and MA0108.2), and the third row is for SPT15 (MA0386.1). A right sidebar contains an 'Analyze selected profiles' section with instructions to select matrix profiles and an 'Add to cart' button. A blue notification box at the bottom right states 'You have 0 profile(s) in your cart. You can add profiles'.

JASPAR²⁰²²

Examples: SPI1, P17676, ChIP-seq, Homo sapiens

Advanced Options

9 profile(s) found

Display 10 profiles

Filter:

	ID	Name	Species	Class	Family	Logo
<input type="checkbox"/>	MA0108.1	TBP		TATA-binding proteins	TBP-related factors	
<input checked="" type="checkbox"/>	MA0108.2	TBP		TATA-binding proteins	TBP-related factors	
<input type="checkbox"/>	MA0386.1	SPT15	Saccharomyces	TATA-	TBP-	

Analyze selected profiles

Please select matrix profiles on the left side to add to your cart or perform the following analysis.

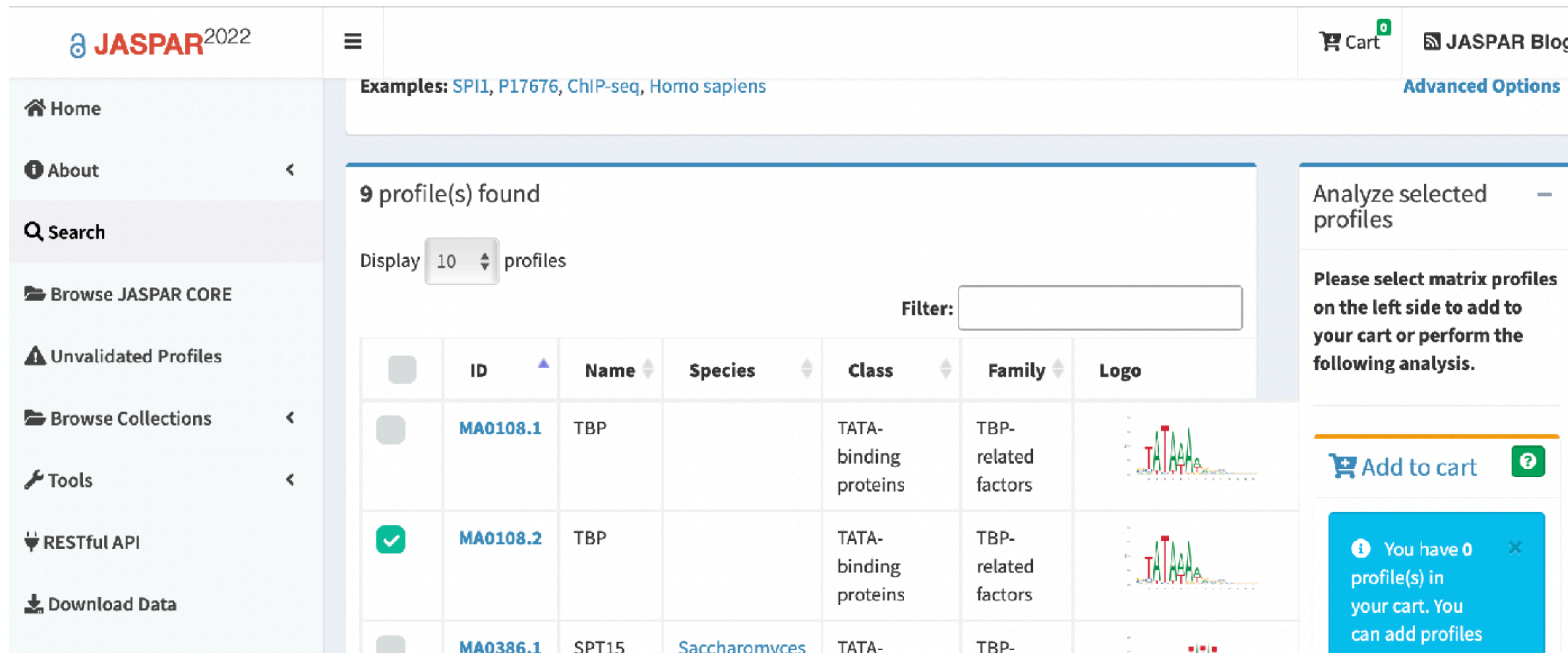
Add to cart

You have 0 profile(s) in your cart. You can add profiles

Applications

Exercise 1 - motif scanning

1. With Jaspar (<https://jaspar.uio.no>) search for TBP PWM matrix and download it in MEME format



The screenshot shows the JASPAR 2022 website interface. The top navigation bar includes the JASPAR 2022 logo, a menu icon, a shopping cart icon with '0' items, and a link to the JASPAR Blog. Below the navigation bar, there are search examples: 'SPI1, P17676, ChIP-seq, Homo sapiens' and a link to 'Advanced Options'. The main content area displays '9 profile(s) found' and a 'Display 10 profiles' dropdown. A search filter box is present. The results are shown in a table with columns for selection, ID, Name, Species, Class, Family, and Logo. The first two rows are for TBP profiles (MA0108.1 and MA0108.2), and the third row is for SPT15 (MA0386.1). A right sidebar contains an 'Analyze selected profiles' section with instructions to select matrix profiles and an 'Add to cart' button. A blue notification box at the bottom right states 'You have 0 profile(s) in your cart. You can add profiles'.

JASPAR²⁰²²

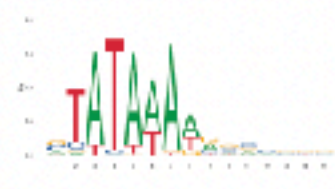
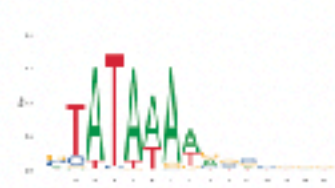

Examples: SPI1, P17676, ChIP-seq, Homo sapiens

Advanced Options

9 profile(s) found

Display 10 profiles

Filter:

	ID	Name	Species	Class	Family	Logo
<input type="checkbox"/>	MA0108.1	TBP		TATA-binding proteins	TBP-related factors	
<input checked="" type="checkbox"/>	MA0108.2	TBP		TATA-binding proteins	TBP-related factors	
<input type="checkbox"/>	MA0386.1	SPT15	Saccharomyces	TATA-	TBP-	

Analyze selected profiles

Please select matrix profiles on the left side to add to your cart or perform the following analysis.

Add to cart

You have 0 profile(s) in your cart. You can add profiles

Applications

Exercise 1 - motif scanning

1. With Jaspar (<https://jaspar.uio.no>) search for TBP PWM matrix and download it in MEME format

The screenshot displays the JASPAR 2022 website interface. The left sidebar contains navigation links: Home, About, Search, Browse JASPAR CORE, Unvalidated Profiles, Browse Collections, Tools, RESTful API, Download Data, Matrix Clusters, Genome Tracks, and Enrichment Analysis. The main content area is titled "Detailed information of matrix profile MA0108.2".

Profile summary (Add)

Name:	TBP
Matrix ID:	MA0108.2
Class:	TATA-binding proteins
Family:	TBP-related factors
Collection:	CORE
Taxon:	Vertebrates
Species:	
Data Type:	
Validation:	2329577
Uniprot ID:	P20226
Source:	

Sequence logo (Download SVG)

Frequency matrix (JASPAR, TRANSFAC, MEME, RAW PFM, Reverse comp.)

A [61	16	352	3	354	268	360	222	155	56	83	82	82	68	77]
C [145	46	0	10	0	0	3	2	44	135	147	127	118	107	101]

Transcription Factor promoter analysis

Exercise 1 - motif scanning

2. With Biomart (<https://www.ensembl.org/biomart/martview>) download promoter sequences of all genes located in Mouse Mitochondrial chromosome

The screenshot shows the Ensembl BioMart interface. The top navigation bar includes the Ensembl logo and links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. Below the navigation bar, there are tabs for New, Count, Results, URL, XML, Perl, and Help. The main content area is divided into two columns. The left column contains a sidebar with sections for Dataset, Filters, and Attributes. The right column contains two dropdown menus for selecting the dataset and filters. The first dropdown menu is set to 'Ensembl Genes 107' and the second dropdown menu is set to 'Mouse genes (GRCm39)'. The 'Attributes' section in the sidebar lists the following attributes: Gene stable ID, Gene stable ID version, Transcript stable ID, and Transcript stable ID version.

Transcription Factor promoter analysis

Exercise 1 - motif scanning

2. With Biomart (<https://www.ensembl.org/biomart/martview>) download promoter sequences of all genes located in Mouse Mitochondrial chromosome

The screenshot shows the Ensembl Biomart interface. The top navigation bar includes the Ensembl logo and links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. Below this, there are buttons for 'New', 'Count', and 'Results', along with options for 'URL', 'XML', 'Perl', and 'Help'. The main content area is titled 'Please restrict your query using criteria below' and includes a warning '(If filter values are truncated in any lists, hover over the list item to see the full list)'. The 'REGION' filter is expanded, showing a checked box for 'Chromosome/scaffold' and a dropdown menu with the following options: JH584298.1, JH584299.1, JH584303.1, JH584304.1, MT (highlighted), X, and Y. Below the 'Chromosome/scaffold' filter, there is an unchecked box for 'Coordinates' and a 'Start' label with a text input field containing the number '1'.

Transcription Factor promoter analysis

Exercise 1 - motif scanning

2. With Biomart (<https://www.ensembl.org/biomart/martview>) download promoter sequences of all genes located in Mouse Mitochondrial chromosome

The screenshot shows the Ensembl Biomart interface. The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located in the top right corner. Below the navigation bar, there are buttons for 'New', 'Count', and 'Results'. The main content area is divided into a left sidebar and a main configuration panel.

Dataset: Mouse genes (GRCm39)

Filters: Chromosome/scaffold: MT

Attributes: Flank-coding region (Gene), Gene name, Upstream flank [500]

Dataset: [None Selected]

Configuration Panel:

Please select columns to be included in the output and hit 'Results' when ready

Missing non coding genes in your mart query output, please check the following [FAQ](#)

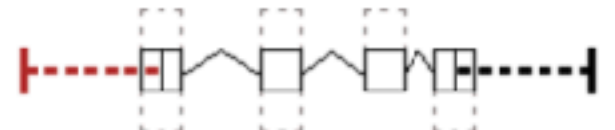
Features Variant (Germline)

Structures Sequences

Homologues (Max select 6 orthologues)

SEQUENCES:

Sequences (max 1)



Unspliced (Transcript) 5' UTR

Unspliced (Gene) 3' UTR

Flank (Transcript) Exon sequences

Flank (Gene) cDNA sequences

Flank-coding region (Transcript) Coding sequence

Flank-coding region (Gene) Peptide

Upstream flank

Upstream flank

Transcription Factor promoter analysis

Exercise 1 - motif scanning

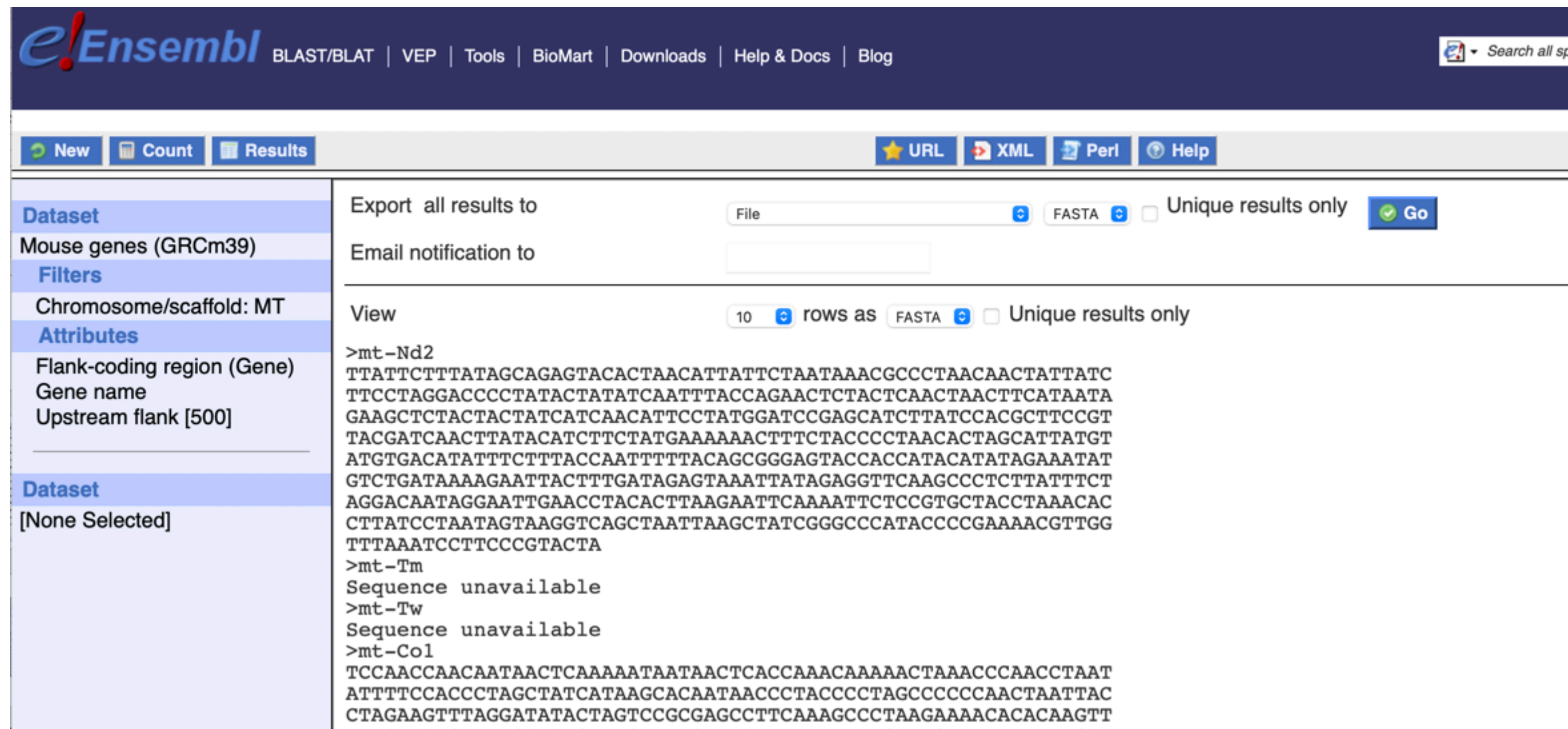
2. With Biomart (<https://www.ensembl.org/biomart/martview>) download promoter sequences of all genes located in Mouse Mitochondrial chromosome

The screenshot shows the Ensembl Biomart interface. The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located in the top right corner. Below the navigation bar, there are buttons for 'New', 'Count', and 'Results', along with options for 'URL', 'XML', 'Perl', and 'Help'. The main content area is divided into two columns. The left column shows the 'Dataset' as 'Mouse genes (GRCm39)' and the 'Attributes' as 'Flank-coding region (Gene)', 'Gene name', and 'Upstream flank [500]'. The right column shows the 'HEADER INFORMATION' section with a list of attributes. Under 'Gene Information', 'Gene name' is selected with a checked checkbox. Under 'Transcript Information', 'Transcript start (bp)' is selected with a checked checkbox. Other attributes listed include 'Gene stable ID', 'Gene stable ID version', 'Gene description', 'Source of gene name', 'Chromosome/scaffold name', 'Gene start (bp)', 'Gene end (bp)', 'Gene type', 'Version (gene)', 'UniParc ID', 'UniProtKB/Swiss-Prot ID', 'UniProtKB/TrEMBL ID', 'Protein stable ID version', 'Transcript type', 'Version (transcript)', 'Version (protein)', 'Strand', 'Transcript end (bp)', 'Transcription start site (TSS)', and 'Transcript length (including UTRs and CDS)'. The 'Exon Information' section is partially visible at the bottom.

Transcription Factor promoter analysis

Exercise 1 - motif scanning

2. With Biomart (<https://www.ensembl.org/biomart/martview>) download promoter sequences of all genes located in Mouse Mitochondrial chromosome



The screenshot shows the Ensembl Biomart interface. The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located in the top right corner. Below the navigation bar, there are buttons for 'New', 'Count', and 'Results'. The main content area is divided into two columns. The left column contains a sidebar with 'Dataset' (Mouse genes (GRCm39)), 'Filters' (Chromosome/scaffold: MT), and 'Attributes' (Flank-coding region (Gene), Gene name, Upstream flank [500]). The right column contains export and view options. The 'Export all results to' section has a dropdown menu set to 'File', a 'FASTA' button, a 'Unique results only' checkbox, and a 'Go' button. The 'Email notification to' field is empty. The 'View' section has a dropdown menu set to '10', a 'rows as' dropdown menu set to 'FASTA', a 'Unique results only' checkbox, and a 'Go' button. The main content area displays the following text:

```
>mt-Nd2
TTATTCTTTATAGCAGAGTACACTAACATTATTCTAATAAACGCCCTAACAACTATTATC
TTCCTAGGACCCCTATACTATATCAATTTACCAGAACTCTACTCAACTAACTTCATAATA
GAAGCTCTACTACTATCATCAACATTCCTATGGATCCGAGCATCTTATCCACGCTTCCGT
TACGATCAACTTATACATCTTCTATGAAAAAATTTCTACCCCTAACACTAGCATTATGT
ATGTGACATATTTCTTTACCAATTTTTACAGCGGGAGTACCACCATACATATAGAAATAT
GTCTGATAAAAAGAATTACTTTGATAGAGTAAATTATAGAGGTTCAAGCCCTCTTATTTCT
AGGACAATAGGAATTGAACCTACACTTAAGAATTCAAAATTTCTCCGTGCTACCTAAACAC
CTTATCCTAATAGTAAGGTCAGCTAATTAAGCTATCGGGCCCATACCCCGAAAACGTTGG
TTTAAATCCTTCCCGTACTA
>mt-Tm
Sequence unavailable
>mt-Tw
Sequence unavailable
>mt-Co1
TCCAACCAACAATAACTCAAAAATAATAACTCACCAAAACAAAACTAAACCCAACCTAAT
ATTTTCCACCCTAGCTATCATAAGCACATAACCCTACCCCTAGCCCCCAACTAATTAC
CTAGAAGTTTAGGATATACTAGTCCGCGAGCCTTCAAAGCCCTAAGAAAACACACAAGTT
```

Transcription Factor promoter analysis

Exercise 1 - motif scanning

3. With FIMO (<https://meme-suite.org/meme/tools/fimo>) search TBP binding regions using previous downloaded data

MEME Suite 5.5.0

- ▶ Motif Discovery
- ▶ Motif Enrichment
- ▶ Motif Scanning
- ▶ Motif Comparison
- ▶ Gene Regulation
- ▶ Utilities
- ▶ Manual
- ▶ Guides & Tutorials
- ▶ Sample Outputs
- ▶ File Format Reference
- ▶ Databases
- ▶ Download & Install
- ▶ Help

FIMO

Find Individual Motif Occurrences

Version 5.5.0

FIMO scans a set of sequences for **individual matches** to each of the motifs you provide (sample output for motifs and sequences). See this [Manual](#) or this [Tutorial](#) for more information.

Data Submission Form

Scan a set of sequences for motifs.

Input the motifs

Enter motifs you wish to scan with.

Upload motifs MA0108.2.meme ?

Input the sequences

Specify sequences or select the [database](#) you want to scan for matches to motifs.

Upload sequences mart_export-2.txt ?

Input job details

(Optional) Enter your email address. ?

(Optional) Enter a job description. ?

Transcription Factor promoter analysis

Exercise 1 - motif scanning

3. With FIMO (<https://meme-suite.org/meme/tools/fimo>) search TBP binding regions using previous downloaded data



FIMO
Find Individual Motif Occurrences

MEME Suite 5.5.0

- ▶ Motif Discovery
- ▶ Motif Enrichment
- ▶ Motif Scanning
- ▶ Motif Comparison
- ▶ Gene Regulation
- ▶ Utilities
- ▶ Manual
- ▶ Guides & Tutorials
- ▶ Sample Outputs
- ▶ File Format Reference
- ▶ Databases
- ▶ Download & Install
- ▶ Help

Your FIMO job is complete. The results should be displayed below.

Job Details ...

Results

- [FIMO HTML output](#)
- [FIMO TSV output](#)
- [FIMO XML output](#)
- [FIMO CISML output](#)
- [FIMO GFF output](#)
- [Input Motifs](#)
- [Uploaded Sequences](#)

Status Messages

- Parsing arguments
- Arguments ok
- Starting fimo

Transcription Factor promoter analysis

Exercise 1 - motif scanning

3. With FIMO (<https://meme-suite.org/meme/tools/fimo>) search TBP binding regions using previous downloaded data

DATABASE AND MOTIFS

DATABASE mart_export-2.txt
Database contains 13 sequences, 6500 residues

MOTIFS MA0108.2.meme (DNA)

MOTIF	WIDTH	BEST POSSIBLE MATCH
MA0108.2	15	GTATAAAAGGCGGGG

Random model letter frequencies (--nrdb--):
A 0.275 C 0.225 G 0.225 T 0.275

SECTION I: HIGH-SCORING MOTIF OCCURENCES

- There were 1 motif occurrences with a p-value less than 0.0001. The full set of motif occurrences can be seen in the TSV (tab-delimited values) output file [fimo.tsv](#), the GFF3 file [fimo.gff](#) which may be suitable for uploading to the [UCSC Genome Table Browser](#) (assuming the FASTA input sequences included genomic coordinates in UCSC or Galaxy format), or the XML file [fimo.xml](#).
- The p-value of a motif occurrence is defined as the probability of a random sequence of the same length as the motif matching that position of the sequence with as good or better a score.
- The score for the match of a position in a sequence to a motif is computed by summing the appropriate entries from each column of the position-dependent scoring matrix that represents the motif.
- The q-value of a motif occurrence is defined as the false discovery rate if the occurrence is accepted as significant.
- The table is sorted by increasing p-value.

Motif ID	Alt ID	Sequence Name	Strand	Start	End	p-value	q-value	Matched Sequence
MA0108.2	TBP	mt-Nd6	+	39	53	4.73e-05	0.485	CTATAAATGCTGTGG

Transcription Factor promoter analysis

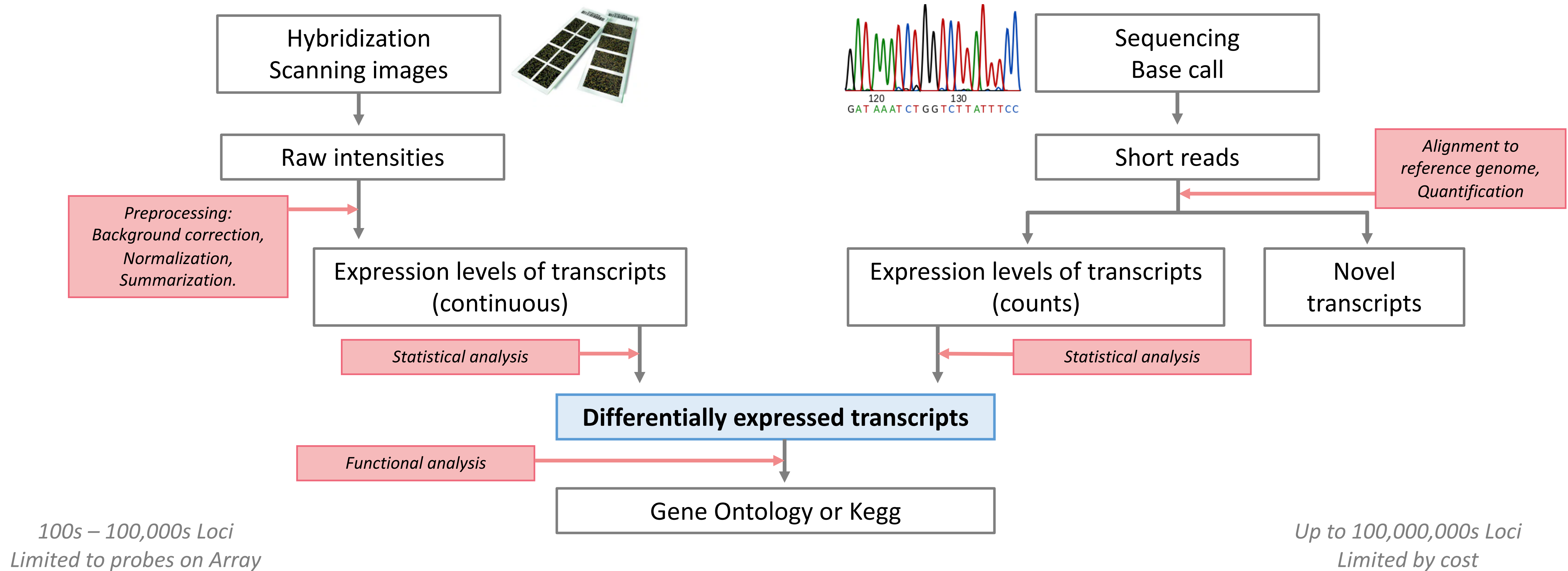
Exercise 2 - motif scanning

Some studies revealed that DAL1, DAL2, DAL4, DAL5, DAL7, DAL80, and GAP1 are strongly co-expressed in *S.Cerevisiae* in more than 200 different conditions (Allocco et al., 2004). It could be the case that those genes are regulated by a common Transcription Factor.

1. With Biomart download the upstream flank-coded sequences (1500 bp upstream) of those genes
2. With FIMO perform a promote analysis scanning all available motifs in *S.Cerevisiae* downloaded from Jaspar
3. Analyse results in Excel making convincing conclusions

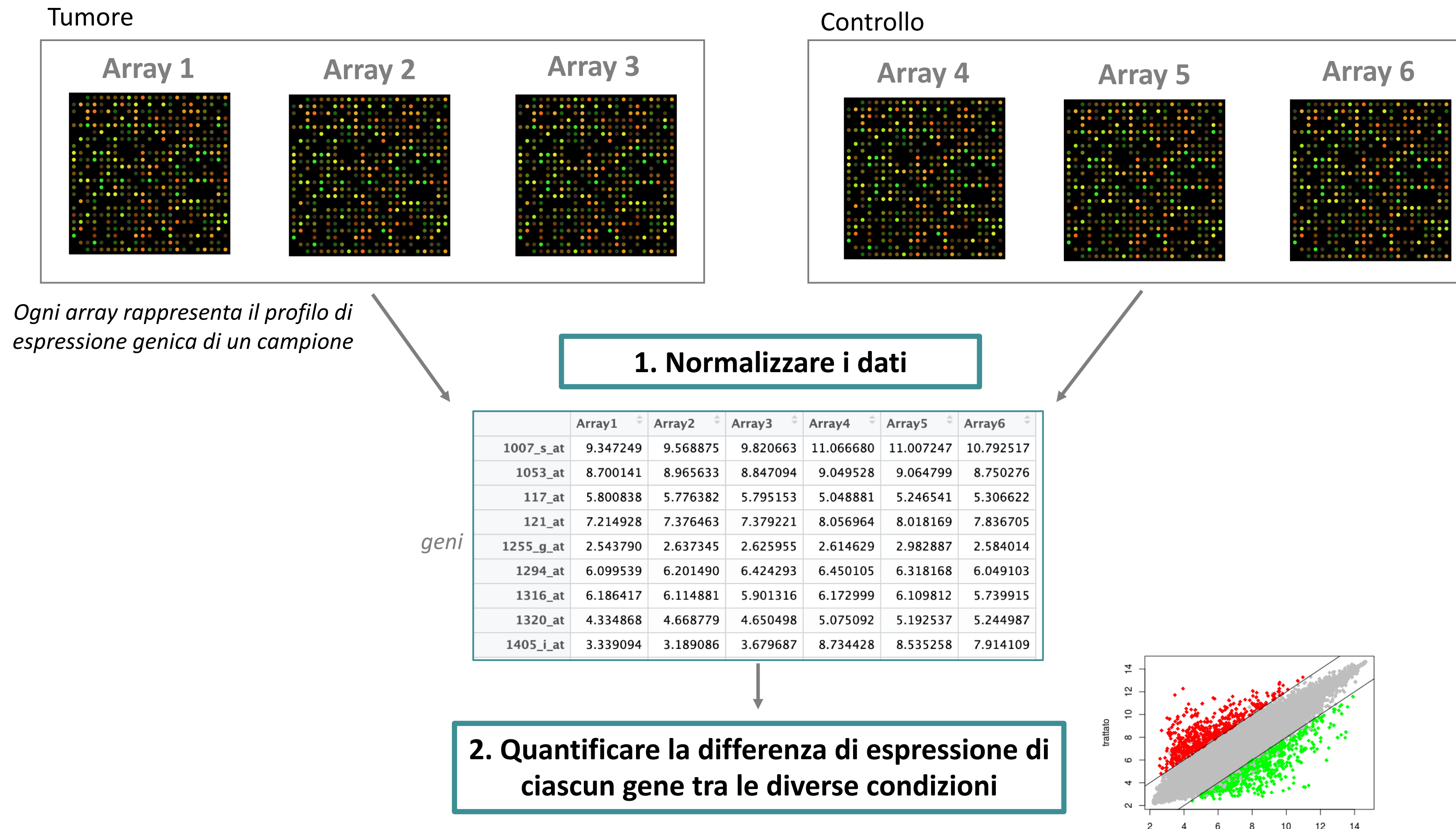
Applications

Differential Expression Analysis



Applications

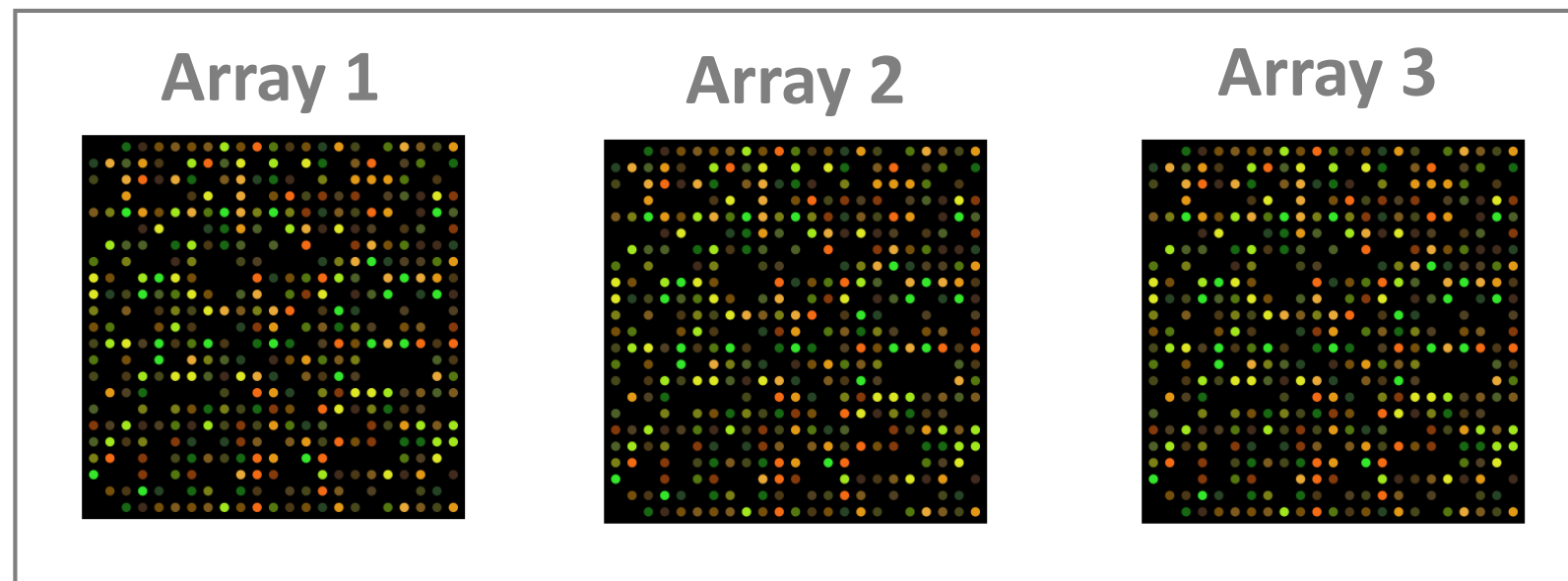
Differential Expression Analysis



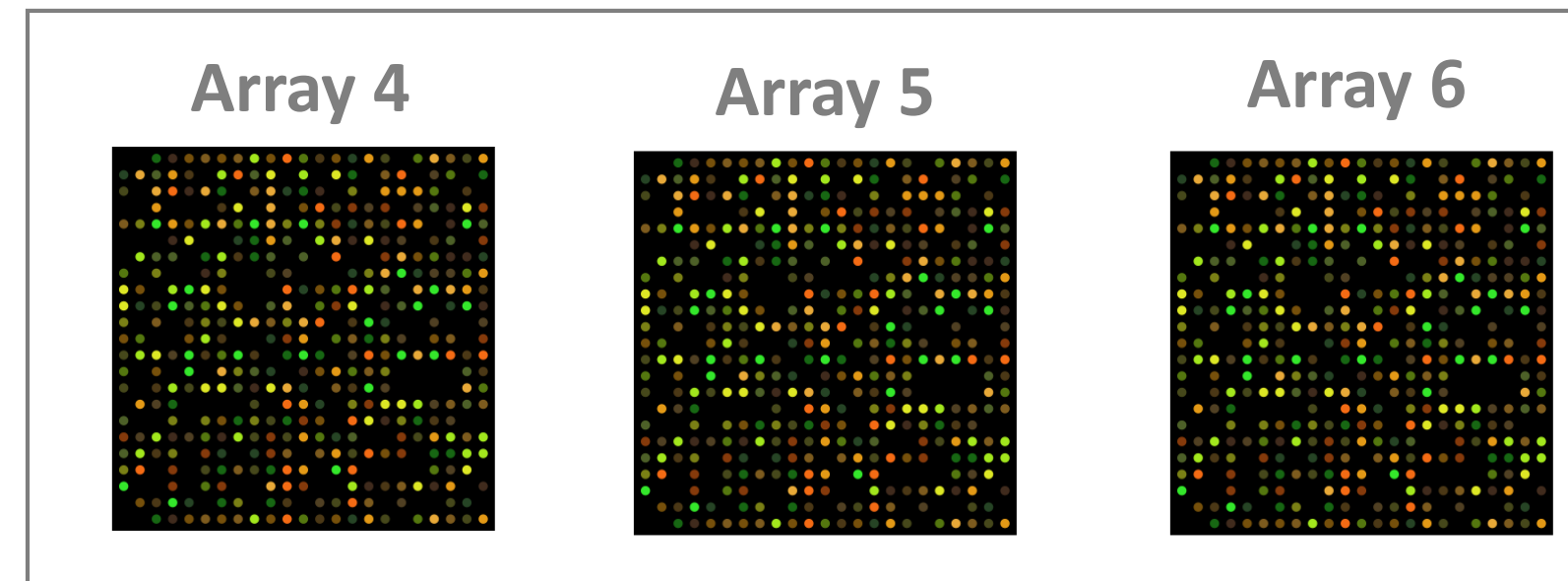
Applications

Differential Expression Analysis

Tumore



Controllo



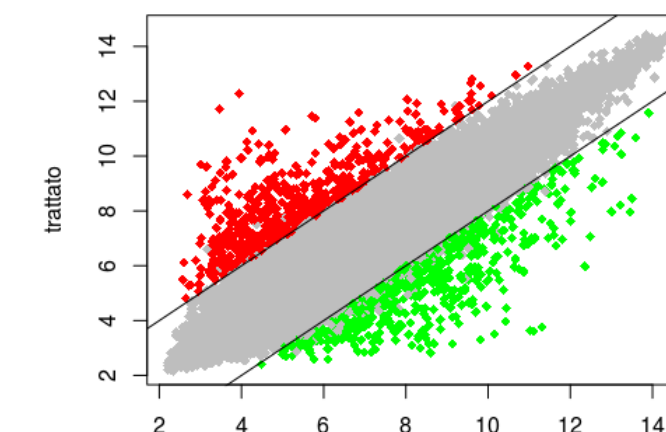
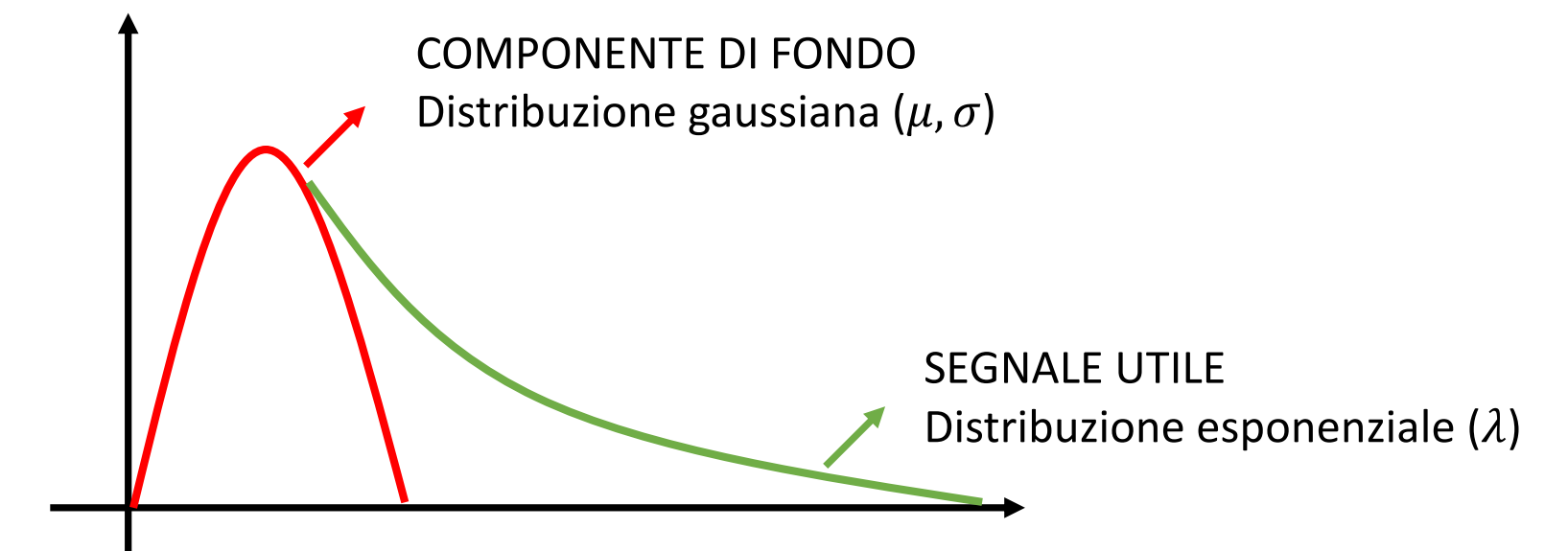
Ogni array rappresenta il profilo di espressione genica di un campione

1. Normalizzare i dati

	Array1	Array2	Array3	Array4	Array5	Array6
1007_s_at	9.347249	9.568875	9.820663	11.066680	11.007247	10.792517
1053_at	8.700141	8.965633	8.847094	9.049528	9.064799	8.750276
117_at	5.800838	5.776382	5.795153	5.048881	5.246541	5.306622
121_at	7.214928	7.376463	7.379221	8.056964	8.018169	7.836705
1255_g_at	2.543790	2.637345	2.625955	2.614629	2.982887	2.584014
1294_at	6.099539	6.201490	6.424293	6.450105	6.318168	6.049103
1316_at	6.186417	6.114881	5.901316	6.172999	6.109812	5.739915
1320_at	4.334868	4.668779	4.650498	5.075092	5.192537	5.244987
1405_i_at	3.339094	3.189086	3.679687	8.734428	8.535258	7.914109

geni

2. Quantificare la differenza di espressione di ciascun gene tra le diverse condizioni



Applications

Differential Expression Analysis

$$FC = \frac{A}{B} = \frac{4}{2} = 2 \quad A \text{ è } 2 \text{ volte } B$$

	Tumore			Controllo		
geni	Array1	Array2	Array3	Array4	Array5	Array6
1007_s_at	9.347249	9.568875	9.820663	11.066680	11.007247	10.792517
1053_at	8.700141	8.965633	8.847094	9.049528	9.064799	8.750276
117_at	5.800838	5.776382	5.795153	5.048881	5.246541	5.306622
121_at	7.214928	7.376463	7.379221	8.056964	8.018169	7.836705
1255_g_at	2.543790	2.637345	2.625955	2.614629	2.982887	2.584014
1294_at	6.099539	6.201490	6.424293	6.450105	6.318168	6.049103
1316_at	6.186417	6.114881	5.901316	6.172999	6.109812	5.739915
1320_at	4.334868	4.668779	4.650498	5.075092	5.192537	5.244987
1405_i_at	3.339094	3.189086	3.679687	8.734428	8.535258	7.914109

Per ciascun gene:

$$FC = \frac{\text{media Condizione1}}{\text{media Condizione2}} = \frac{\text{media Tumore}}{\text{media Controllo}}$$

N.B. i valori sono in scala logaritmica

$$\log FC = (\text{media Tumore}) - (\text{media Controllo})$$

$$\log FC = (3.3 + 3.2 + 3.7)/3 - (8.7 + 8.5 + 7.9)/3 = 3.4 - 8.4 \approx -5$$

Applications

Differential Expression Analysis

$$FC = \frac{A}{B} = \frac{4}{2} = 2 \quad A \text{ è } 2 \text{ volte } B$$

	Tumore			Controllo		
geni	Array1	Array2	Array3	Array4	Array5	Array6
1007_s_at	9.347249	9.568875	9.820663	11.066680	11.007247	10.792517
1053_at	8.700141	8.965633	8.847094	9.049528	9.064799	8.750276
117_at	5.800838	5.776382	5.795153	5.048881	5.246541	5.306622
121_at	7.214928	7.376463	7.379221	8.056964	8.018169	7.836705
1255_g_at	2.543790	2.637345	2.625955	2.614629	2.982887	2.584014
1294_at	6.099539	6.201490	6.424293	6.450105	6.318168	6.049103
1316_at	6.186417	6.114881	5.901316	6.172999	6.109812	5.739915
1320_at	4.334868	4.668779	4.650498	5.075092	5.192537	5.244987
1405_i_at	3.339094	3.189086	3.679687	8.734428	8.535258	7.914109

Per ciascun gene:

$$FC = \frac{\text{media Condizione1}}{\text{media Condizione2}} = \frac{\text{media Tumore}}{\text{media Controllo}}$$

N.B. i valori sono in scala logaritmica

$$\log FC = (\text{media Tumore}) - (\text{media Controllo})$$

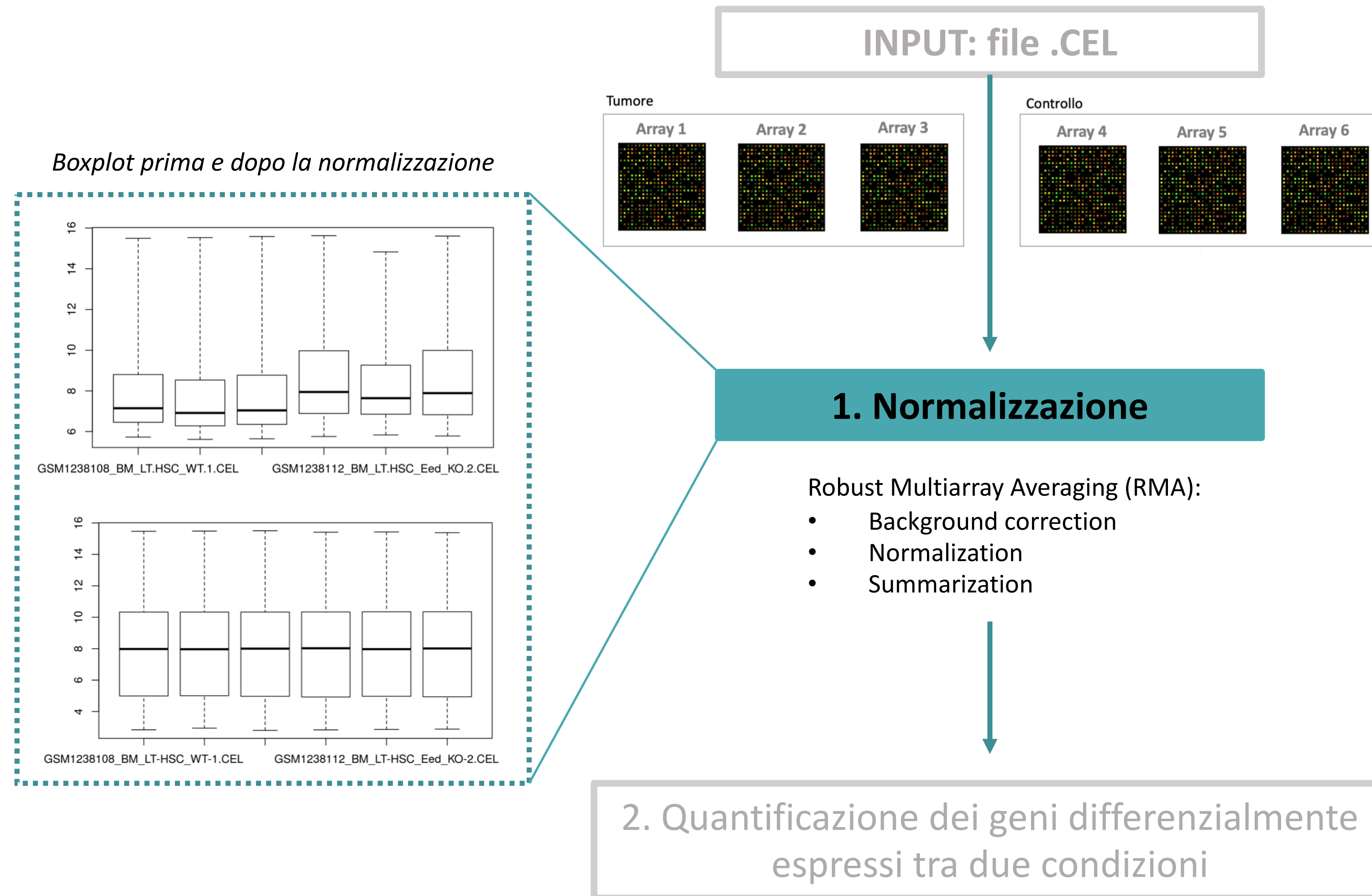
$$\log FC = (3.3 + 3.2 + 3.7)/3 - (8.7 + 8.5 + 7.9)/3 = 3.4 - 8.4 \approx -5$$

logFC > 1 gene up-regolato nel Tumore
logFC < -1 gene down-regolato nel Tumore



Applications

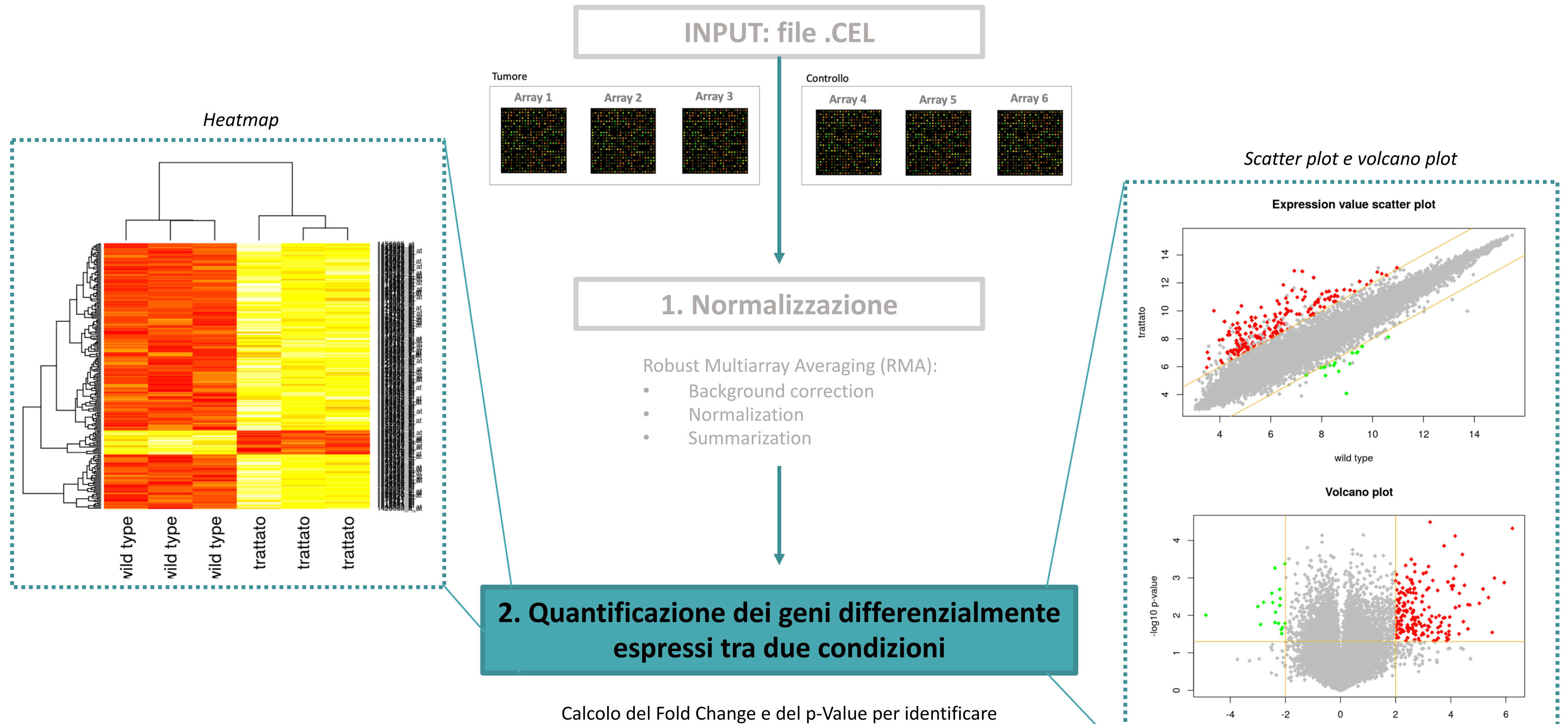
Differential Expression Analysis



Calcolo del Fold Change e del p-Value per identificare

Applications

Differential Expression Analysis



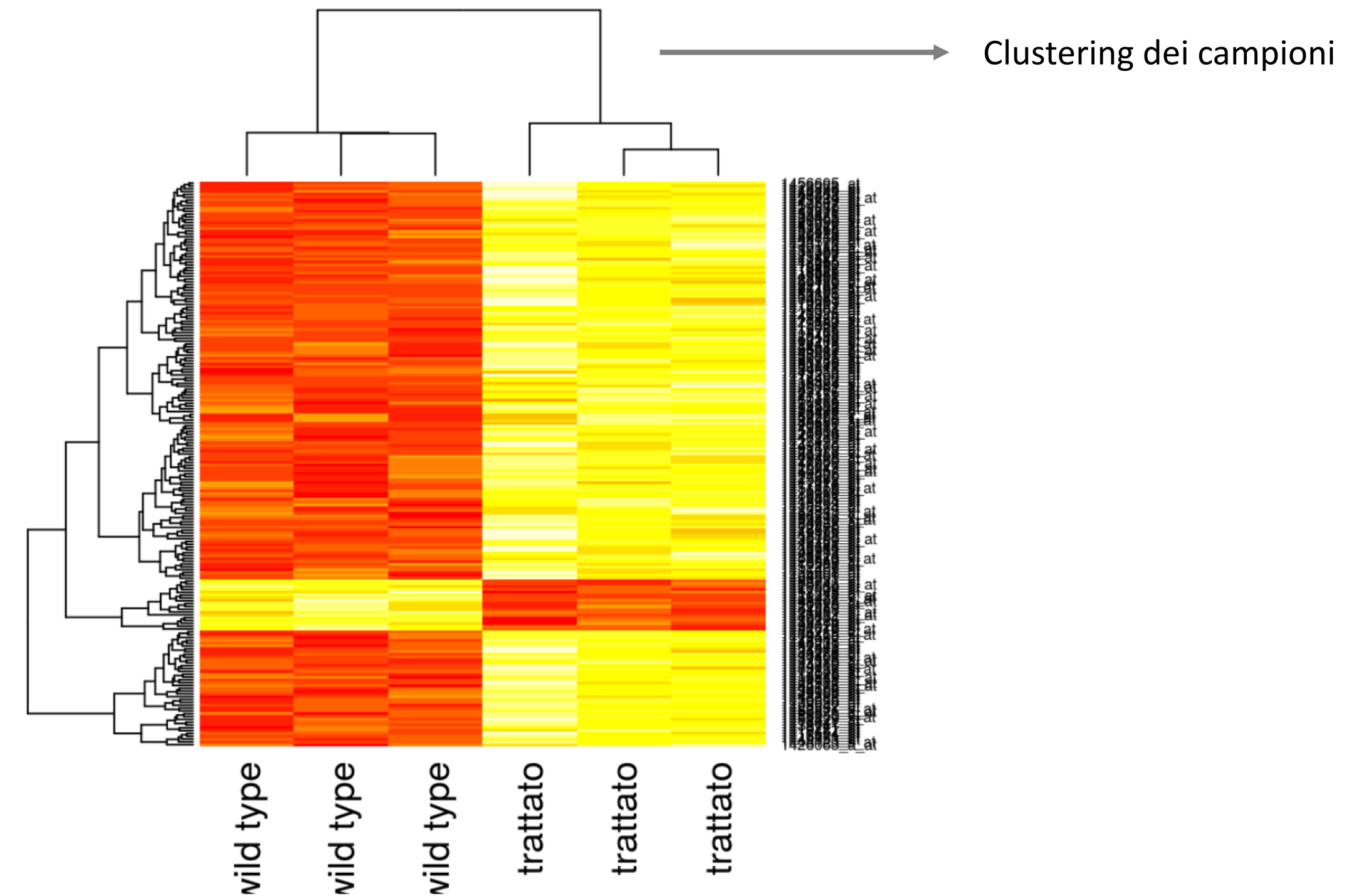
Applications

Differential Expression Analysis

Visualizzazione dei risultati

	Array1	Array2	Array3	Array4	Array5	Array6
1007_s_at	9.347249	9.568875	9.820663	11.066680	11.007247	10.792517
1053_at	8.700141	8.965633	8.847094	9.049528	9.064799	8.750276
117_at	5.800838	5.776382	5.795153	5.048881	5.246541	5.306622
121_at	7.214928	7.376463	7.379221	8.056964	8.018169	7.836705
1255_g_at	2.543790	2.637345	2.625955	2.614629	2.982887	2.584014
1294_at	6.099539	6.201490	6.424293	6.450105	6.318168	6.049103
1316_at	6.186417	6.114881	5.901316	6.172999	6.109812	5.739915
1320_at	4.334868	4.668779	4.650498	5.075092	5.192537	5.244987
1405_i_at	3.339094	3.189086	3.679687	8.734428	8.535258	7.914109

Clustering dei geni



Heatmap: una rappresentazione grafica dei dati dove ogni singolo valore nella matrice di espressione è rappresentato da un colore in una scala di colori (es dal giallo al rosso)

Applications

Differential Expression Analysis

=MEDIA(num1; num2; num3; ...)

Probe.id	trattato1	trattato2	trattato3	controllo1	controllo2	controllo3	mean exp trattato	mean exp controllo
1415670_at	10.32095475	10.3348446	10.32643855	10.0755224	10.103059	10.15155883		
1415671_at	11.55659724	11.21470877	11.6095877	11.33865796	11.33715745	11.47039553		
1415672_at	11.54761726	11.57386469	11.44209389	11.44663589	11.58652392	11.52606895		
1415673_at	9.92847925	7.720741249	7.620332624	7.817102331	8.009357375	7.837974306		
1415674_a_at	10.88265313	10.61796162	10.66808788	10.56361125	10.71527994	10.55999493		
1415675_at	9.733159903	9.807113078	10.03279695	9.858790847	10.03756243	10.00587701		
1415676_a_at	12.2327371	12.08600143	12.39492531	12.12988021	12.15886086	12.09462057		
1415677_at	8.494620928	7.57074354	8.532571262	7.877672468	8.006344756	7.969195083		
1415678_at	10.78227494	10.5624431	10.86456977	10.3011318	10.40717798	10.50016997		
1415679_at	11.69570084	11.63114164	11.67969986	11.60829637	11.6649613	11.64915692		
1415680_at	10.90050271	10.99138343	10.77633482	10.81999107	11.03762005	11.01180531		
1415681_at	10.85069521	11.06834829	10.98970299	10.84059181	10.94414474	10.87226999		
1415682_at	9.556735053	9.315819843	9.64109681	8.307379153	8.426017105	8.541887088		
1415683_at	11.59546026	11.73032524	11.44292373	11.55537457	11.76600437	11.65426006		
1415684_at	8.869762455	8.655886981	9.097590276	8.651620784	9.016728456	8.909072845		

Applications

Differential Expression Analysis

	A	B	C	D	E	F	G	H	I
1	Probe.id	trattato1	trattato2	trattato3	controllo1	controllo2	controllo3	mean exp trattato	mean exp controllo
2	1415670_at	10.32095475	10.3348446	10.32643855	10.0755224	10.103059	10.15155883	=MEDIA(B2:D2)	
3	1415671_at	11.55659724	11.21470877	11.6095877	11.33865796	11.33715745	11.47039553	MEDIA(num1; [num2]; ...)	
4	1415672_at	11.54761726	11.57386469	11.44209389	11.44663589	11.58652392	11.52606895		
5	1415673_at	9.92847925	7.720741249	7.620332624	7.817102331	8.009357375	7.837974306		
6	1415674_a_at	10.88265313	10.61796162	10.66808788	10.56361125	10.71527994	10.55999493		
7	1415675_at	9.733159903	9.807113078	10.03279695	9.858790847	10.03756243	10.00587701		

	A	B	C	D	E	F	G	H	I
1	Probe.id	trattato1	trattato2	trattato3	controllo1	controllo2	controllo3	mean exp trattato	mean exp controllo
2	1415670_at	10.32095475	10.3348446	10.32643855	10.0755224	10.103059	10.15155883	10.32741263	=MEDIA(E2:G2)
3	1415671_at	11.55659724	11.21470877	11.6095877	11.33865796	11.33715745	11.47039553		
4	1415672_at	11.54761726	11.57386469	11.44209389	11.44663589	11.58652392	11.52606895		
5	1415673_at	9.92847925	7.720741249	7.620332624	7.817102331	8.009357375	7.837974306		
6	1415674_a_at	10.88265313	10.61796162	10.66808788	10.56361125	10.71527994	10.55999493		
7	1415675_at	9.733159903	9.807113078	10.03279695	9.858790847	10.03756243	10.00587701		
8	1415676_a_at	12.2327371	12.08600143	12.39492531	12.12988021	12.15886086	12.09462057		

Applications

Differential Expression Analysis

	A	B	C	D	E	F	G	H	I	J
1	Probe.id	trattato1	trattato2	trattato3	controllo1	controllo2	controllo3	mean exp trattato	mean exp controllo	Fold change
2	1415670_at	10.32095475	10.3348446	10.32643855	10.0755224	10.103059	10.15155883	10.32741263	10.11004674	=H2-I2
3	1415671_at	11.55659724	11.21470877	11.6095877	11.33865796	11.33715745	11.47039553	11.4602979	11.38207031	
4	1415672_at	11.54761726	11.57386469	11.44209389	11.44663589	11.58652392	11.52606895	11.52119195	11.51974292	
5	1415673_at	9.92847925	7.720741249	7.620332624	7.817102331	8.009357375	7.837974306	8.423184375	7.888144671	
6	1415674_a_at	10.88265313	10.61796162	10.66808788	10.56361125	10.71527994	10.55999493	10.72290088	10.61296204	
7	1415675_at	9.733159903	9.807113078	10.03279695	9.858790847	10.03756243	10.00587701	9.857689977	9.967410096	
8	1415676_a_at	12.2327371	12.08600143	12.39492531	12.12988021	12.15886086	12.09462057	12.23788795	12.12778722	
9	1415677_at	8.494620928	7.57074354	8.532571262	7.877672468	8.006344756	7.969195083	8.19931191	7.951070769	
10	1415678_at	10.78227494	10.5624431	10.86456977	10.3011318	10.40717798	10.50016997	10.73642927	10.40282659	

Applications

Differential Expression Analysis

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Probe.id	trattato1	trattato2	trattato3	controllo1	controllo2	controllo3	mean exp trattato	mean exp controllo	Fold change	pValue		
2	1415670_at	10.32095475	10.3348446	10.32643855	10.0755224	10.103059	10.15155883	10.32741263	10.11004674	0.217365888	=TESTT(B2:D2)		
3	1415671_at	11.55659724	11.21470877	11.6095877	11.33865796	11.33715745	11.47039553	11.4602979	11.38207031	0.07822759	TESTT(matrice1; matrice2; coda; tipo)		
4	1415672_at	11.54761726	11.57386469	11.44209389	11.44663589	11.58652392	11.52606895	11.52119195	11.51974292	0.001449029			
5	1415673_at	9.92847925	7.720741249	7.620332624	7.817102331	8.009357375	7.837974306	8.423184375	7.888144671	0.535039704			
6	1415674_a_at	10.88265313	10.61796162	10.66808788	10.56361125	10.71527994	10.55999493	10.72290088	10.61296204	0.109938836			

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Probe.id	trattato1	trattato2	trattato3	controllo1	controllo2	controllo3	mean exp trattato	mean exp controllo	Fold change	pValue		
2	1415670_at	10.32095475	10.3348446	10.32643855	10.0755224	10.103059	10.15155883	10.32741263	10.11004674	0.217365888	=TESTT(B2:D2; E2:G2)		
3	1415671_at	11.55659724	11.21470877	11.6095877	11.33865796	11.33715745	11.47039553	11.4602979	11.38207031	0.07822759	TESTT(matrice1; matrice2; coda; tipo)		
4	1415672_at	11.54761726	11.57386469	11.44209389	11.44663589	11.58652392	11.52606895	11.52119195	11.51974292	0.001449029			
5	1415673_at	9.92847925	7.720741249	7.620332624	7.817102331	8.009357375	7.837974306	8.423184375	7.888144671	0.535039704			
6	1415674_a_at	10.88265313	10.61796162	10.66808788	10.56361125	10.71527994	10.55999493	10.72290088	10.61296204	0.109938836			

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Probe.id	trattato1	trattato2	trattato3	controllo1	controllo2	controllo3	mean exp trattato	mean exp controllo	Fold change	pValue		
2	1415670_at	10.32095475	10.3348446	10.32643855	10.0755224	10.103059	10.15155883	10.32741263	10.11004674	0.217365888	=TESTT(B2:D2; E2:G2; 2; 3)		
3	1415671_at	11.55659724	11.21470877	11.6095877	11.33865796	11.33715745	11.47039553	11.4602979	11.38207031	0.07822759	TESTT(matrice1; matrice2; coda; tipo)		
4	1415672_at	11.54761726	11.57386469	11.44209389	11.44663589	11.58652392	11.52606895	11.52119195	11.51974292	0.001449029			
5	1415673_at	9.92847925	7.720741249	7.620332624	7.817102331	8.009357375	7.837974306	8.423184375	7.888144671	0.535039704			
6	1415674_a_at	10.88265313	10.61796162	10.66808788	10.56361125	10.71527994	10.55999493	10.72290088	10.61296204	0.109938836			

Applications

Exercise 1 - differential expression analysis

The Transcription factor Xbp1 was over-expressed in the mouse adipocyte cell line F442A through adenoviral infection. A microarray experiment (GEO series GSE46178) was performed to compare gene expression profiles in cells in which Xbp1 is over-expressed compared to control cells. Four replicates were conducted for each of the two treatments

1. How many genes are upregulated? And how many genes are downregulated? Generate the plots of the differentially expressed genes in the two conditions.
2. With FIMO (<https://meme-suite.org/meme/tools/fimo>) find among the up genes which is directly regulated by Xbp1

Applications

Exercise 1 - differential expression analysis



[GEO Publications](#) [FAQ](#) [MIAME](#) [Email GEO](#)

NCBI » GEO » GEO2R » GSE46178

[Login](#)

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. [Full instructions](#) [YouTube](#)

GEO accession **Set** [Overexpression of spliced XBP1 in cultured Adipocytes](#)

► **Samples**

► [Define groups](#)

Selected **8** out of **8** samples

GEO2R

[Options](#)

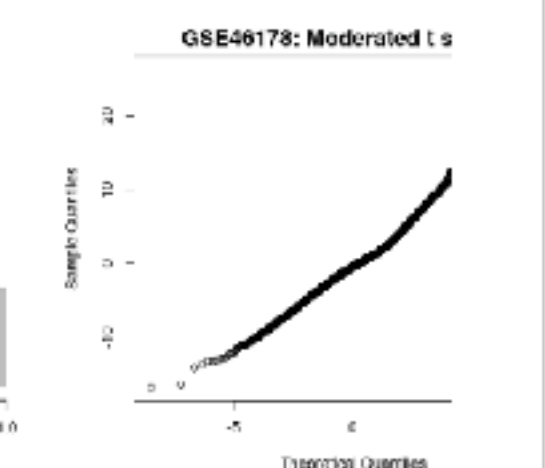
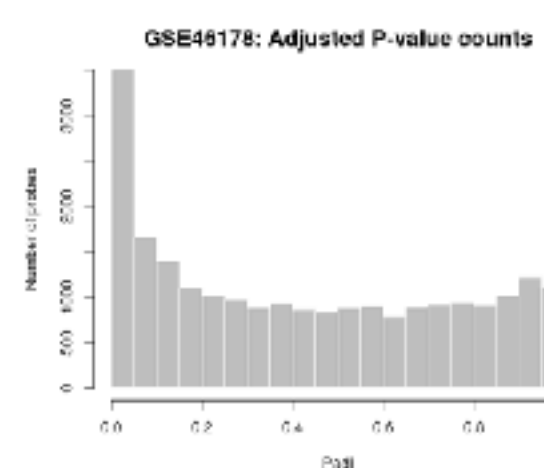
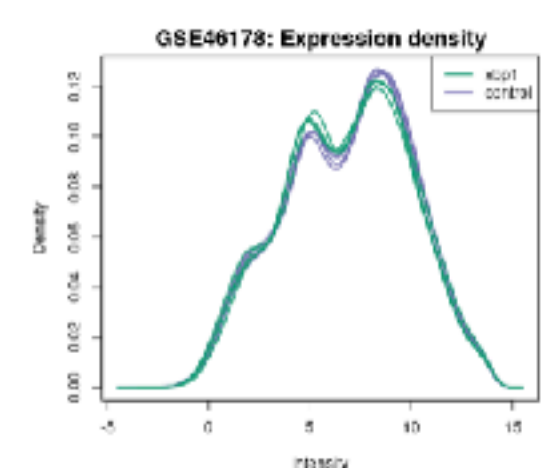
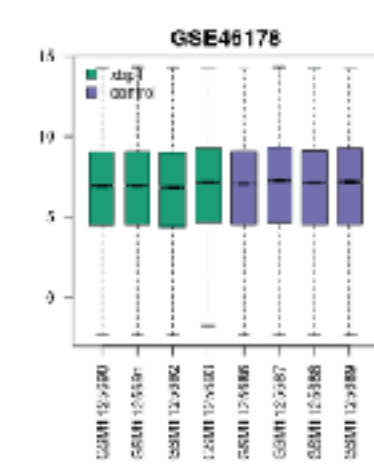
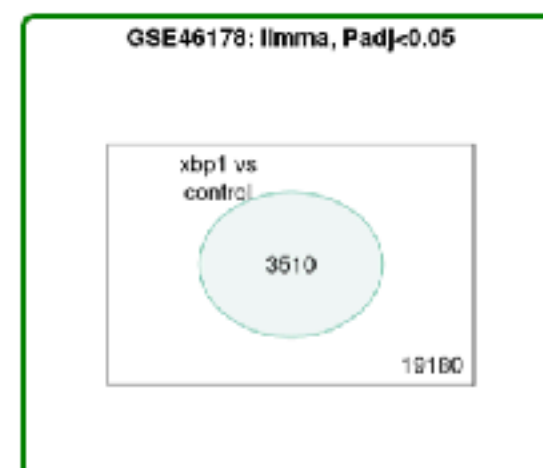
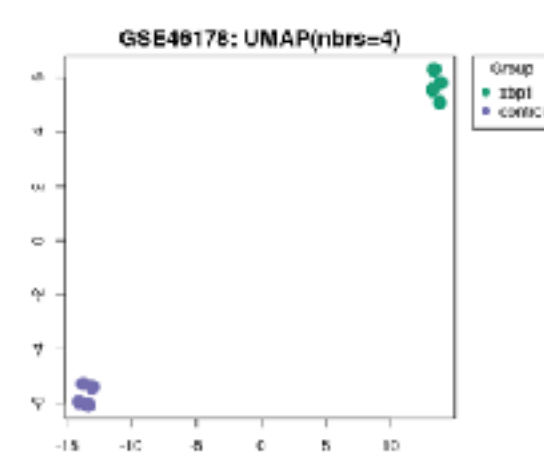
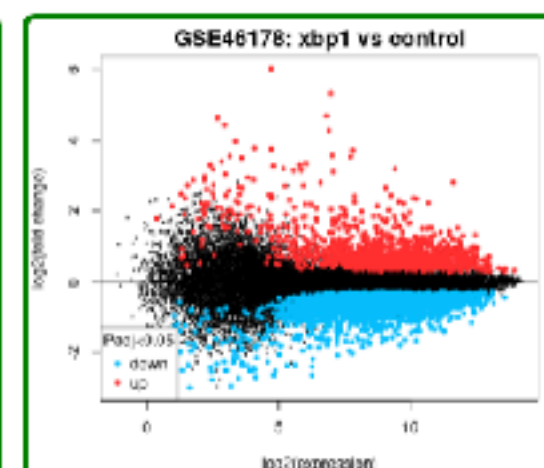
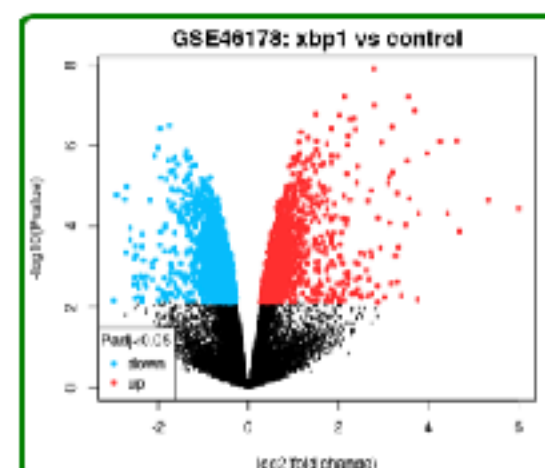
[Profile graph](#)

[R script](#)

Log-transformation has been applied to the data. You can change this in the Options tab.

[Reanalyze](#) if you changed any options.

Visualization [?](#)



Applications

Exercise 3 - Master Regulator analysis

1. Perform a Master regulator off the previous exercise (GEO series GSE46178) to find top regulator mechanism altered by Id2 knockout (Mouse regulons provided from STRING)
2. Get insights about top regulators with <https://www.genecards.org>

Applications

Exercise 3 - differential expression analysis and Master Regulator

Eye development and photoreceptor maintenance requires the retinal pigment epithelium (RPE), a thin layer of cells that underlies the neural retina (GEO series GSE5048). Three independent replicates of Zebrafish retina with RPE attached consisting of ten samples each at 52hpf (WRR52) were compared with three independent pure retinal samples consisting of ten retinas each at 52hpf (WR52).

1. Analyse the dataset performing a Master Regulator Analysis (Danio Rerio regulons file provided from STRING)
2. Get insights about top regulators with <https://www.genecards.org>