

Sequencing-Data from the Public Domain Bioinformatics for Wetlab-Scientists

GRADE Brain Workshop

Vera Laub
PhD candidate
Edinger Institute
25th of May, 2022

Recap of ChIP-seq Analysis

What did you find out about the binding of ARNTL to genomic regions?

Which questions remained?

Agenda

Day 1: ChIP-seq

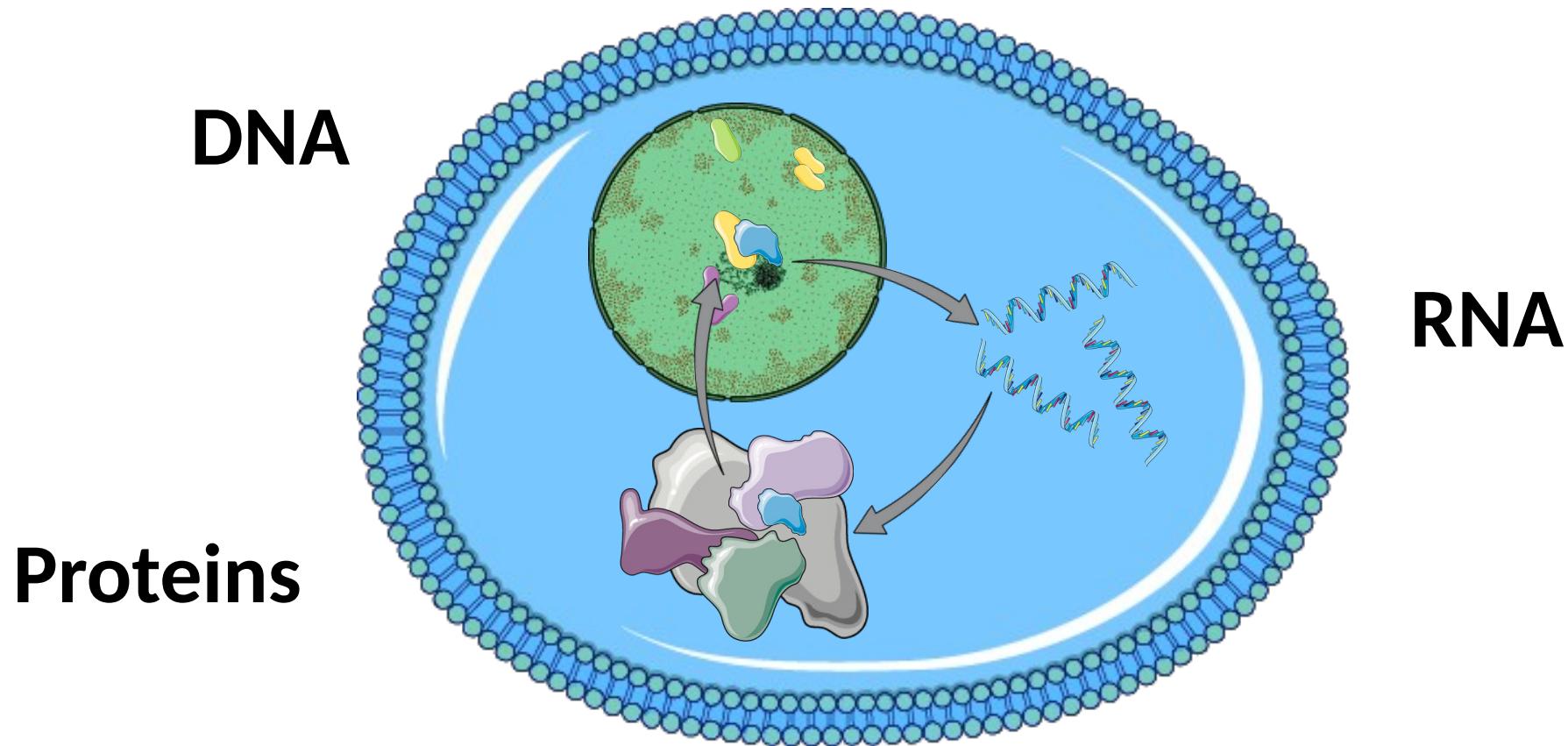
1. Introduction
2. Genomics Basics
3. ChIP-seq
4. Public Repositories
5. Data Visualization: IGV
6. Peak Annotation:
ChIPseeker

Day 2: RNA-seq

1. RNA-seq
2. Data Intersection: Venny
3. GO Term Analysis:
PANTHER
4. Break
5. GO Term Analysis:
STRING
6. Network Analysis: STRING

What are your expectations for today?

Data Integration

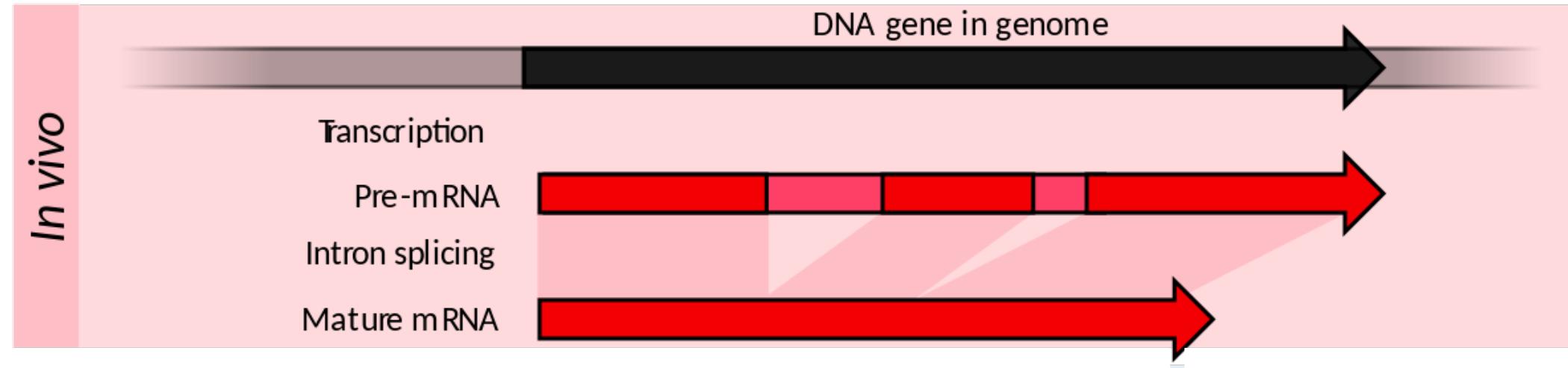


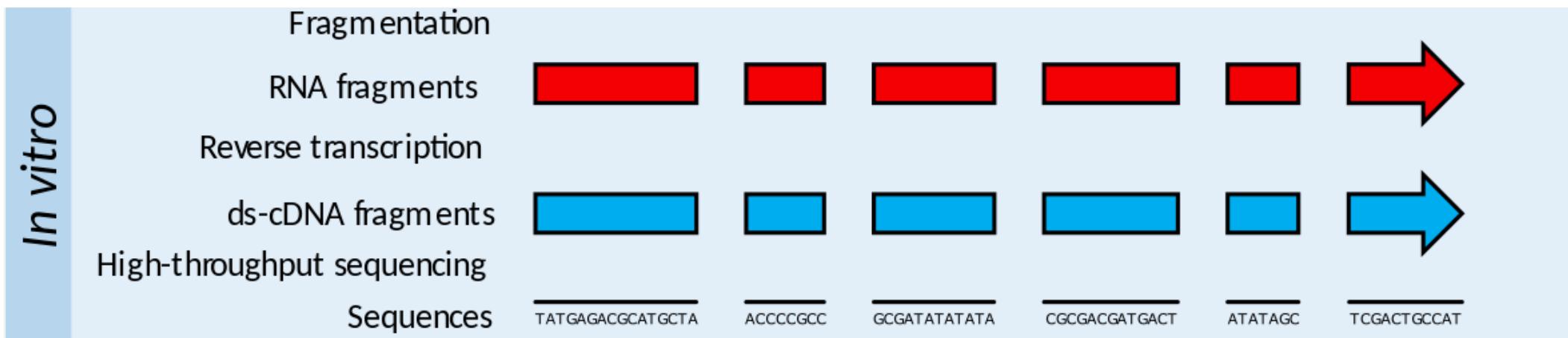
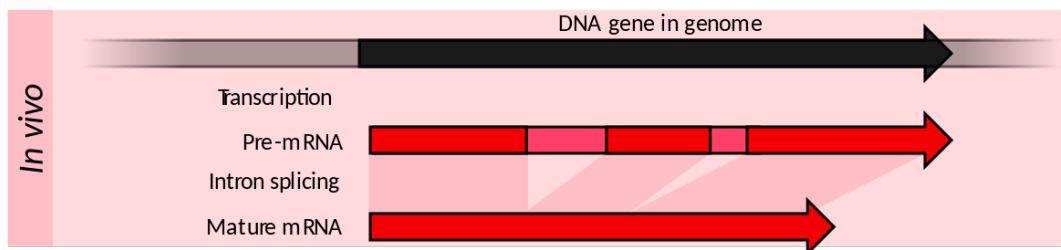
- Identification and quantification of
 - transcripts and their splice isoforms

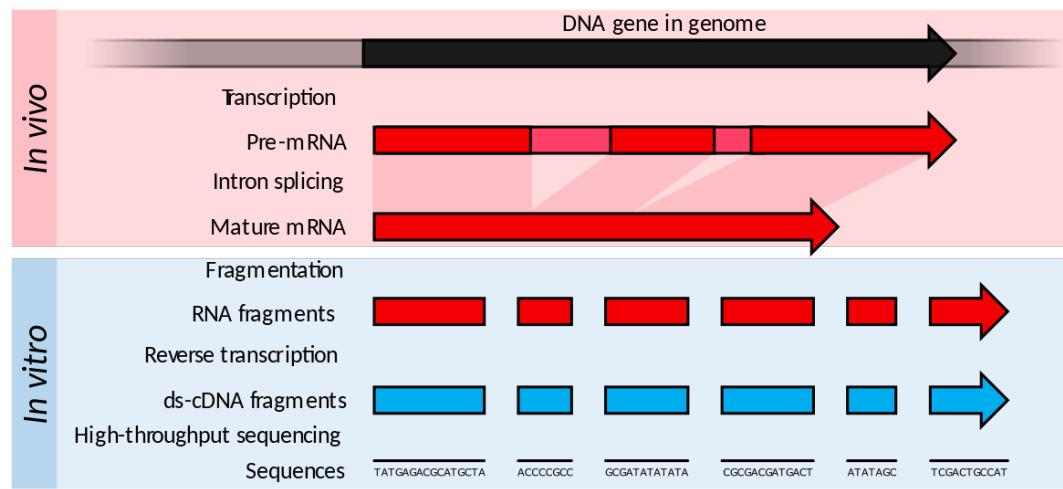
- Identification and quantification of
 - transcripts and their splice isoforms
 - novel transcriptionally active regions

- Identification and quantification of
 - transcripts and their splice isoforms
 - novel transcriptionally active regions
 - RNA editing products

- Identification and quantification of
 - transcripts and their splice isoforms
 - novel transcriptionally active regions
 - RNA editing products
 - mapping of transcribed regions



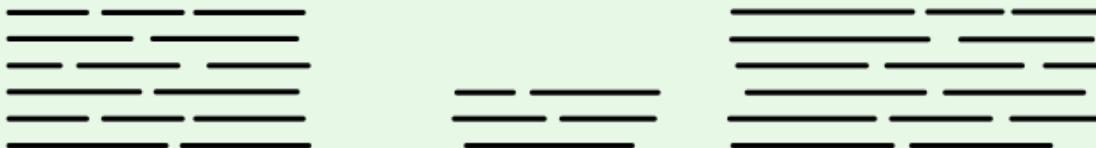




In silico

Sequence processing

Alignment

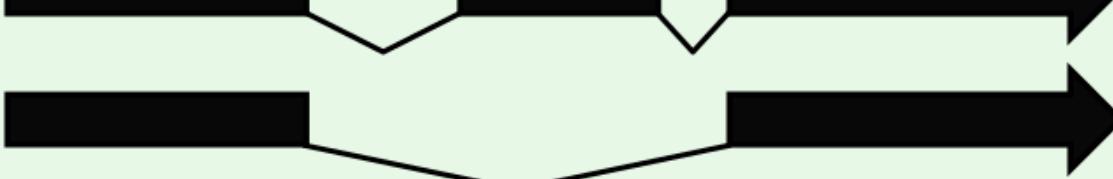


Genome sequence

Splice variant A

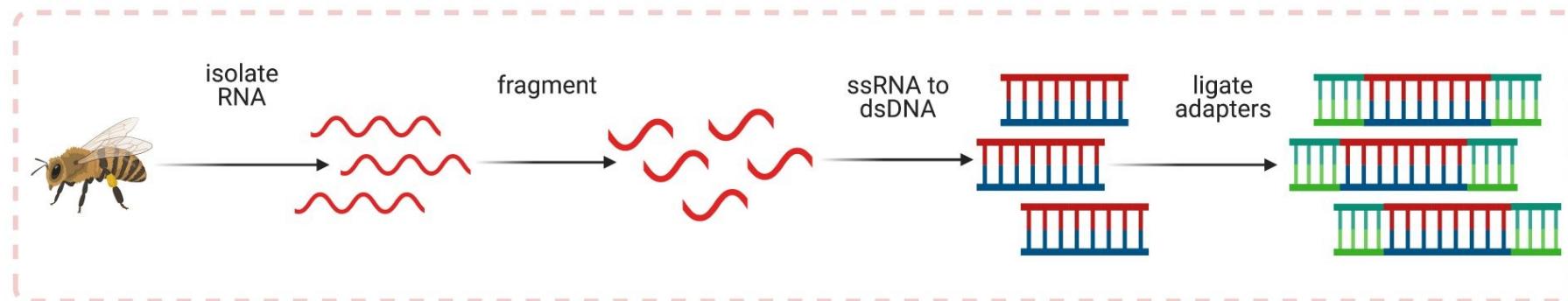


Splice variant B

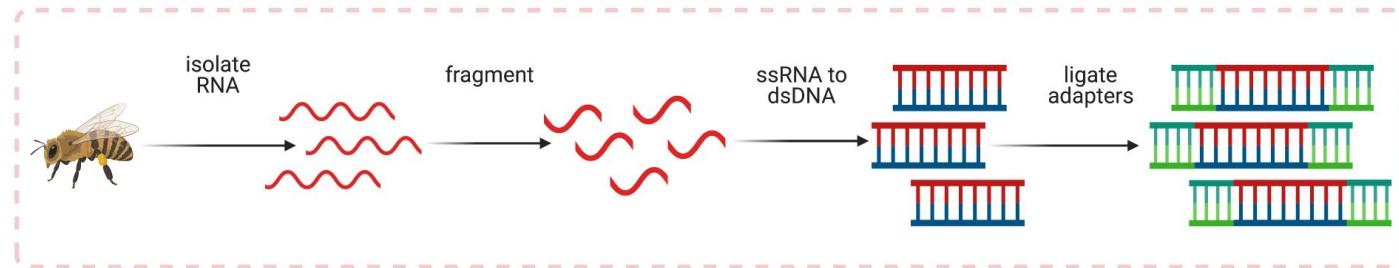


RNA-seq: Workflow

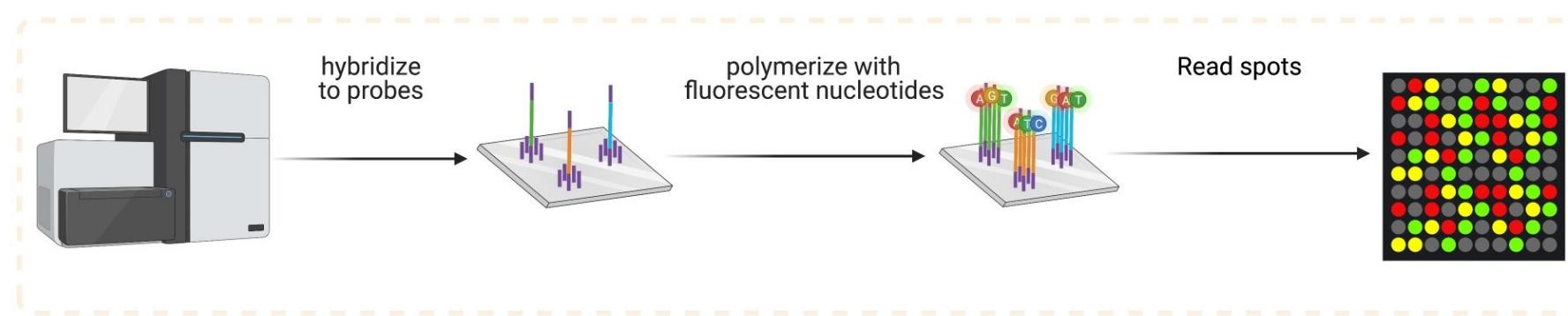
(I) RNA-seq library preparation



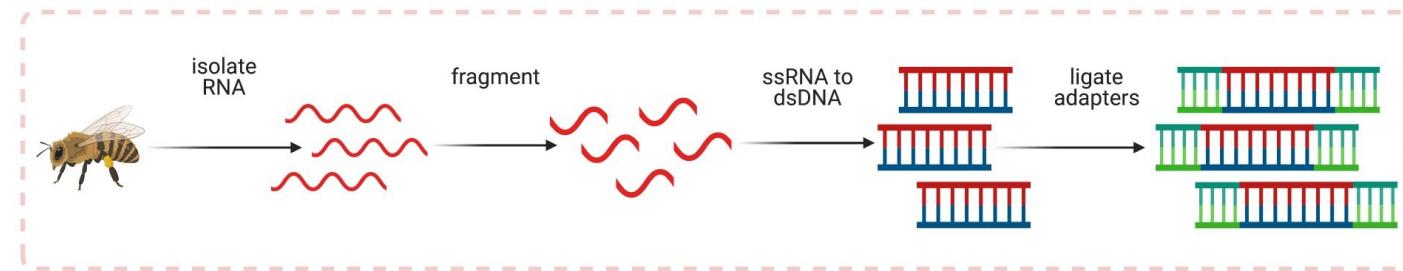
(I) RNA-seq library preparation



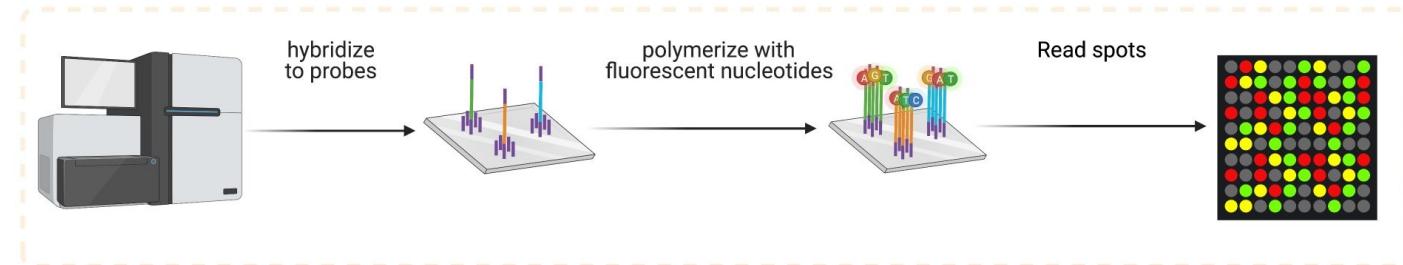
(II) RNA-seq



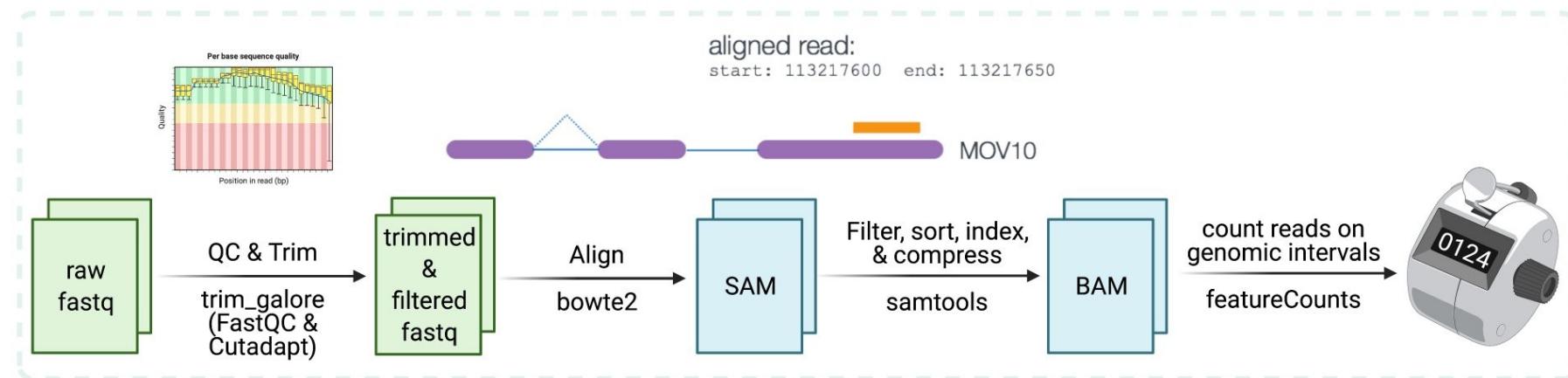
(I) RNA-seq library preparation



(II) RNA-seq



(III) Gene expression quantification



Retrieval of RNA-seq Data

- no standardized deposition in public repositories

Retrieval of RNA-seq Data

- **no standardized deposition** in public repositories
→ analyzed data often in supplementary material (as .xlsx)

Retrieval of RNA-seq Data

- no standardized deposition in public repositories
→ analyzed data often in supplementary material (as .xlsx)

Functional & Integrative Genomics (2021) 21:1–16
<https://doi.org/10.1007/s10142-020-00757-6>

ORIGINAL ARTICLE

=ARNTL
Brain and Mouse ARNT-Like 1

Transcriptome analysis of the circadian clock gene **BMAL1** deletion with opposite carcinogenic effects



Retrieval of RNA-seq Data

ESM 2

(DOCX 32 kb)

Table S1.

List of proteasomal or polymerase-related genes excluded from network analysis

(XLSX 18 kb)

Table S2. → download and inspect content

List of common DEGs in CIS and DOX treatments used to construct PPI network

(XLSX 33 kb)

ENSEMBL ID
= unique identifier

	A Gene stable ID	B Gene name	C Gene type
1	ENSG00000170161	AL512625.1	lncRNA
2	ENSG00000180066	LINC02870	lncRNA
3	ENSG00000197536	IRF1-AS1	lncRNA
4	ENSG00000223891	OSER1-DT	lncRNA
5	ENSG00000223960	CHROMR	lncRNA
6	ENSG00000225756	DBH-AS1	lncRNA
7	ENSG00000226007	BX005266.2	lncRNA
8	ENSG00000229956	ZRANB2-AS2	lncRNA
9	ENSG00000247708	STX18-AS1	lncRNA
10	ENSG00000254929	AL591684.2	lncRNA
11	ENSG00000257453	AC011611.3	lncRNA
12	ENSG00000258017	AC011603.3	lncRNA
13	ENSG00000259623	AC125257.1	lncRNA
14	ENSG00000260917	AL158212.3	lncRNA
15	ENSG00000261534	AL596244.1	lncRNA
16	ENSG00000265415	AC099850.4	lncRNA
17	ENSG00000272084	AL137127.1	lncRNA
18	ENSG00000275964	AL355001.2	lncRNA
19	ENSG00000276077	CU633904.2	lncRNA
20	ENSG00000277011	AC148477.4	lncRNA
21	ENSG00000277067	CU634019.2	lncRNA
22	ENSG00000274012	RN7SL2	misc_RNA
23	ENSG00000276168	RN7SL1	misc_RNA

gene names are
not unique
(pseudonyms, e.g.
ARNTL = **BMAL1**,
JAP3, **bHLH5**)

Stable IDs:

aim to be **unambiguous** and **consistent**

- Ensembl ID
 - ENS[species][feature][unique 11 digit number],
e.g. ENSMUSG00000055116 = Arntl mouse

Gene Identifiers

Stable IDs:

aim to be unambiguous and consistent

- Ensembl ID
 - ENS[species][feature][unique 11 digit number]

**When working with pre-analyzed public datasets
we have to work with what we find!**

Data Intersection with Venny

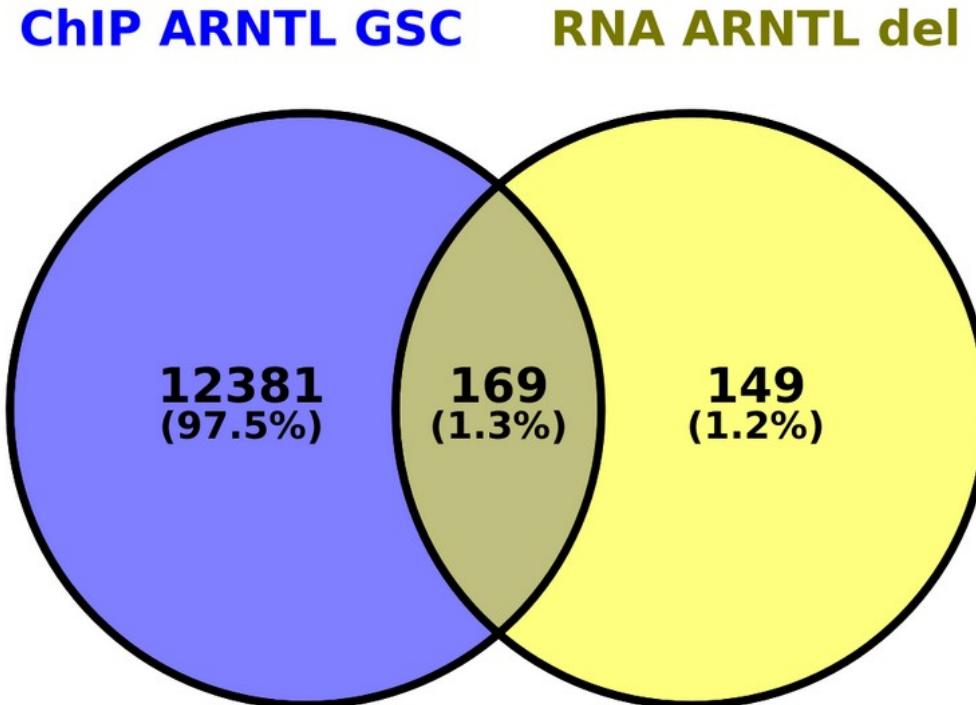
- tool to intersect up to 4 datasets
→ require common ID

Data Intersection with Venny

- tool to intersect up to 4 datasets
 - require common ID
- Output:
 - Venn-Diagram
 - list of all items in each intersection

Data Intersection with Venny

- tool to intersect up to 4 datasets
→ require common ID
- Output:
 - Venn-Diagram
 - list of all items in each intersection



Why intersect sequencing datasets
from different sources?

Dataset Intersection with Venny

VENNY 2.1 By Juan Carlos Oliveros
BioinfoGP, CNB-CSIC

1. Paste up to four lists. One element per row ([example](#)),
2. Click the numbers to see the results,
3. Right-click the figure to view and save it
(actual size in pixels: 1280x1280)

UPPERCASE lowercase ←cannot be undone!

List 1 0 List 2 0 ← assign name to dataset

Input dataset 1

Input dataset 2

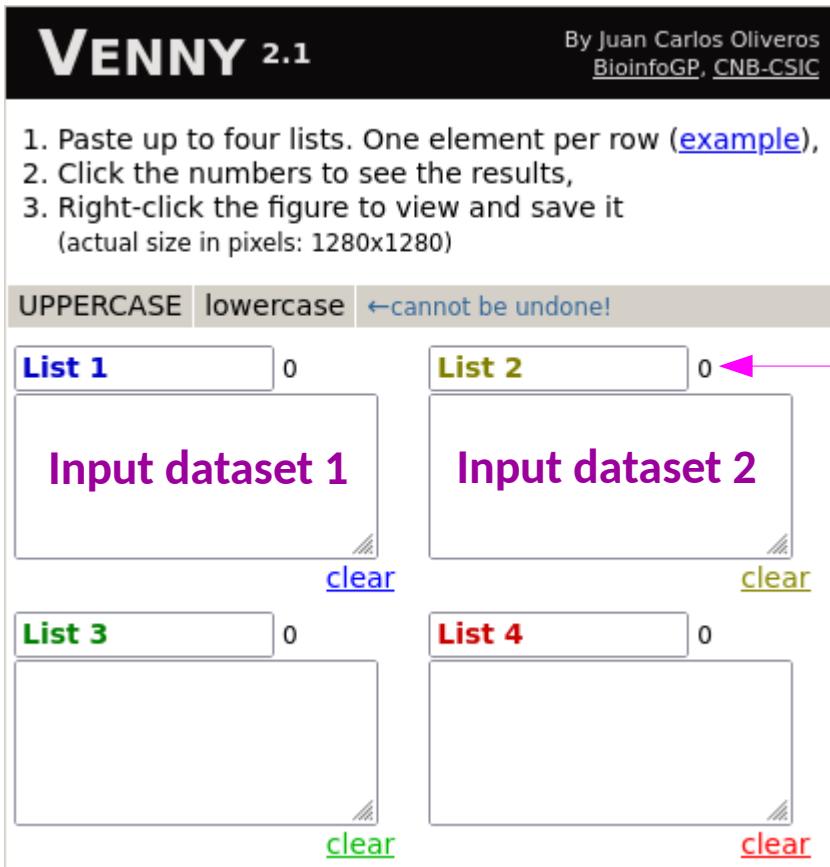
clear clear

List 3 0 List 4 0

Input dataset 3

Input dataset 4

clear clear



1. Paste up to four lists. One element per row ([example](#)),
 2. Click the numbers to see the results,
 3. Right-click the figure to view and save it
- (actual size in pixels: 1280x1280)

UPPERCASE lowercase ←cannot be undone!

ChIP ARNTL GSC 12550

RNA5-8S5
MIR4734
FP671120.2
U7
RP11-292F9.1

clear

RNA ARNTL del 318

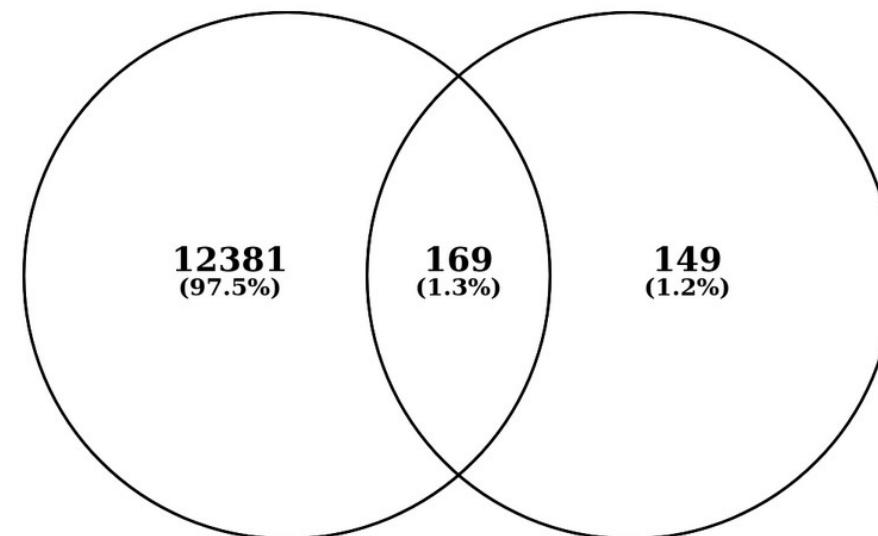
AC141586.1
ANAPC1P4
UBBP4
MST02P
AC083899.1

clear

customize Venn-Diagram

ChIP ARNTL GSC

RNA ARNTL del



UPPERCASE lowercase ←cannot be undone!

ChIP ARNTL GSC 12550

RNA5-8S5
MIR4734
FP671120.2
U7
RP11-292F9.1

[clear](#)

RNA ARNTL del 318

AC141586.1
ANAPC1P4
UBBP4
MST02P
AC083899.1

[clear](#)

Results:

169 common elements in "ChIP ARNTL GSC" and "RNA ARNTL del"

HECW1

MAP3K14

JARID2

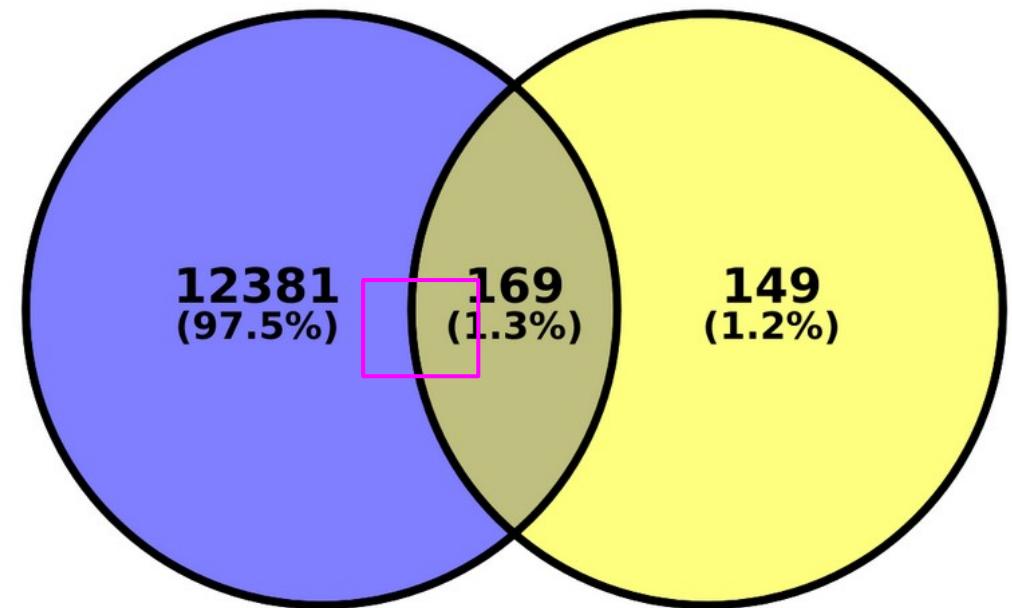
TUBG2

INPP4A

ARID1B



ChIP ARNTL GSC **RNA ARNTL del**



click on intersection of circles to view results

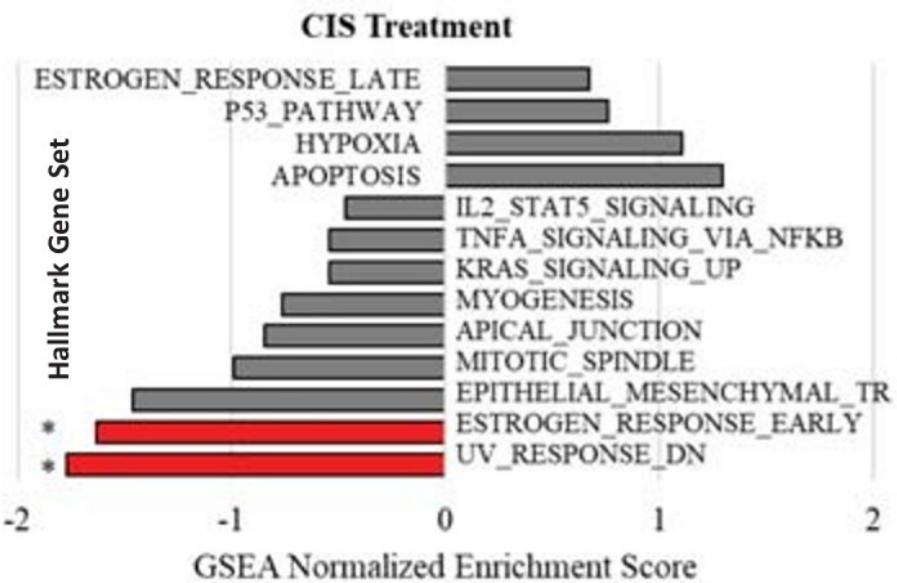
Practical Session

Task: Intersect 2 datasets with Venny
(ChIPseeker analysis, RNA-seq)

What problems did you encounter?

GO Term Analysis

Enriched MSigDB



KEGG_GOTerm for CISPLATIN

Wnt signaling pathway

Axon guidance

Adherens junction

Mucin type O-glycan biosynthesis

Other types of O-glycan biosynthesis

PI3K-Akt signaling pathway

Focal adhesion

ECM-receptor interaction

Human papillomavirus infection

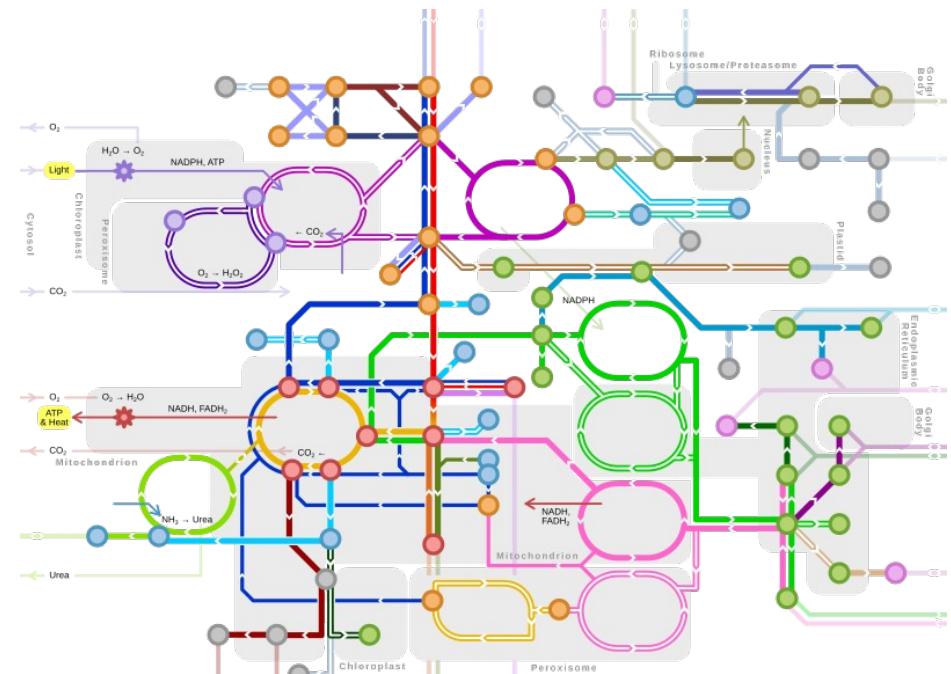
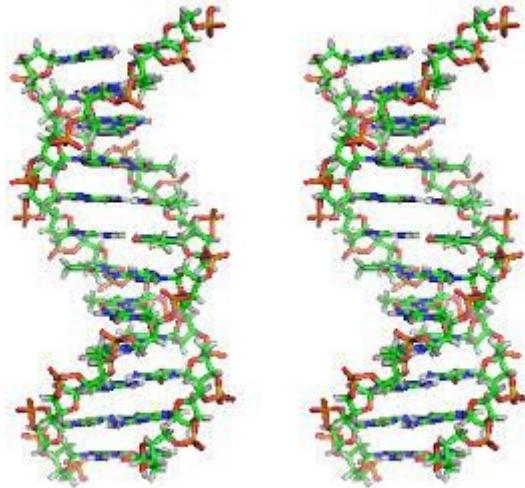
How have you come across or performed GO Term analysis before?

Gene Ontology

- representation of **CURRENT** knowledge on genes

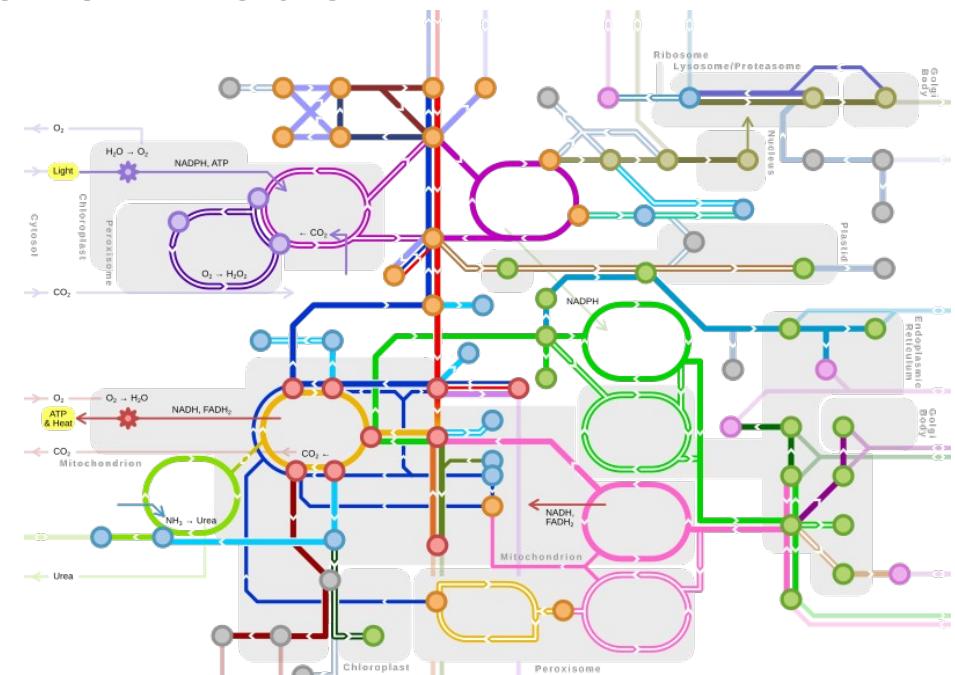
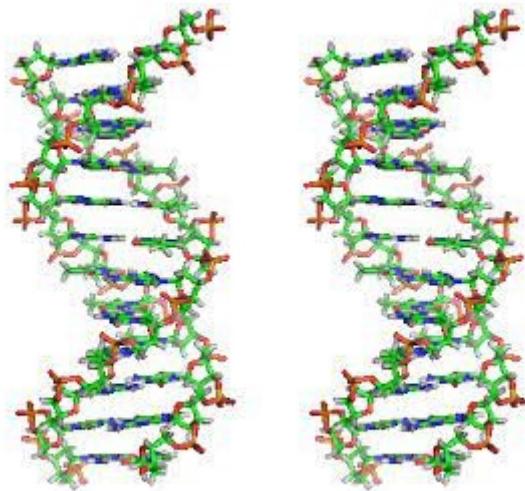
Gene Ontology

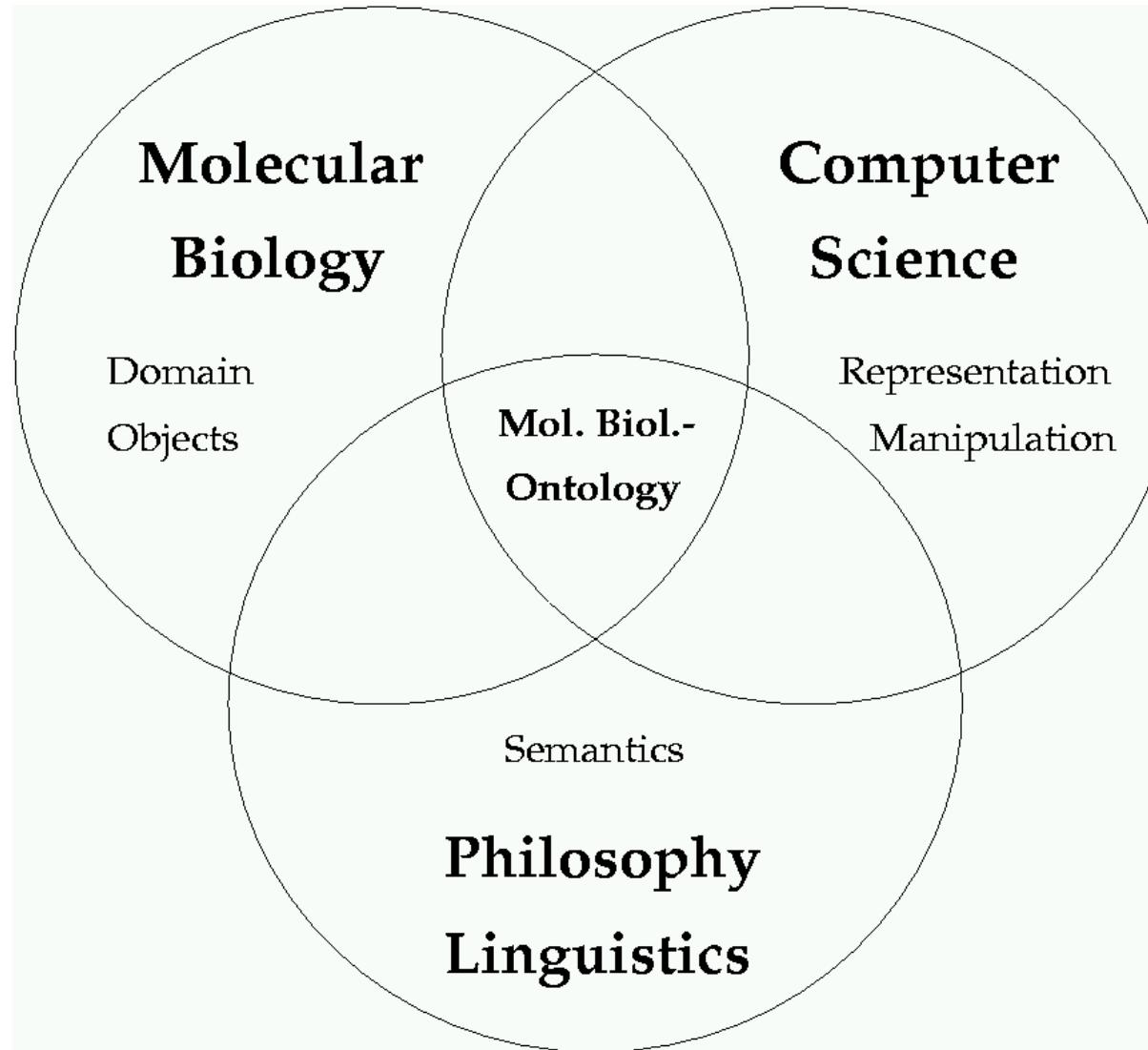
- representation of **CURRENT** knowledge on genes



Gene Ontology

- representation of **CURRENT** knowledge on genes
- **comprehensive**, computational model





- integrates resources from a variety of research groups



GENEONTOLOGY
Unifying Biology

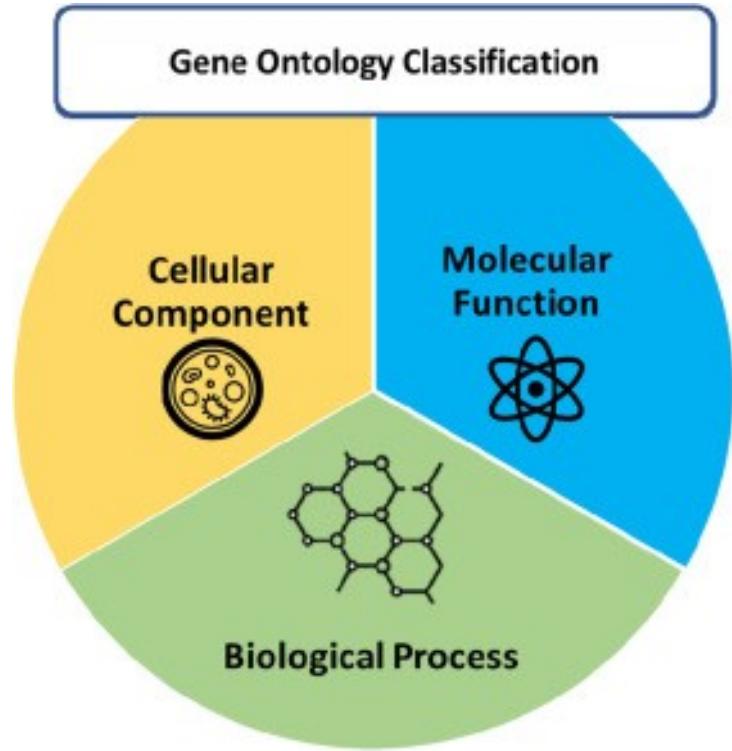
- integrates resources from a variety of research groups
- different model organisms

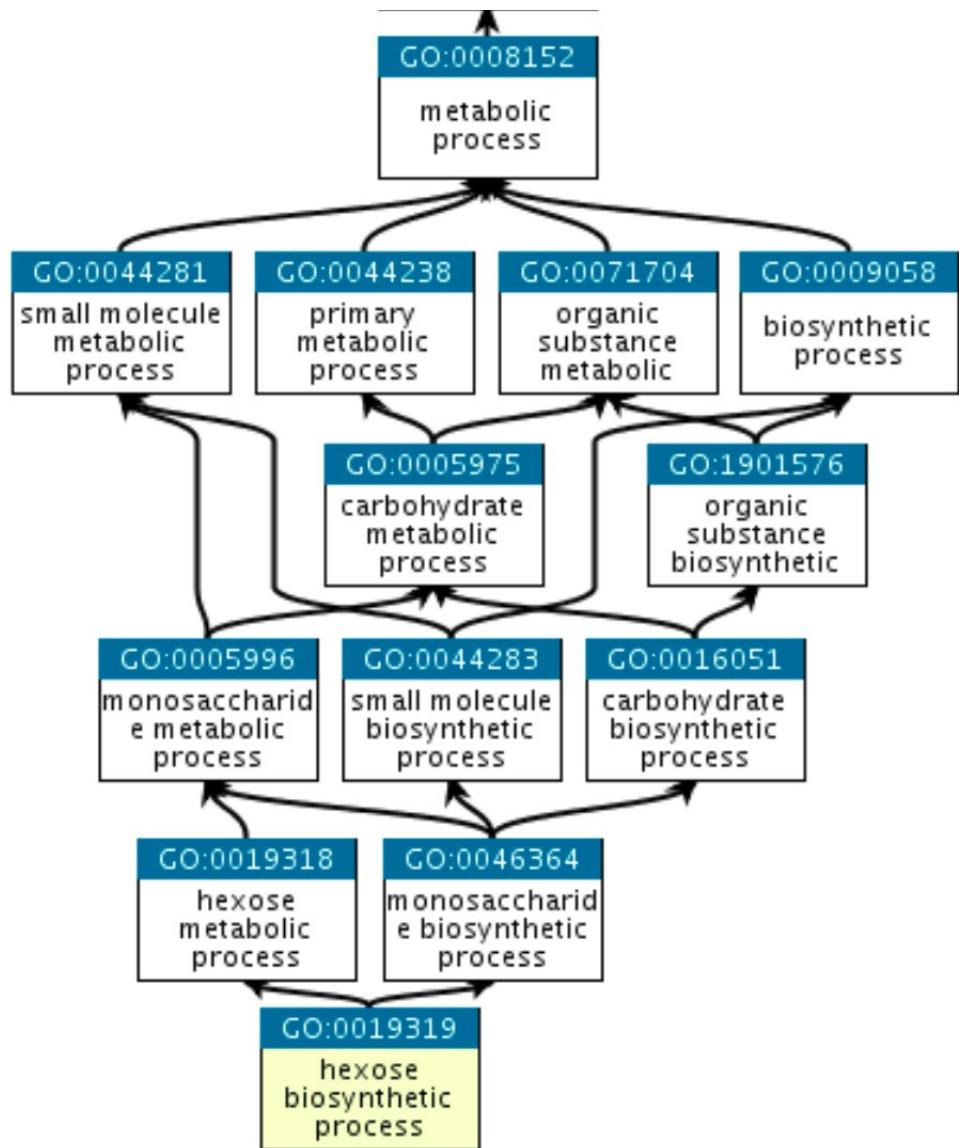
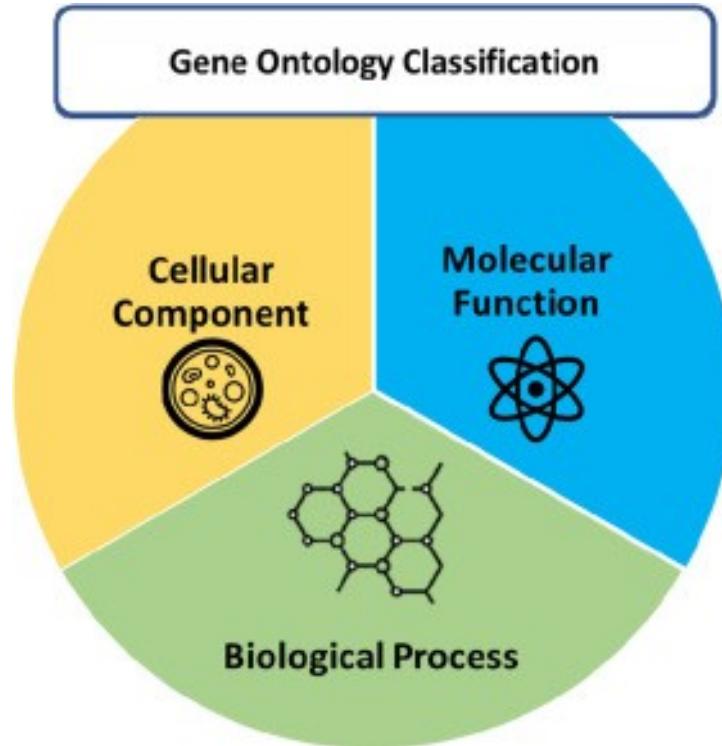


GENEONTOLOGY
Unifying Biology

- integrates resources from a variety of research groups
- different model organisms
- gene and protein databases







Which pitfalls do you see in GO as method to decipher biological function?



GO Term Analysis with PANTHER

= Protein ANalysis THrough Evolutionary Relationships

GO Term Analysis with PANTHER

= Protein ANalysis THrough Evolutionary Relationships

Input: List of genes

= Protein ANalysis THrough Evolutionary Relationships

Input: List of genes

Output:

- Functional classification: Gene list or graphic chart
- Statistical tests (overrepresentation, enrichment)

enter dataset of interest

Gene List Analysis

Browse

Sequence Search

cSNP Scoring

Keyword Search

Please refer to our article in [Nature Protocols](#) for detailed instructions on how to use this page.

Help Tips

Steps:

- 1. Select list and list type to analyze
- 2. Select Organism
- 3. Select operation

Using enhancer data

1. Enter ids and or select file for batch upload. Else enter ids or select file or list from workspace for comparing to a reference list.

Enter IDs:

[Supported](#)

[IDs](#)

HECW1
MAP3K14
JARID2

separate IDs by a space or comma

Upload

IDs:

[File](#)

[format](#)

[Browse...](#) No file selected.

Select

List Type:

ID List

Previously exported text search results

Workspace list

PANTHER Generic Mapping

ID's from Reference Proteome Genome

Organism for id list [Absidia glauca \(ABSGL\)](#)

VCF File Flanking region [20 Kb](#) Search Enhancer Data

Functional Classification

2.

Select organism.

Homo sapiens

Mus musculus

Rattus norvegicus

Gallus gallus

Danio rerio

assign organism of
dataset

Functional Classification

2. Select organism.

Homo sapiens

Mus musculus

Rattus norvegicus

Gallus gallus

Danio rerio

assign organism of dataset

3. Select Analysis.

Functional classification viewed in gene list

Functional classification viewed in graphic charts

Bar chart

Pie chart

Statistical overrepresentation test

Statistical enrichment test

choose analysis, then submit

Functional Classification

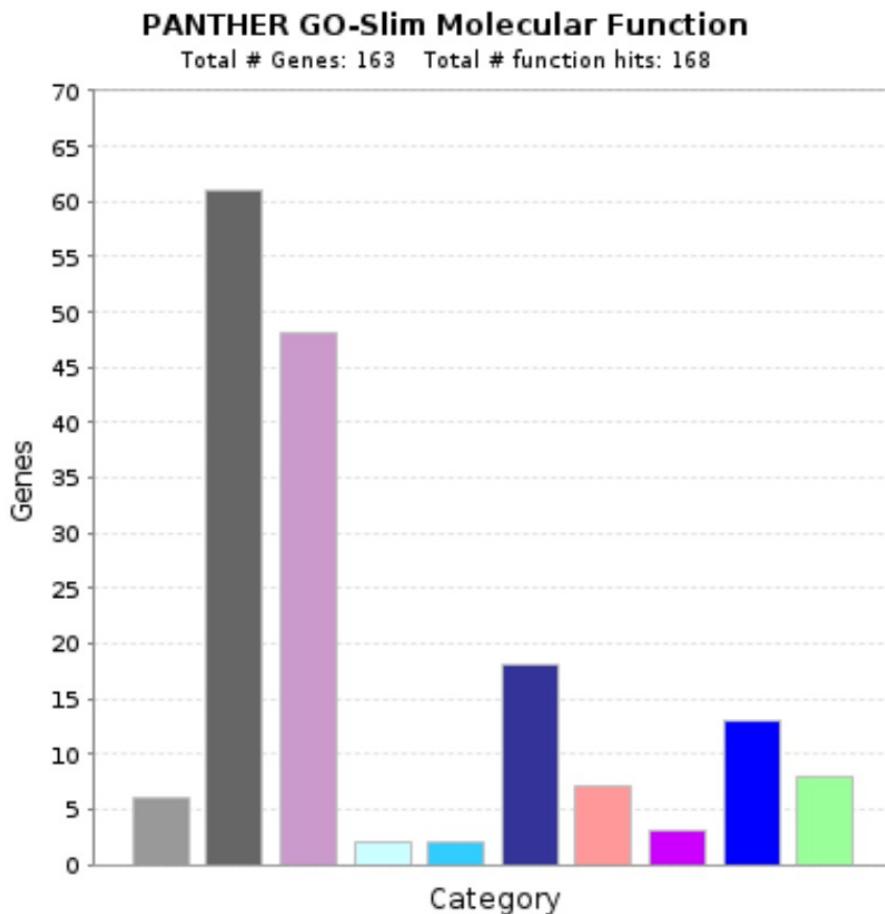
Select Ontology:

Molecular Function ▾

View: 100% ▾

[choose ontology](#)

Functional Classification



Select Ontology:

Molecular Function ▾

View: 100% ▾

choose ontology

Click to get gene list for a category:

- █ [ATP-dependent activity \(GO:0140657\)](#)
- █ [binding \(GO:0005488\)](#)
- █ [catalytic activity \(GO:0003824\)](#)
- █ [cytoskeletal motor activity \(GO:0003774\)](#)
- █ [molecular adaptor activity \(GO:0060090\)](#)
- █ [molecular function regulator \(GO:0098772\)](#)
- █ [molecular transducer activity \(GO:0060089\)](#)
- █ [structural molecule activity \(GO:0005198\)](#)
- █ [transcription regulator activity \(GO:0140110\)](#)
- █ [transporter activity \(GO:0005215\)](#)

Statistical Overrepresentation Test

3. Select Analysis.

- Functional classification viewed in gene list
 - Functional classification viewed in graphic charts
 - Statistical overrepresentation test
 - Statistical enrichment test
- GO biological process complete ▼ ?
- choose test and submit

Statistical Overrepresentation Test

Use a Reference List that includes all genes from a whole genome

Default whole-genome lists:

Homo sapiens genes



then choose reference (right side of screen)

Analysis Summary: Please report in publication [?](#)

Analysis Type: PANTHER Overrepresentation Test (Released 20220202)

Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.6399963 Released 2022-03-22

Analyzed List: Client Text Box Input (Homo sapiens) [Change](#)

Reference List: Homo sapiens (all genes in database) [Change](#)

Annotation Data Set: GO biological process complete [?](#)

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing [?](#) No correction

check summary, choose test type, then launch analysis

Statistical Overrepresentation Test: Output

Results

Uniquely Mapped IDs:

Unmapped IDs:

Multiple mapping information:

Reference list

20589 out of 20589

0

0

successfully mapped genes

Client Text Box Input

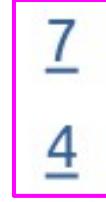
162 out of 163

7

4

Statistical Overrepresentation Test: Output

Results

	Reference list	Client Text Box Input
Uniquely Mapped IDs:	<u>20589</u> out of 20589	<u>162</u> out of 163
Unmapped IDs:	0	 7 4
Multiple mapping information:	0	

unmapped/ multiple mapped genes, click to view

Bonferroni count: 9201

Export [Table](#) [XML with user input ids](#) [JSON with user input ids](#)

[click to export](#)

Displaying only results for Bonferroni-corrected for P < 0.05, [click here to display all results](#)

	Homo sapiens (REF)	Client Text Box Input (▼ Hierarchy NEW! ?)				
	#	#	expected	Fold Enrichment	+/-	P value
GO biological process complete						
regulation of axon extension	96	8	.76	10.53	+	1.53E-02
↳ regulation of developmental process	2463	52	19.50	2.67	+	1.86E-07
↳ biological regulation	12469	127	98.72	1.29	+	3.29E-02
↳ regulation of extent of cell growth	111	8	.88	9.10	+	4.20E-02
↳ regulation of cellular component organization	2345	41	18.57	2.21	+	1.29E-02
↳ regulation of cellular process	11275	121	89.26	1.36	+	4.27E-03
↳ regulation of cell size	184	10	1.46	6.86	+	2.97E-02
↳ regulation of cellular component size	363	15	2.87	5.22	+	2.93E-03
↳ regulation of anatomical structure size	497	16	3.93	4.07	+	2.79E-02
↳ regulation of biological quality	3730	56	29.53	1.90	+	9.72E-03

terms with
p < 0.05
click to view
details for term

Bonferroni count: 9201

Export [Table](#) [XML with user input ids](#) [JSON with user input ids](#)

Displaying only results for Bonferroni-corrected for P < 0.05, [click here to display all results](#)

	Homo sapiens (REF)	Client Text Box Input (▼ Hierarchy NEW! ?)					
		#	#	expected	Fold Enrichment	+/-	P value
GO biological process complete							
regulation of axon extension	96	8	.76	10.53	+	1.53E-02	
↳ regulation of developmental process	2463	52	19.50	2.67	+	1.86E-07	
↳ biological regulation	12469	127	98.72	1.29	+	3.29E-02	
↳ regulation of extent of cell growth	111	8	.88	9.10	+	4.20E-02	
↳ regulation of cellular component organization	2345	41	18.57	2.21	+	1.29E-02	
↳ regulation of cellular process	11275	121	89.26	1.36	+	4.27E-03	
↳ regulation of cell size	184	10	1.46	6.86	+	2.97E-02	
↳ regulation of cellular component size	363	15	2.87	5.22	+	2.93E-03	
↳ regulation of anatomical structure size	497	16	3.93	4.07	+	2.79E-02	
↳ regulation of biological quality	3730	56	29.53	1.90	+	9.72E-03	

regulation of axon extension

Term Information

Accession GO:0030516

Name regulation of axon extension

Ontology biological_process

Synonyms None

Alternate IDs None

Definition Any process that modulates the rate, direction or extent of axon extension. *Source: GOC:go_curators*

Comment None

History See term [history for GO:0030516](#) at QuickGO

Subset None

Related [Link](#) to all **genes and gene products** annotated to regulation of axon extension (**excluding "regulates"**).

[Link](#) to all direct and indirect **annotations** to regulation of axon extension (**excluding "regulates"**).

[Link](#) to all direct and indirect **annotations download** (limited to first 10,000) for regulation of axon extension (**excluding "regulates"**).

[Include "regulates"](#)

For more information, please see the [ontology relation documentation](#).

regulation of axon extension

Term Information

Accession GO:0030516

Name regulation of axon extension

Ontology biological_process

Synonyms None

Alternate IDs None

Definition Any process that modulates the rate, direction or extent of axon extension. *Source: GOC:go_curators*

Comment None

History See term [history for GO:0030516](#) at QuickGO

Subset None

Related [Link](#) to all **genes and gene products** annotated to regulation of axon extension (**excluding "regulates"**).

[Link](#) to all direct and indirect **annotations** to regulation of axon extension (**excluding "regulates"**).

[Link](#) to all direct and indirect **annotations download** (limited to first 10,000) for regulation of axon extension (**excluding "regulates"**).

more info regarding genes in this GO

[Include "regulates"](#)

For more information, please see the [ontology relation documentation](#).

Bonferroni count: 9201

Export [Table](#) [XML with user input ids](#) [JSON with user input ids](#)

Displaying only results for Bonferroni-corrected for P < 0.05, [click here to display all results](#)

	Homo sapiens (REF)	Client Text Box Input (▼ Hierarchy NEW! ?)				
	#	#	expected	Fold Enrichment	+/-	P value
GO biological process complete						
regulation of axon extension	96	8	.76	10.53	+	1.53E-02
↳ regulation of developmental process	2463	52	19.50	2.67	+	1.86E-07
↳ biological regulation	12469	127	98.72	1.29	+	3.29E-02
↳ regulation of extent of cell growth	111	8	.88	9.10	+	4.20E-02
↳ regulation of cellular component organization	2345	41	18.57	2.21	+	1.29E-02
↳ regulation of cellular process	11275	121	89.26	1.36	+	4.27E-03
↳ regulation of cell size	184	10	1.46	6.86	+	2.97E-02
↳ regulation of cellular component size	363	15	2.87	5.22	+	2.93E-03
↳ regulation of anatomical structure size	497	16	3.93	4.07	+	2.79E-02
↳ regulation of biological quality	3730	56	29.53	1.90	+	9.72E-03

click number to view genes associated with each term

	<u>Gene ID</u>	Mapped IDs	<u>Gene Name</u> <u>Gene Symbol</u> <u>Ortholog</u>	<u>PANTHER Family/Subfamily</u>	<u>PANTHER Protein Class</u>	<input checked="" type="checkbox"/> <u>Species</u>
<input type="checkbox"/>	1. HUMAN HGNC=6893 UniProtKB=P10636	MAPT	Microtubule-associated protein tau MAPT ortholog	MICROTUBULE-ASSOCIATED PROTEIN TAU (PTHR11501:SF14)	-	Homo sapiens
<input type="checkbox"/>	2. HUMAN HGNC=29284 UniProtKB=Q9P265	DIP2B	Disco-interacting protein 2 homolog B DIP2B ortholog	DISCO-INTERACTING PROTEIN 2 HOMOLOG B (PTHR22754:SF38)	-	Homo sapiens
<input type="checkbox"/>	3. HUMAN HGNC=10731 UniProtKB=Q9C0C4	SEMA4C	Semaphorin-4C SEMA4C ortholog	SEMAPHORIN-4C (PTHR11036:SF16)	membrane-bound signaling molecule	Homo sapiens
<input type="checkbox"/>	4. HUMAN HGNC=9681 UniProtKB=Q13332	PTPRS	Receptor-type tyrosine-protein phosphatase S PTPRS ortholog	RECEPTOR-TYPE TYROSINE-PROTEIN PHOSPHATASE S (PTHR19134:SF449)	protein phosphatase	Homo sapiens

click to view subfamily information

Practical Session

Task: Perform GO Term analysis of intersected genes with PANTHER.

Breakout room 1: Anna, Sara, Miles

Breakout room 2: Leon, Felix, Tobias

Breakout room 3: Rebecca, Minju, Finn

Use links in the Handout/ Chat.

What problems did you encounter?

More tools for GO Term Analysis



There are multiple ways to perform GO Term analysis!

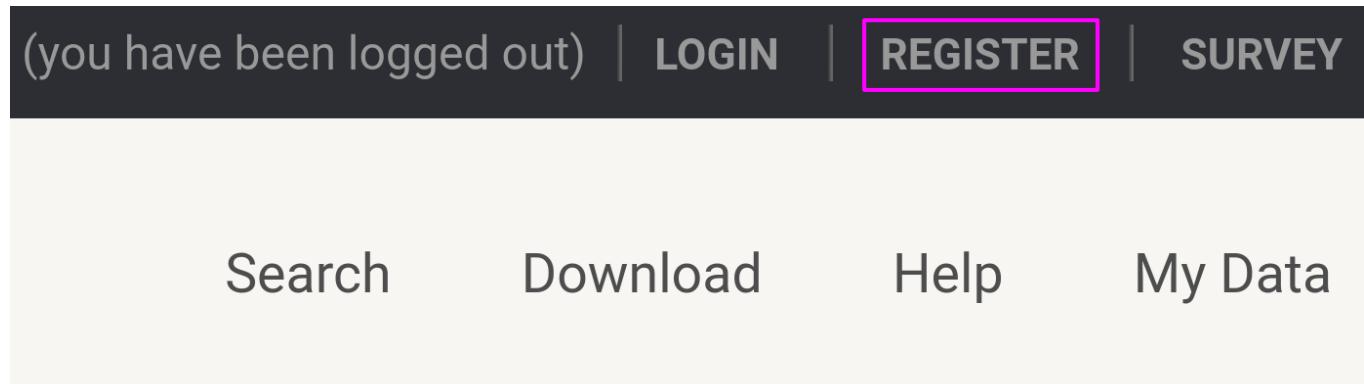
More tools for GO Term Analysis



GO Term analysis for
ChIP-seq → Extra Slides

There are multiple ways to
perform GO Term analysis!

GO Term analysis with STRING



Task: Create STRING account

Use links in the Handout/ Chat.

GO Term analysis with STRING

VERA.LAUB@GMX.DE | LOGOUT | SURVEY

Search Download Help My Data

click to view uploaded data

Your browsing history ➤

Your stored gene sets ▾

reference	names	genes	db version	last updated				
intersection_ChIPSeeker_ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363_q5_RNA-seq_BMAL1del_10142_2020_757_MOESM4_ESM	169	161	11.5	2022-05-09, 13:38:06	show	analyze	edit	delete
downregulated_genes_Pbx1+E2akd_aNS	23	22	11.5	2022-04-11, 08:52:49	show	analyze	edit	delete
upregulated_genes_Pbx1+E2akd_aNS	118	115	11.5	2022-04-11, 08:50:03	show	analyze	edit	delete
downregulated_genes_E2Akd_aNS	198	194	11.5	2022-03-14, 12:24:39	show	analyze	edit	delete
upregulated_genes_E2Akd_aNS	431	420	11.5	2022-03-14, 12:23:32	show	analyze	edit	delete
diff.regulated_genes_E2Akd_aNS	630	615	11.5	2022-03-14, 10:24:35	show	analyze	edit	delete
downregulated_genes_Pbx1kd_aNS	121	118	11.0	2021-07-13, 14:33:55	archived			delete
upregulated_genes_Pbx1kd_aNS	369	365	11.0	2021-07-13, 12:22:44	archived			delete
ChIP-Seq_Pbx1_aNS_annotation_VM23	25314	14869	11.0	2021-03-09, 17:07:49	archived			delete
ChIPseeker_PBX1_foldEnrichment>100Peaks	728	581	11.0	2021-03-09, 16:12:09	archived			delete
add new geneset								delete all

click to add new dataset

Create a New Gene Set

1) name your new set ...

name for future reference:

intersection_ChipSeeker_Chip-seq_GliomaStemCells_ARNTL_hg38_GSM3

choose name for new dataset

Create a New Gene Set

1) name your new set ...

name for future reference:

intersection_ChipSeeker_Chip-seq_GliomaStemCells_ARNTL_hg38_GSM3

2) identify your organism ...

Homo sapiens

choose organism of dataset

Create a New Gene Set

1) name your new set ...

name for future reference:

intersection_ChipSeeker_Chip-seq_GliomaStemCells_ARNTL_hg38_GSM3

2) identify your organism ...

Homo sapiens



3) provide identifiers ...

list of gene/protein names:

(one identifier per line; examples: #1 #2 #3)

HECW1
MAP3K14
JARID2
TUBG2
INPP4A
ARID1B
FOYNR

**insert dataset
(various formats possible)**

Create a New Gene Set

1) name your new set ...

name for future reference:

intersection_ChipSeeker_Chip-seq_GliomaStemCells_ARNTL_hg38_GSM3

2) identify your organism ...

Homo sapiens



3) provide identifiers ...

list of gene/protein names:

HECW1
MAP3K14
JARID2
TUBG2
INPP4A
ARID1B
FOYNR

click to continue

<- BACK

CONTINUE ->

2) please review the mappings ...

1) 'HECW1':

- HECW1 - E3 ubiquitin-protein ligase HECW1; E3 ubiquitin-protein ligase that mediates ubiquitination and subsequent degradation of DVL1. Also targets the...

2) 'MAP3K14':

- MAP3K14 - Mitogen-activated protein kinase kinase kinase 14; Lymphotoxin beta-activated kinase which seems to be exclusively involved in the activation...
- TRAF1 - TNF receptor-associated factor 1; Adapter molecule that regulates the activation of NF- kappa-B and JNK. Plays a role in the regulation of cell sur...
- IKBKAP - Elongator complex protein 1; May act as a scaffold protein that may assemble active IKK-MAP3K14 complexes (IKKA, IKKB and MAP3K14/NIK); ...

review gene mappings

Create a New Gene Set

info

Your geneset has been uploaded.

OK

Practical Session

Task: Upload intersecting genes
(ChIPseeker, RNA-seq)

GO Term analysis with STRING

Your stored gene sets ▾

reference	names	genes	db version	last updated	show	analyze	edit	delete
intersection_ChIPSeeker_ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363_q5_RNA-seq_ARNTL-deletion	169	161	11.5	2022-05-11, 08:33:21	show	analyze	edit	delete

click to perform
GO Term analysis

Geneset Analysis

Interaction Stats at Default Settings

number of nodes: 161

number of edges: 163

average node degree: 2.02

avg. local clustering coefficient: 0.372

expected number of edges: 84

PPI enrichment p-value: 2.25e-14

*your network has significantly more interactions
than expected (what does that mean?)*

[click for more information](#)

Geneset Analysis

Interaction Stats at Default Settings

number of nodes: 161
number of edges: 163
average node degree: 2.02
avg. local clustering coefficient: 0.372

expected number of edges: 84
PPI enrichment p-value: 2.25e-14

your network has significantly more interactions than expected (what does that mean?)

click for more information

Interaction Enrichment

This means that your proteins have more interactions among themselves than what would be expected for a random set of proteins of the same size and degree distribution drawn from the genome. Such an enrichment indicates that the proteins are at least partially biologically connected, as a group.

Functional enrichments in your set of proteins

[explain columns](#)

Biological Process (Gene Ontology)					
GO-term	description	count in network	strength	false discovery rate	
GO:0045780	Positive regulation of bone resorption	4 of 19	1.41	0.0077	
GO:0045124	Regulation of bone resorption	5 of 43	1.15	0.0084	
GO:0071526	Semaphorin-plexin signaling pathway	4 of 37	1.12	0.0366	
GO:0106027	Neuron projection organization	4 of 39	1.1	0.0410	
GO:0002763	Positive regulation of myeloid leukocyte differentiation	5 of 58	1.02	0.0223	
GO:0030516	Regulation of axon extension	7 of 99	0.93	0.0067	
GO:0030279	Negative regulation of ossification	5 of 75	0.91	0.0458	
GO:0048814	Regulation of dendrite morphogenesis	6 of 95	0.89	0.0235	
GO:0050770	Regulation of axonogenesis	11 of 187	0.85	0.00070	

[click for more information](#)

regulation of axonogenesis

Term Information

Accession GO:0050770

Data health 

Name regulation of axonogenesis

Ontology biological_process

Synonyms None

Alternate IDs None

Definition Any process that modulates the frequency, rate or extent of axonogenesis, the generation of an axon, the long process of a neuron. Source: GOC:ai

Comment None

History See term [history for GO:0050770](#) at QuickGO

Subset None

Related [Link](#) to all **genes and gene products** annotated to regulation of axonogenesis (**excluding "regulates"**).

[Link](#) to all direct and indirect **annotations** to regulation of axonogenesis (**excluding "regulates"**).

[Link](#) to all direct and indirect **annotations download** (limited to first 10,000) for regulation of axonogenesis (**excluding "regulates"**).

[Include "regulates"](#)

For more information, please see the [ontology relation documentation](#).

KEGG Pathways

<i>pathway</i>	<i>description</i>	<i>count in network</i>	<i>strength</i>	<i>false discovery rate</i>
hsa05225	Hepatocellular carcinoma	8 of 160	0.78	0.0245

several more GO
enrichments available

Save / Export

Biological Process (Gene Ontology)	download	126 GO-terms significantly enriched; file-format: tab-delimited
Molecular Function (Gene Ontology)	download	7 GO-terms significantly enriched; file-format: tab-delimited
Cellular Component (Gene Ontology)	download	9 GO-terms significantly enriched; file-format: tab-delimited
KEGG Pathways	download	one single pathway is enriched; file-format: tab-delimited
WikiPathways	download	7 pathways significantly enriched; file-format: tab-delimited
Subcellular localization (COMPARTMENTS)	download	9 compartments significantly enriched; file-format: tab-delimited
Annotated Keywords (UniProt)	download	13 keywords significantly enriched; file-format: tab-delimited
Protein Domains (Pfam)	download	one single domain is enriched; file-format: tab-delimited
Protein Domains and Features (InterPro)	download	5 domains significantly enriched; file-format: tab-delimited
Protein Domains (SMART)	download	one single domain is enriched; file-format: tab-delimited
All enriched terms	download	179 enriched terms in 10 categories; file-format: tab-delimited

select to download

Task: Perform GO Term analysis with STRING

Breakout room 1: Anna, Finn, Tobias, Rebecca

Breakout room 2: Sara, Miles, Felix

What differences do you find in comparison to PANTHER analysis?

Which problems did you encounter?

What can we learn about biological assemblages by GO Term analysis?

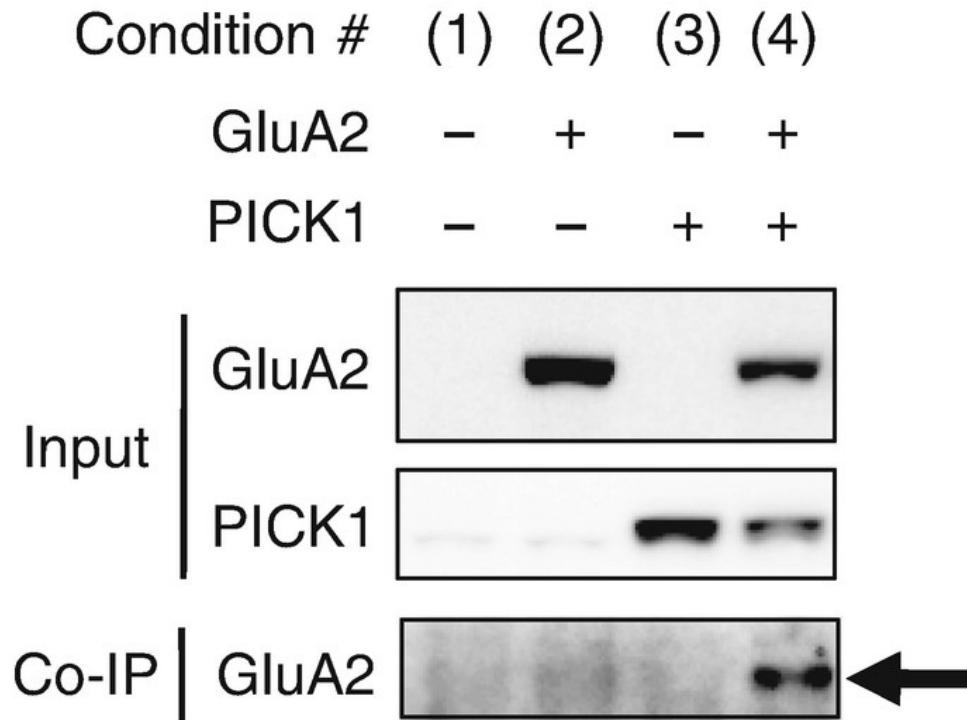
Protein-Protein Interaction Networks

The networks of protein complexes formed by biochemical events and/or electrostatic forces and that serve a **distinct biological function as a complex.**

Protein-Protein Interaction Networks

In vitro

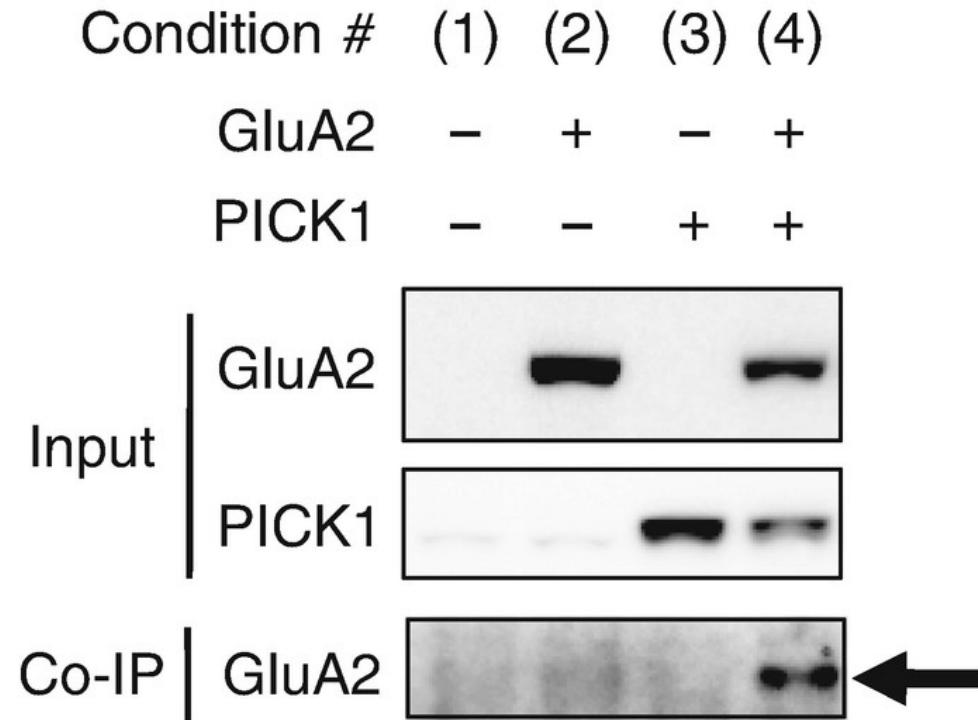
- Co-Immunoprecipitation



Protein-Protein Interaction Networks

In vitro

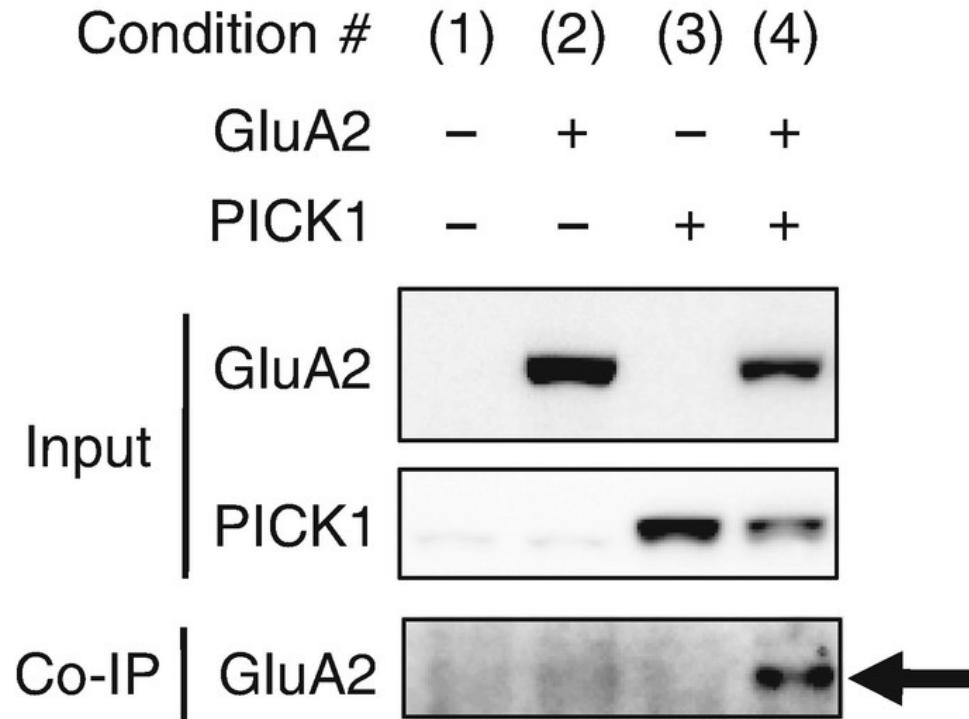
- Co-Immunoprecipitation
- Pull-down assays



Protein-Protein Interaction Networks

In vitro

- Co-Immunoprecipitation
- Pull-down assays
- Mass-Spectrometry



IP: anti-PICK1
IB: anti-GluA2

Protein-Protein Interaction Networks

In silico: STRING

SEARCH

Single Protein by Name / Identifier

Protein Name:

(examples: #1 #2 #3)

ARNTL

Organism:

Homo sapiens



[Advanced Settings](#)

SEARCH

Protein-Protein Interaction Networks

In silico: STRING

SEARCH

Single Protein by Name / Identifier

Protein Name: (examples: #1 #2 #3)

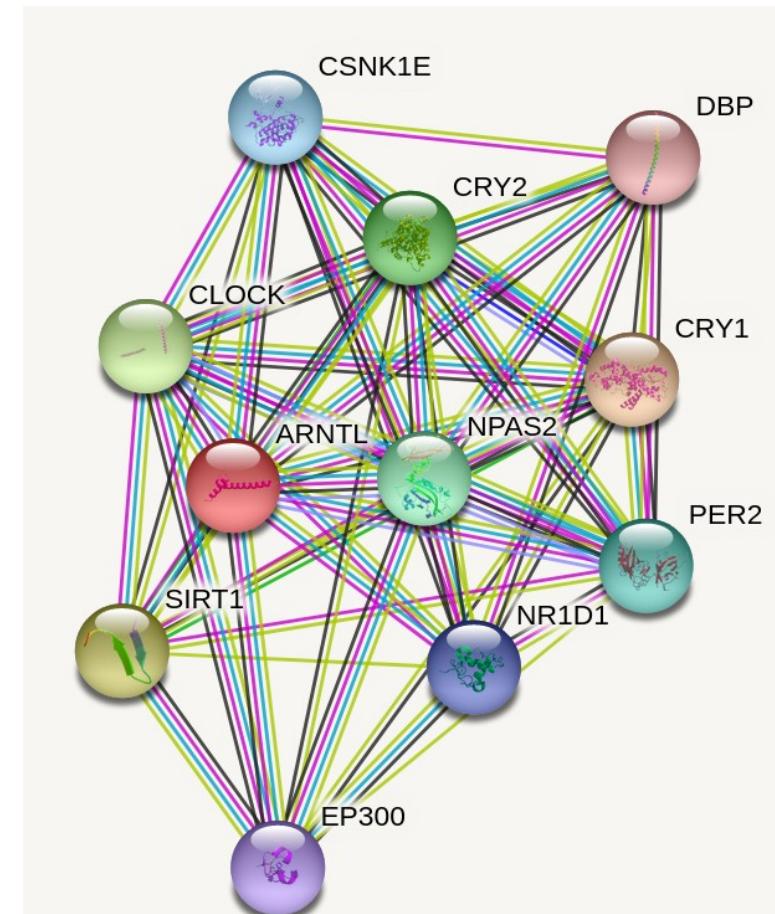
ARNTL

Organism:

Homo sapiens

[Advanced Settings](#)

SEARCH



Protein-Protein Interaction Networks

In silico: STRING

SEARCH

Single Protein by Name / Identifier

Protein Name: (examples: #1 #2 #3)

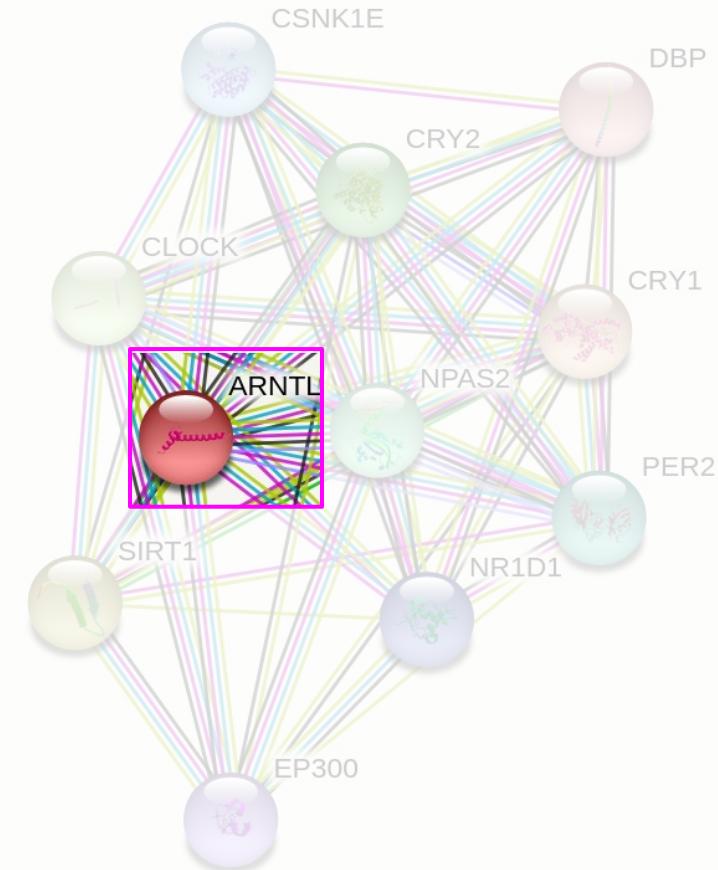
ARNTL

Organism:

Homo sapiens

[Advanced Settings](#)

SEARCH



Information

Aryl hydrocarbon receptor nuclear translocator-like protein 1; Transcriptional activator which forms a core component of the circadian clock. The circadian clock, an internal time-keeping system, regulates various physiological processes through the generation of approximately 24 hour circadian rhythms in gene expression, which are translated into rhythms in metabolism and behavior. It is derived from the Latin roots 'circa' (about) and 'diem' (day) and acts as an important regulator of a wide array of physiological functions including metabolism, sleep, body temperature, blood pressure [...]

Identifier: ENSP00000384517, ARNTL

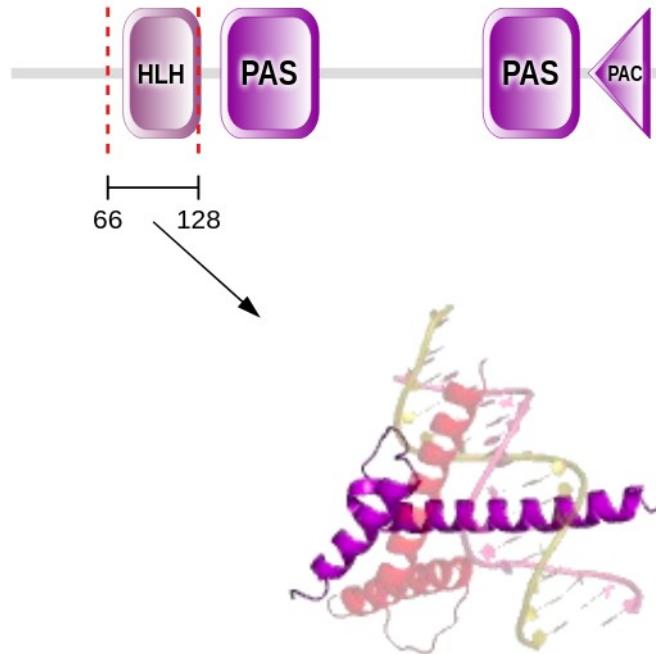
Organism: Homo sapiens



- show protein sequence
- homologs among STRING organisms
- Pathways, Functions, Resources (GeneCards)

Enable node color mode New

Show this node's terms in the analysis table



PDB
1 of 4
PDB structure (4h10)
identity: 100%

Protein-Protein Interaction Networks with STRING

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Protein-Protein Interaction Networks with STRING

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Data Sources

Interactions in STRING are derived from five main sources:



Genomic Context
Predictions



High-throughput Lab
Experiments



(Conserved) Co-
Expression



Automated
Textmining



Previous Knowledge in
Databases

Protein-Protein Interaction Networks with STRING

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Data Sources

Interactions in STRING are derived from five main sources:



Genomic Context
Predictions



High-throughput Lab
Experiments



(Conserved) Co-
Expression



Automated
Textmining



Previous Knowledge in
Databases

Coverage

The STRING database currently covers 24'584'628 proteins from 5'090 organisms.

Protein-Protein Interaction Networks with STRING

[Search](#)[Download](#)[Help](#)[My Data](#)

[click to view data](#)

Protein-Protein Interaction Networks with STRING



STRING

Search Download Help My Data

reference	names	genes	db version	last updated				
RNA-seq_ARNTL-deletion_allDEG	318	259	11.5	2022-05-11, 11:00:11	show	analyze	edit	delete
intersection_ChIPSeeker_ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363_q5_RNA-seq_ARNTL-deletion	169	161	11.5	2022-05-11, 08:33:21	show	analyze	edit	delete

click to view data

Protein-Protein Interaction Networks with STRING



STRING

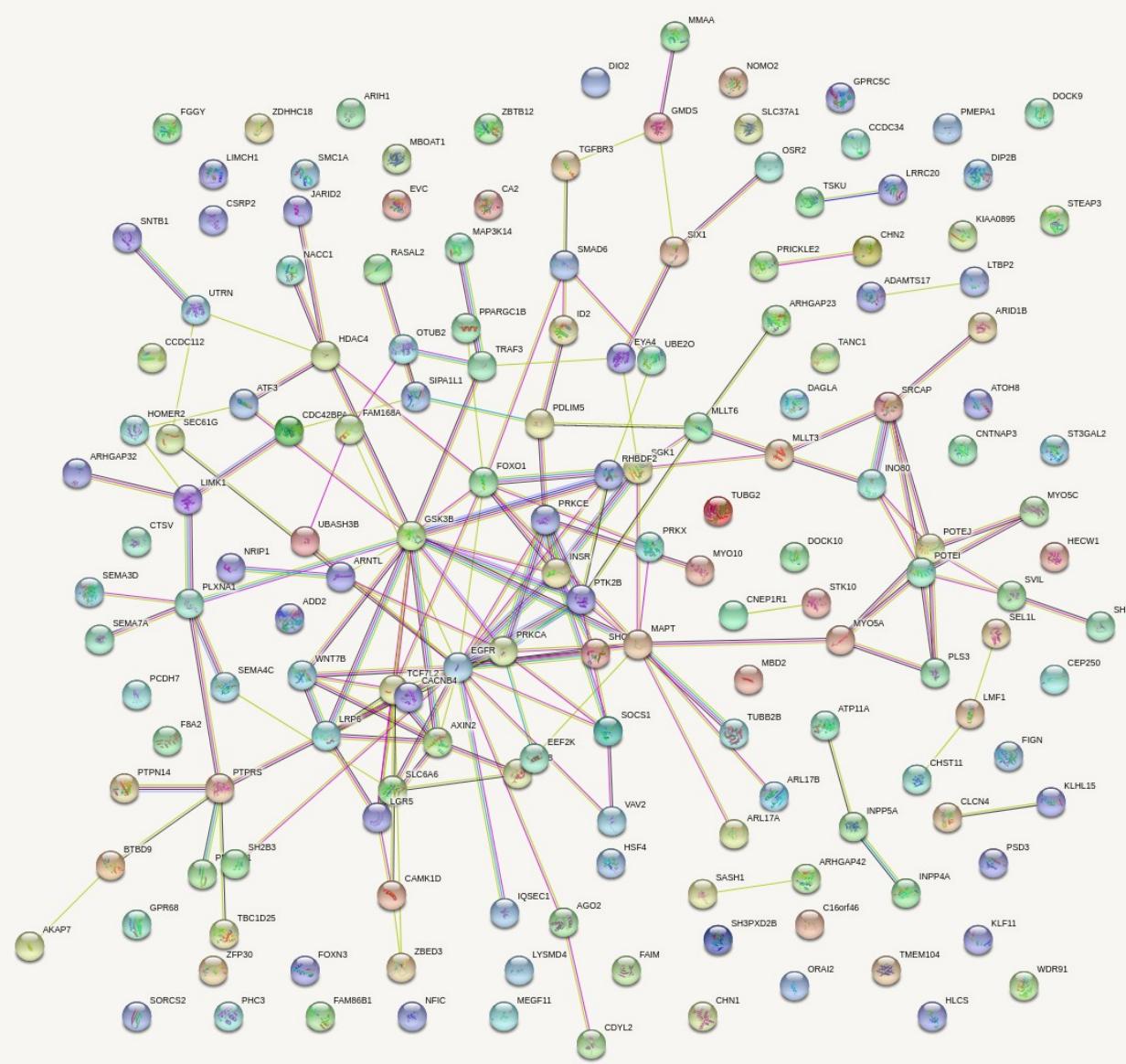
Search Download Help My Data

reference	names	genes	db version	last updated				
RNA-seq_ARNTL-deletion_allDEG	318	259	11.5	2022-05-11, 11:00:11	show	analyze	edit	delete
intersection_ChIPSeeker_ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363_q5_RNA-seq_ARNTL-deletion	169	161	11.5	2022-05-11, 08:33:21	show	analyze	edit	delete

reference:	"intersection_ChIPSeeker_ChIP-seq_GliomaStemCells_ARNTL_hg38_GSM3982997_SRX6602363_q5_RNA-seq_ARNTL-deletion"
organism:	Homo sapiens
distinct input items:	169
distinct genes mapped:	161
enable statistics use:	yes
last updated:	2022-05-11, 08:33:21

[VIEW AS INTERACTION NETWORK](#)

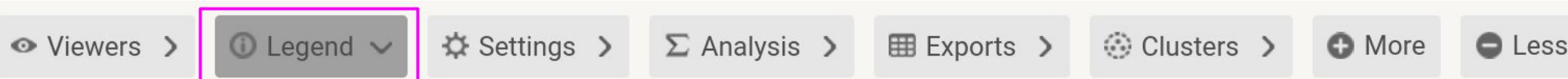
click to view
interaction network



**Eeeeehm....
WHAT?**

Customization of Network

click to view legend



Nodes:

Network nodes represent proteins

splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Color



*colored nodes:
query proteins and first shell of interactors*



*white nodes:
second shell of interactors*

Node Content



*empty nodes:
proteins of unknown 3D structure*



*filled nodes:
some 3D structure is known or predicted*

Edges:

Edges represent protein-protein associations

associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding to each other.

Known Interactions



from curated databases



experimentally determined

Predicted Interactions



gene neighborhood



gene fusions



gene co-occurrence

Others



textmining



co-expression



protein homology

click to view settings

Viewers >

Legend >

Settings ▾

Analysis >

Exports >

Clusters >

More

- Less

Network type:

- full STRING network (the edges indicate both functional and physical protein associations)
 - physical subnetwork (the edges indicate that the proteins are part of a physical complex)
-

meaning of network edges:

- evidence ( line color indicates the type of interaction evidence)
 - confidence ( line thickness indicates the strength of data support)
-

customize network

active interaction sources:

- Textmining
 - Experiments
 - Databases
 - Co-expression
 - Neighborhood
 - Gene Fusion
 - Co-occurrence
-

minimum required interaction score:

highest confidence (0.900) 

set interaction score

max number of interactors to show:

1st shell: no more than 10 interactors

2nd shell: - none -

set number of
interactors per shell

Advanced Settings

network display mode:



static png



(network is a simple bitmap image; not interactive)



interactive svg



(network is a scalable vector graphic [SVG]; interactive)

network display options:



disable structure previews inside network bubbles



hide disconnected nodes in the network



show input protein names



hide protein names

**change appearance
and exclude
unconnected nodes**

click to update
network

UPDATE

Advanced Settings

network display mode:



static png



(network is a simple bitmap image; not interactive)



interactive svg



(network is a scalable vector graphic [SVG]; interactive)

network display options:



disable structure previews inside network bubbles



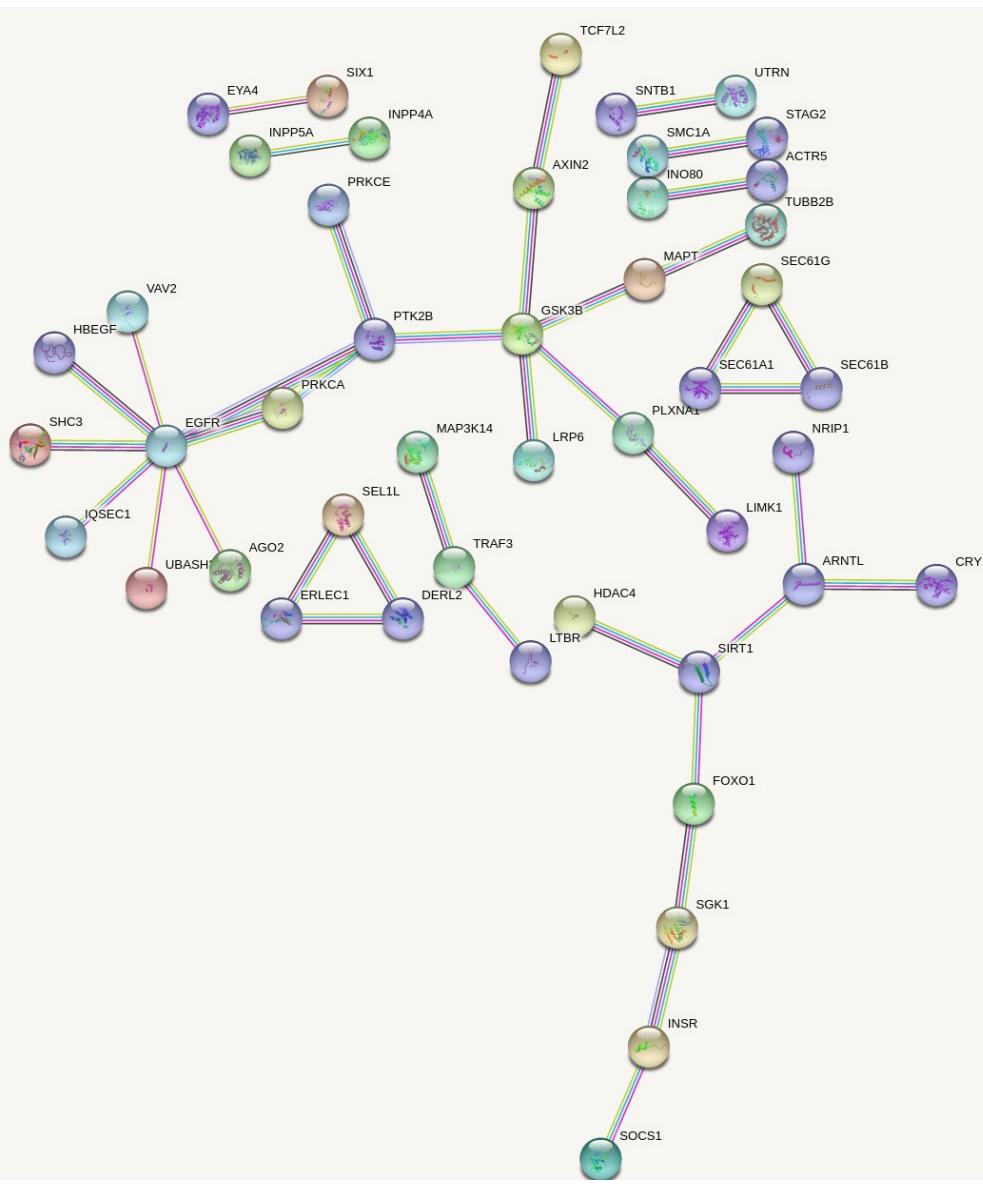
hide disconnected nodes in the network



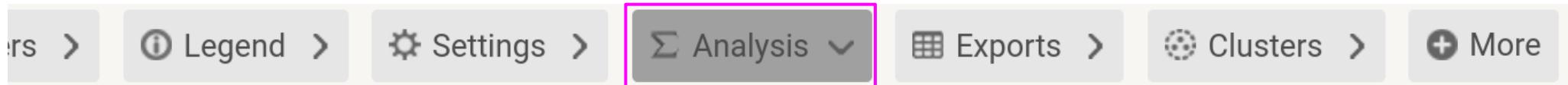
show input protein names



hide protein names



click to view GO Term analysis results



Network Stats

number of nodes: 171
number of edges: 39
average node degree: 0.456
avg. local clustering coefficient: 0.2

expected number of edges: 28
PPI enrichment p-value: 0.0271
your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

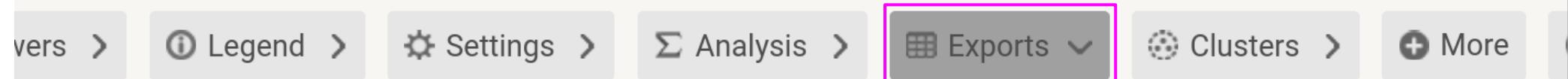
Note: some enrichments may be expected here (why?)

[explain columns](#)

Biological Process (Gene Ontology)					
GO-term	description	count in network	strength	false discovery rate	
GO:0031204	Posttranslational protein targeting to membrane, translocati...	3 of 7	1.69	0.0088	
GO:0030970	Retrograde protein transport, er to cytosol	4 of 18	1.41	0.0056	
GO:0070914	UV-damage excision repair	3 of 14	1.39	0.0315	
GO:0045780	Positive regulation of bone resorption	4 of 19	1.38	0.0060	
GO:0045124	Regulation of bone resorption	5 of 43	1.12	0.0074	

(more ...)

click to export



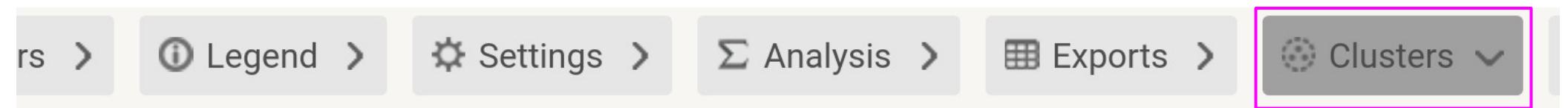
Export your current network:



[Send network to Cytoscape](#)

- ... as a bitmap image: [download](#) file format is 'PNG': portable network graphic
- ... as a high-resolution bitmap: [download](#) same PNG format, but at higher resolution
- ... as a vector graphic: [download](#) SVG: scalable vector graphic - can be opened and edited in Illustrator, CorelDraw, Dia, etc
- ... as short tabular text output: [download](#) TSV: tab separated values - can be opened in Excel and Cytoscape (lists only one-way edges: A-B)
- ... as tabular text output: [download](#) TSV: tab separated values - can be opened in Excel (lists reciprocal edges: A-B,B-A)
- ... as an XML summary: [download](#) structured XML interaction data, according to the 'PSI-MI' data standard
- ... protein node degrees: [download](#) node degree of proteins in your network (given the current score cut-off)
- ... network coordinates: [download](#) a flat-file format describing the coordinates and colors of nodes in the network
- ... protein sequences: [download](#) MFA: multi-fasta format - containing the aminoacid sequences in the network
- ... protein annotations: [download](#) a tab-delimited file describing the names, domains and descriptions of proteins in your network
- ... functional annotations: [download](#) a tab-delimited file containing all known functional terms of proteins in your network

click to perform cluster analysis



Clustering Options



no clustering (network is shown as it is)



kmeans clustering (network is clustered to a specified number of clusters)

number of clusters:

3



edges between clusters:

Dotted line



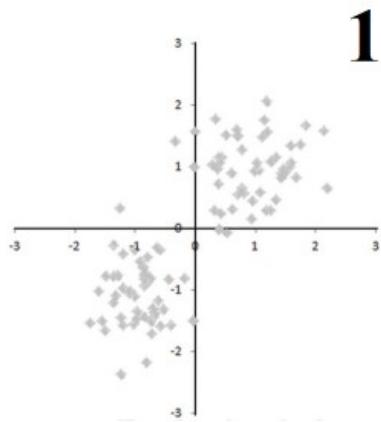
MCL clustering

(network is clustered to a specified "MCL inflation parameter")

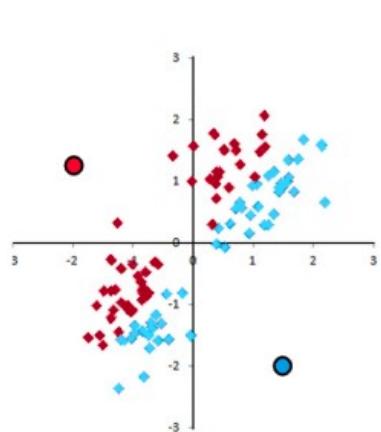
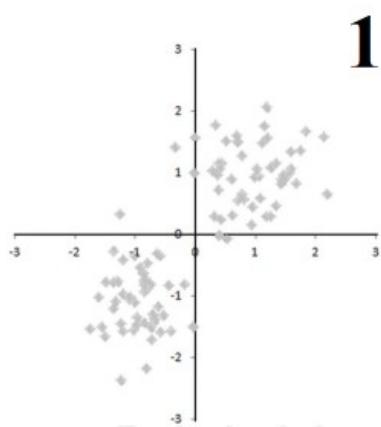
APPLY

customize clustering parameters

Kmeans Clustering

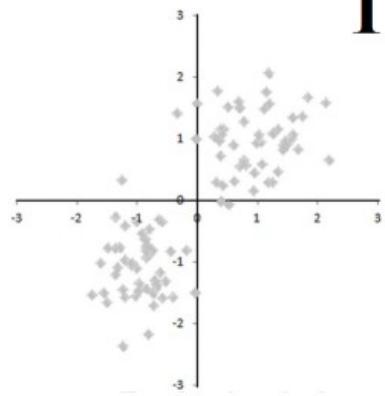


Kmeans Clustering

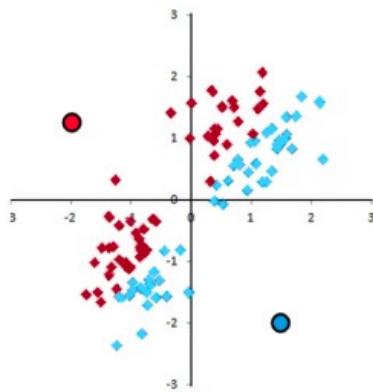


Kmeans Clustering

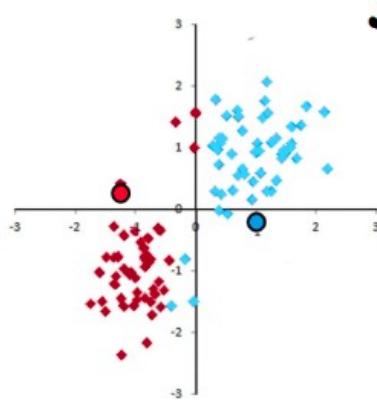
1



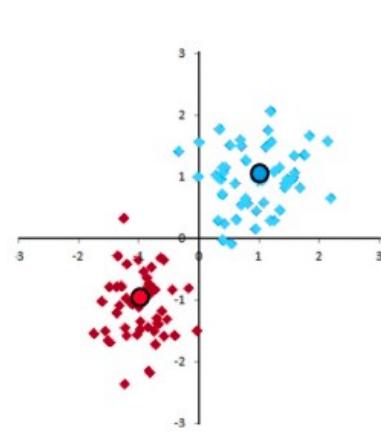
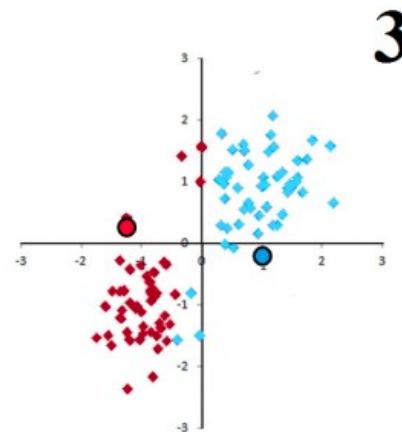
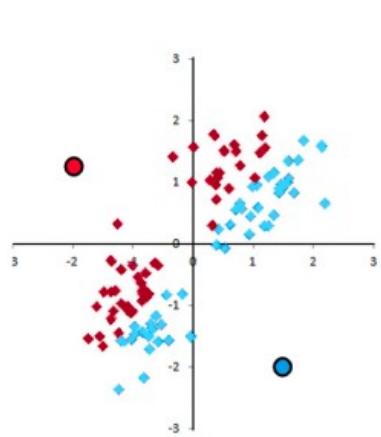
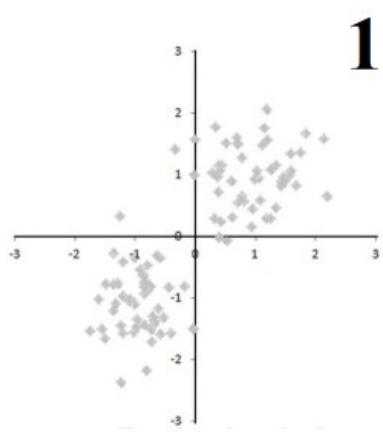
2



3



Kmeans Clustering



Clustering Options



no clustering (network is shown as it is)



kmeans clustering (network is clustered to a specified number of clusters)

number of clusters:

3



edges between clusters:

Dotted line



MCL clustering

(network is clustered to a specified "MCL inflation parameter")

APPLY

Whether or not to perform clustering depends on your research question.
Only because you **CAN** does not necessarily mean that you **SHOULD** :)

Clustering Options



no clustering (network is shown as it is)



kmeans clustering (network is clustered to a specified number of clusters)

number of clusters:

3



edges between clusters:

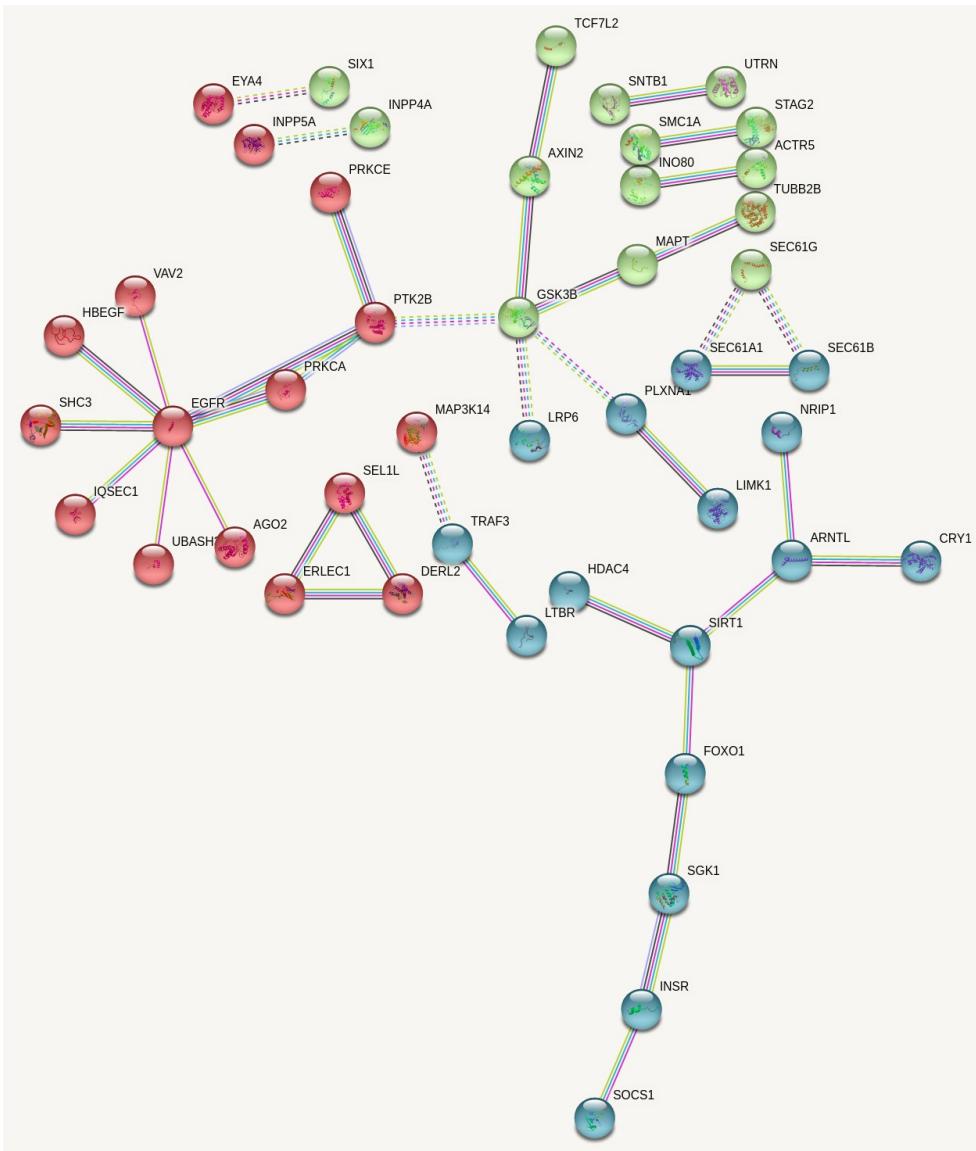
Dotted line



MCL clustering (network is clustered to a specified "MCL inflation parameter")

APPLY

apply clustering



Task: Perform Interaction Network analysis with STRING

Breakout room 1: Anna, Finn, Tobias

Breakout room 2: Miles, Rebecca, Sara, Felix

What can you learn about the intersection dataset?

Which problems did you encounter?

Which problems did you encounter?

What can you learn about the intersection dataset?

Which problems did you encounter?

What can you learn about the intersection dataset?

What wetlab experiments could follow?

Final questions?

Thank you for your
attention!



Practical Training Session: Task

1. Retrieve ChIP-seq dataset relevant for your project
2. Visualize in IGV
3. Perform Peak Annotation with ChIPSeeker
4. Retrieve RNA-seq dataset for relevant context
5. Intersect genes affected by ChIP-seq and RNA-seq
6. Perform GO Term analysis and network analysis of commonly affected genes

Extra Slides

GO Term Analysis with GREAT

GREAT: Genomic Regions Enrichment of Annotations Tool

GREAT predicts functions of *cis*-regulatory regions.

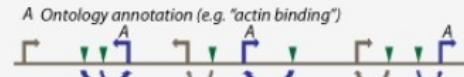
1. **Input:** A set of Genomic Regions (such as transcription factor binding events identified by ChIP-Seq).



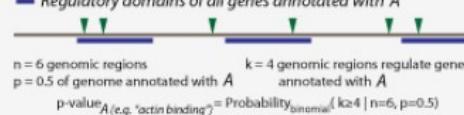
2. GREAT associates both proximal and distal input Genomic Regions with their putative target genes.



3. GREAT uses gene Annotations from numerous ontologies to associate genomic regions with annotations.



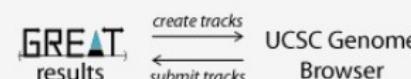
4. GREAT calculates statistical Enrichments for associations between Genomic Regions and Annotations.



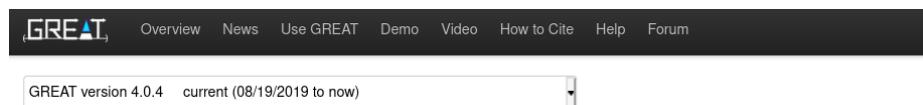
5. **Output:** Annotation terms that are significantly associated with the set of input Genomic Regions.

Ontology term	p-value
Actin cytoskeleton	10^{-9}
FOS gene family	10^{-8}
TRAIL signaling	10^{-7}

6. Users can create UCSC custom tracks from term-enriched subsets of Genomic Regions. Any track can be directly submitted to GREAT from the UCSC Table Browser.



- **GO analysis tool** provided by Bejerano lab, Stanford University
- **coding genes are well annotated with their biological functions, non-coding regions lack such annotation**
→ GREAT assigns biological meaning to a set of non-coding genomic regions by analyzing the annotations of the **nearby genes**



GREAT predicts functions of cis-regulatory regions.

GO Term Analysis with GREAT

Species Assembly

- Human: GRCh38 (UCSC hg38, Dec. 2013)
- Human: GRCh37 (UCSC hg19, Feb. 2009)
- Mouse: GRCm38 (UCSC mm10, Dec. 2011)
- Mouse: NCBI build 37 (UCSC mm9, Jul. 2007)

[Can I use a different species or assembly?](#)

select the correct assembly of your input data

Test regions

- BED file: [Browse...](#) bedtool...NS.bed
- BED data:

browse the data (in .bed format) you want to perform GO term analysis with...

... or copy/paste

Background regions

- Whole genome
- BED file: [Browse...](#) No file selected.
- BED data:

use whole genome as background

When should I use a background set?
What should my background regions file contain?

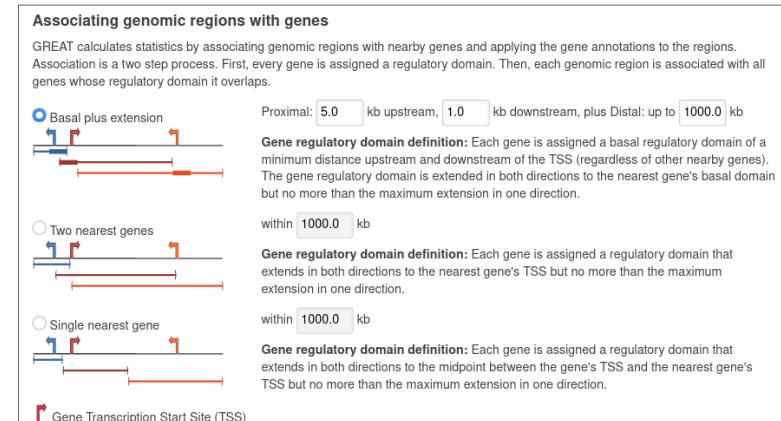
Association rule settings

[Show settings »](#)

Submit

Reset

leave rest at default and submit



GO Term Analysis with GREAT

If extended .bed formats are used as input, you might encounter this error:

GREAT encountered a user error.

GREAT encountered an error in your test set "bedtools_intersect_Chip-Seq_CTCF_ESC_Pbx1_aNS.bed".
Please modify your data and re-submit it. If you need to report this error, please include this id: 20210303-public-4.0.4-ID8Pv2

Error details:

Line 1: The sixth field (strand) must be either '+', '-', or '.' (not '3672159'). The seventh field (thickStart) must be a non-negative integer (not peak_5). Input data: 'chr1 3671454 3671616 chr1 3670408 3672159 peak_5 162'
Line 2: The sixth field (strand) must be either '+', '-', or '.' (not '4808425'). The seventh field (thickStart) must be a non-negative integer (not peak_12). Input data: 'chr1 4808051 4808182 chr1 4807422 4808425 peak_12 131'
Line 3: The sixth field (strand) must be either '+', '-', or '.' (not '5019013'). The seventh field (thickStart) must be a non-negative integer (not peak_20). Input data: 'chr1 5017923 5018174 chr1 5017817 5019013 peak_20 251'
Line 4: The sixth field (strand) must be either '+', '-', or '.' (not '5070611'). The seventh field (thickStart) must be a non-negative integer (not peak_26). Input data: 'chr1 5070300 5070499 chr1 5070276 5070611 peak_26 199'
Line 5: The sixth field (strand) must be either '+', '-', or '.' (not '6272662'). The seventh field (thickStart) must be a non-negative integer (not peak_38). Input data: 'chr1 6272282 6272575 chr1 6272157 6272662 peak_38 293'
Line 6: The sixth field (strand) must be either '+', '-', or '.' (not '6383295'). The seventh field (thickStart) must be a non-negative integer (not peak_41). Input data: 'chr1 6382820 6382986 chr1 6382576 6383295 peak_41 166'
Line 7: The sixth field (strand) must be either '+', '-', or '.' (not '6406647'). The seventh field (thickStart) must be a non-negative integer (not peak_42). Input data: 'chr1 6406421 6406630 chr1 6406152 6406647 peak_42 209'
Line 8: The sixth field (strand) must be either '+', '-', or '.' (not '6410393'). The seventh field (thickStart) must be a non-negative integer (not peak_43). Input data: 'chr1 6410194 6410358 chr1 6409817 6410393 peak_43 164'
Line 9: The sixth field (strand) must be either '+', '-', or '.' (not '6484052'). The seventh field (thickStart) must be a non-negative integer (not peak_45). Input data: 'chr1 6483370 6483539 chr1 6483029 6484052 peak_45 169'
Line 10: The sixth field (strand) must be either '+', '-' or '' (not '7398597'). The seventh field (thickStart) must be a non-negative integer (not

- software varies in definition of columns 4-12 in .bed files
- for GO term analysis we only need information in columns 1-3 (chromosome, peak start, peak end)
 - to circumvent this problem, remove all additional columns of your input file with text editor/ spreadsheet and resubmit to GREAT

GO Term Analysis with GREAT

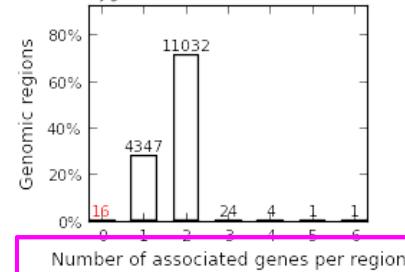
+ Region-Gene Association Graphs

What do these graphs illustrate?

Number of associated genes per region

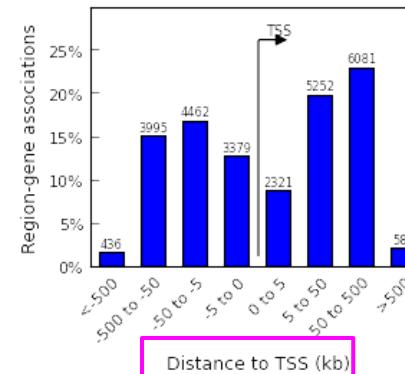
Download as PDF.

Genomic regions associated with one or more genes
Genomic regions not associated with any genes



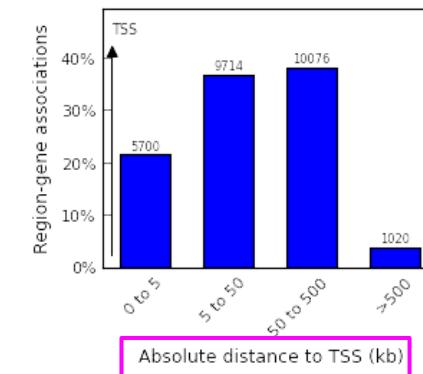
Binned by orientation and distance to TSS

Download as PDF.



Binned by absolute distance to TSS

Download as PDF.



click to download as .pdf (alternative: screenshot, but worse quality/ resolution than .pdf)

+ Global Controls

Global Export

Which data is exported by each option?

click to expand/collapse

GO Term Analysis with GREAT

Global Controls Global Export Which data is exported by each option?

Ontologies GO Phenotype Genes

Show Non-Empty GO Molecular Function Mouse Phenotype Ensembl Genes
Expand All GO Biological Process Mouse Phenotype Single KO
Collapse All GO Cellular Component Human Phenotype

Click a link above to jump to the corresponding table.

Top Rows/Table: 20 Set Default Minimum Region-based Fold Enrichment: 2 Set Default Observed Gene Hits: 1 Set Default Term Name Filter: Where is my term? Filter Clear

Term Annotation Count: Min: 1 Max: 1000 Set Default What is the term annotation count?

Statistical Significance: Raw P-Value Bonferroni FDR Threshold: 0.05 Set Default Significant P-Values are bold in the tables below. Terms are bold when significant by both binomial and hypergeometric tests. Which statistics does GREAT provide?

View: Significant By Both Significant By Region-based Binomial Ignore statistical significance Which view should I use?

Reset view

check/uncheck to customize represented features in table

General Binomial over Regions Hypergeometric over Genes

Show All Ontology Term ID

Rank Rank
Raw P-Value Raw P-Value
Bonferroni P-Value Bonferroni P-Value
FDR Q-Value FDR Q-Value
Fold Enrichment (Obs/Exp) Fold Enrichment (Obs/Exp)
Expected (n * p) Expected (n * K/N)
Observed Region Hits (k) Observed Gene Hits (k)
Genome Fraction (p) Total Genes (K)
Region Set Coverage (k/n) Gene Set Coverage (k/K)
Term Gene Coverage (k/K)

Click on a link above to sort all tables by the named column. What do each of the columns mean?

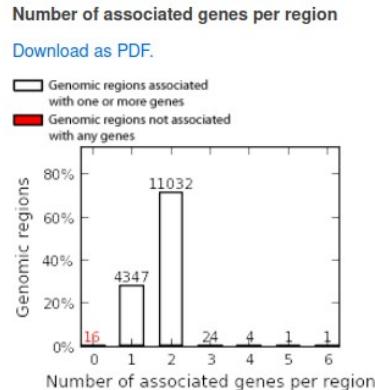
click to reset to default view

click for more information

GO Term Analysis with GREAT

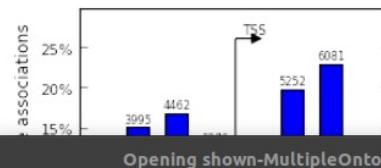
+ Region-Gene Association Graphs

What do these graphs illustrate?



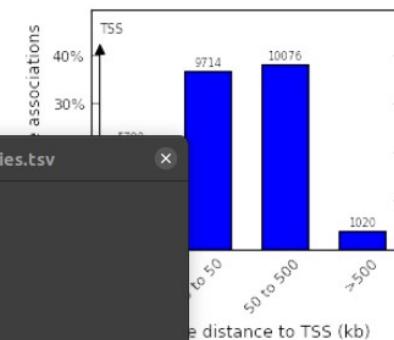
Binned by orientation and distance to TSS

Download as PDF.



Binned by absolute distance to TSS

Download as PDF.



+ Global Controls

Global Export

Which data is exported?

+ Ensembl Genes ()

Global Export

All data as .tsv

+ GO Biological Processes

All shown data as HTML

All shown data as .tsv

Table controls: Export

View all region-gene associations

You have chosen to open:

shown-MultipleOntologies.tsv

which is: plain text document
from: <http://great.stanford.edu>

What should Firefox do with this file?

Open with Text Editor (default)

Save File

Do this automatically for files like this from now on.

Cancel

OK

click to download global analysis export file

GO Term Analysis with GREAT

GO Biological Process (19 terms)													Global controls
Term Name	Binom Rank	Binom Raw P-Value		Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Enrichment	Visualize this table:		
		Set	Min: 1 Max: Inf								[select one]	[select one]	Bar chart of current sorted value
regulation of mitochondrial membrane permeability	73	4.9864e-23	8.9414e-21	2.7699	127	0.82%	811	5.5964e-5	1.5165	46	53	0.38%	
regulation of membrane permeability	85	1.9256e-21	2.9654e-19	2.4994	141	0.91%	736	1.5351e-5	1.5143	52	60	0.42%	
heterochromatin organization	189	6.3422e-13	4.3926e-11	3.4138	49	0.32%	1,640	1.3664e-2	1.6380	15	16	0.12%	
neutral amino acid transport	266	4.0572e-10	1.9966e-8	2.4256	64	0.41%	1,292	3.1419e-3	1.4976	30	35	0.24%	
regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway	276	6.4365e-10	3.0527e-8	2.8490	47	0.30%	1,738	2.0642e-2	1.5633	17	19	0.14%	
regulation of toll-like receptor signaling pathway	299	2.2306e-9	9.7654e-8	2.0352	86	0.56%	1,727	1.8689e-2	1.3343	42	55	0.34%	
positive regulation of membrane permeability	319	4.3869e-9	1.8001e-7	2.6754	47	0.30%	1,738	2.0642e-2	1.5633	17	19	0.14%	
positive regulation of mitochondrial membrane permeability	370	2.3971e-8	8.4806e-7	2.5869	45	0.29%	1,866	3.0330e-2	1.5531	16	18	0.13%	
negative regulation of Notch signaling pathway	374	2.7880e-8	9.7582e-7	2.0456	73	0.47%	806	5.1621e-5	1.6414	31	33	0.25%	
activation of MAPKKK activity	446	2.1196e-7	6.2209e-6	2.1238	58	0.38%	1,519	8.9260e-3	1.6445	16	17	0.13%	
positive regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway	546	1.1737e-6	2.8138e-5	2.9665	27	0.18%	1,828	2.6962e-2	1.7472	10	10	0.08%	
glutamate metabolic process	584	2.7910e-6	6.2559e-5	2.0361	52	0.34%	1,771	2.1870e-2	1.4784	22	26	0.18%	
translational elongation	661	8.1444e-6	1.6129e-4	2.0533	46	0.30%	1,710	1.7839e-2	1.4560	25	30	0.20%	

click to visualize as bar chart

comprehensive overview table for associated GO terms

GO Term Analysis with GREAT

Table Visualization

Job Description and Directions [hide]

Job ID: 20210309-public-4.0.4-EHkCRr

Display name: bedtools_intersect_ChIP-Seq_CTCF_ESC_Pbx1_aNS.bed

This visualization generates a bar chart based on a single ontology-specific table from a GREAT job. The metric plotted is Binomial p value, which is what this ontology-specific table was sorted by in the previous page.

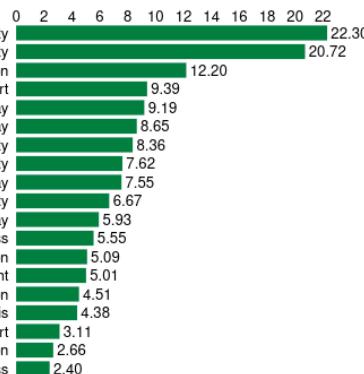
Use the controls at the bottom of the page to edit the table. Terms with long names may be truncated.

Export (will open a pop-up window): [SVG](#) | [PNG](#) | [PDF](#) | [PostScript](#)

Job ID: 20210309-public-4.0.4-EHkCRr
Display name: bedtools_intersect_ChIP-Seq_CTCF_ESC_Pbx1_aNS.bed

GO Biological Process

-log10(Binomial p value)



click to save in appropriate format

Chart title:

Bar height:

Font size:

Bar width:

Axis Guidelines: [Show axis guidelines](#)

Label length: [Full length](#)

Value labels: [Hide value labels](#)

Bar color:

make changes to customize graph

GO Term Analysis with GREAT

GO Biological Process (19 terms)													Global controls		
Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Enrichment	Export	Set	Term annotation count: Min: 1 Max: Inf	Visualize this table:	[select one]	[select one]
		▲	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
regulation of mitochondrial membrane permeability	73	4.9864e-23	8.9414e-21	2.7699	127	0.82%	811	5.5964e-5	1.5165	46	53	0.38%	20	Set	Bar chart of current sorted value
regulation of membrane permeability	85	1.9256e-21	2.9654e-19	2.4994	141	0.91%	736	1.5351e-5	1.5143	52	60	0.42%	Inf	Set	Visualize shown terms in hierarchy
heterochromatin organization	189	6.3422e-13	4.3926e-11	3.4138	49	0.32%	1,640	1.3664e-2	1.6380	15	16	0.12%			
neutral amino acid transport	266	4.0572e-10	1.9966e-8	2.4256	64	0.41%	1,292	3.1419e-3	1.4976	30	35	0.24%			
regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway	276	6.4365e-10	3.0527e-8	2.8490	47	0.30%	1,738	2.0642e-2	1.5633	17	19	0.14%			
regulation of toll-like receptor signaling pathway	299	2.2306e-9	9.7654e-8	2.0352	86	0.56%	1,727	1.8689e-2	1.3343	42	55	0.34%			
positive regulation of membrane permeability	319	4.3869e-9	1.8001e-7	2.6754	47	0.30%	1,738	2.0642e-2	1.5633	17	19	0.14%			
positive regulation of mitochondrial membrane permeability	370	2.3971e-8	8.4806e-7	2.5869	45	0.29%	1,866	3.0330e-2	1.5531	16	18	0.13%			
negative regulation of Notch signaling pathway	374	2.7880e-8	9.7582e-7	2.0456	73	0.47%	806	5.1621e-5	1.6414	31	33	0.25%			
activation of MAPKKK activity	446	2.1196e-7	6.2209e-6	2.1238	58	0.38%	1,519	8.9260e-3	1.6445	16	17	0.13%			
positive regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway	546	1.1737e-6	2.8138e-5	2.9665	27	0.18%	1,828	2.6962e-2	1.7472	10	10	0.08%			
glutamate metabolic process	584	2.7910e-6	6.2559e-5	2.0361	52	0.34%	1,771	2.1870e-2	1.4784	22	26	0.18%			
translational elongation	661	8.1444e-6	1.6129e-4	2.0533	46	0.30%	1,710	1.7839e-2	1.4560	25	30	0.20%			

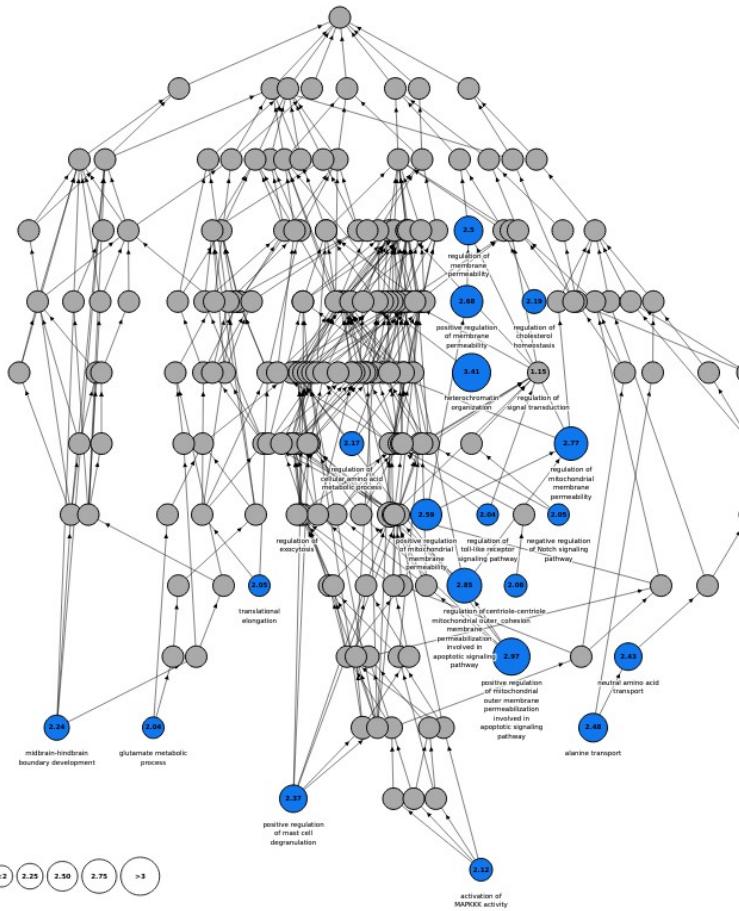
click to view in ontology visualizer

GO Term Analysis with GREAT

job ID: 20210310-public-4-D4-0f&CRr
Display name: bedtools_intersect_CtCF_Seq_ESC_Pas1_wMS.bed

Local DAG for enriched terms in GO Biological Process

Nodes sized according to Binomial Fold Enrichment



Directed Acyclic Graph = DAG

- based on the enriched terms from a single ontology-specific table from a GREAT job
 - enriched terms are shown in blue. Nodes have been sized according to Binomial Fold Enrichment.

Change min/max fold enrichment scale

Min val: 2

Max val: 3

customize graph (using information below)

- click and drag nodes to move them horizontally
 - hold "ctrl" and click a node to toggle whether the name/statistic of the term is shown. Initially only enriched term details are shown, but all nodes except for the root allow selection.
 - click once on a node to select it, which will display data in the table below
 - hold shift and click on multiple nodes to select many terms

GO Term Analysis with GREAT

click to view details
for GO term

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Enrichment	Bar chart of current sorted value	Visualize shown terms in hierarchy
regulation of mitochondrial membrane permeability	73	4.9864e-23	8.9414e-21	2.7699	127	0.82%	811	5.5964e-5	1.5165	46	53 0.38%
regulation of membrane permeability	85	1.9256e-21	2.9654e-19	2.4994	141	0.91%	736	1.5351e-5	1.5143	52	60 0.42%
heterochromatin organization	189	6.3422e-13	4.3926e-11	3.4138	49	0.32%	1,640	1.3664e-2	1.6380	15	16 0.12%
neutral amino acid transport	266	4.0572e-10	1.9966e-8	2.4256	64	0.41%	1,292	3.1419e-3	1.4976	30	35 0.24%
regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway	276	6.4365e-10	3.0527e-8	2.8490	47	0.30%	1,738	2.0642e-2	1.5633	17	19 0.14%
regulation of toll-like receptor signaling pathway	299	2.2306e-9	9.7654e-8	2.0352	86	0.56%	1,727	1.8689e-2	1.3343	42	55 0.34%
positive regulation of membrane permeability	319	4.3869e-9	1.8001e-7	2.6754	47	0.30%	1,738	2.0642e-2	1.5633	17	19 0.14%
positive regulation of mitochondrial membrane permeability	370	2.3971e-8	8.4806e-7	2.5869	45	0.29%	1,866	3.0330e-2	1.5531	16	18 0.13%
negative regulation of Notch signaling pathway	374	2.7880e-8	9.7582e-7	2.0456	73	0.47%	806	5.1621e-5	1.6414	31	33 0.25%
activation of MAPKKK activity	446	2.1196e-7	6.2209e-6	2.1238	58	0.38%	1,519	8.9260e-3	1.6445	16	17 0.13%
positive regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway	546	1.1737e-6	2.8138e-5	2.9665	27	0.18%	1,828	2.6962e-2	1.7472	10	10 0.08%
glutamate metabolic process	584	2.7910e-6	6.2559e-5	2.0361	52	0.34%	1,771	2.1870e-2	1.4784	22	26 0.18%
translational elongation	661	8.1444e-6	1.6129e-4	2.0533	46	0.30%	1,710	1.7839e-2	1.4560	25	30 0.20%

Summary for GO Biological Process analysis
(to be found underneath)

The test set of 15,425 genomic regions picked 12,245 (57%) of all 21,395 genes.

GO Biological Process has 13,090 terms covering 17,925 (84%) of all 21,395 genes, and 1,163,819 term - gene associations.

13,090 ontology terms (100%) were tested using an annotation count range of [1, Inf].

GO Term Analysis with GREAT

Term: heterochromatin organization (ID: GO:0070828) from [GO Biological Process](#)

Job ID: 20210303-public-4.0.4-irVHIA

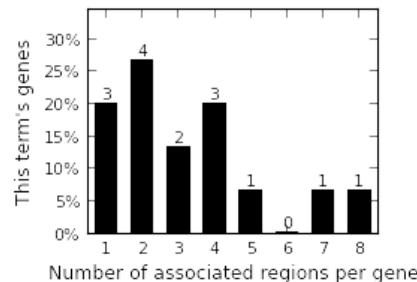
Display Name: bedtools_intersect_Chip-Seq_CTCF_ESC_Pbx1_aNS.bed

• This term's genomic region-gene association graphs

What do these graphs illustrate?

Number of associated regions per gene

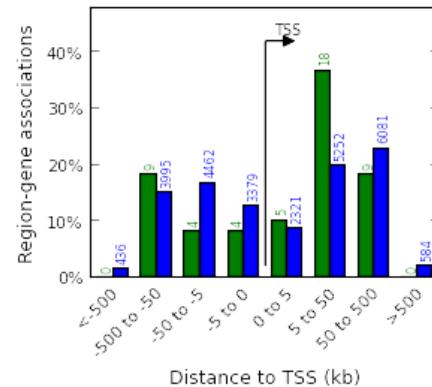
[Download as PDF.](#)



Binned by orientation and distance to TSS

This term's region-gene associations
Set-wide region-gene associations

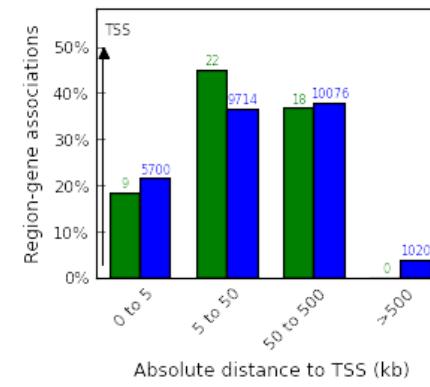
[Download as PDF.](#)



Binned by absolute distance to TSS

This term's region-gene associations
Set-wide region-gene associations

[Download as PDF.](#)



... some summary statistics in comparison to loci associated with other terms

GO Term Analysis with GREAT

• This term's genomic region-gene association tables (49 regions, 15 genes)

[What do these tables show?](#)

This term's genomic region -> gene association table

[Download table as text.](#)

Region	Gene (distance to TSS)
unnamed	H3f3a (+49,609)
unnamed	H3f3a (+44,924)
unnamed	Hmga2 (+219,002)
unnamed	Hmga2 (+205,449)
unnamed	Hmga2 (+87,930)
unnamed	Hmga2 (+44,136)
unnamed	Hmga2 (-800)
unnamed	Hmga2 (-190,740)
unnamed	Hmga2 (-225,602)
unnamed	Hmga2 (-396,246)
unnamed	Baz2a (-5,071)
unnamed	Baz2a (-579)
unnamed	Baz2a (+1,068)
unnamed	Baz2a (+5,112)
unnamed	H3f3b (+9,642)
unnamed	H3f3b (+4,000)
unnamed	Bahcc1 (-63,537)
unnamed	Bahcc1 (+7,512)
unnamed	Bahcc1 (+29,791)
unnamed	Bahcc1 (+47,466)
unnamed	Bahcc1 (+72,059)

This term's gene -> genomic region association table

[Download table as text.](#)

Gene	Region (distance to TSS)
Bahcc1	unnamed (-63,537), unnamed (+7,512), unnamed (+29,791), unnamed (+47,466), unnamed (+72,059), unnamed (+74,803), unnamed (+75,234)
Bahd1	unnamed (-17,922), unnamed (+192), unnamed (+1,349)
Baz2a	unnamed (-5,071), unnamed (-579), unnamed (+1,068), unnamed (+5,112)
Cdkn2a	unnamed (-120), unnamed (+12,355), unnamed (+148,102), unnamed (+149,032)
H3f3a	unnamed (+44,924), unnamed (+49,609)
H3f3b	unnamed (+4,000), unnamed (+9,642)
Hells	unnamed (+51,229)
Hmga1	unnamed (+1,202), unnamed (+9,206), unnamed (+21,097)
Hmga1-rs1	unnamed (+13,486)
Hmga2	unnamed (-396,246), unnamed (-225,602), unnamed (-190,740), unnamed (-800), unnamed (+44,136), unnamed (+87,930), unnamed (+205,449), unnamed (+219,002)
Hp1bp3	unnamed (-109,189), unnamed (-97,163)
Mthfr	unnamed (+19,351), unnamed (+24,796)
Setd7	unnamed (-87,566), unnamed (-22,899), unnamed (-589), unnamed (+20,895), unnamed (+49,793)
Tnrc18	unnamed (-61,921), unnamed (-59,862), unnamed (-24,858), unnamed (+49,521)
Wdhd1	unnamed (+17,688)

... some more details of genes associated with this term

→ can be used to identify candidate genes for further experiments

... some more details about GO term

• Additional ontology term information

If the content of the frame below is incorrect [report it](#) | [Why might the content below be incorrect?](#) | [open in a new tab](#)

AmiGO 2 Home Search ▾ Browse Tools & Resources Help Feedback About

Quick search Search

heterochromatin organization

Term Information

Accession GO:0070828
Name heterochromatin organization
Ontology biological_process
Synonyms heterochromatin organisation
Alternate IDs None
Definition Any process that results in the specification, formation or maintenance of the physical structure of eukaryotic heterochromatin, a compact and highly condensed form of chromatin. Source: GOC:mah
Comment None
History See term history for GO:0070828 at QuickGO
Subset None

GO Term Analysis with GREAT

• GO Cellular Component (3 terms)

Global controls

Table controls:

Export

Shown top rows in this table:

20

Set

Term annotation count: Min:

1

Max:

Inf

Set

Visualize this table:

[select one]

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
mitochondrial outer membrane translocase complex	109	2.8401e-8	4.4139e-7	3.9684	24	0.16%	245	4.5501e-2	1.7472	9	9	0.07%
endoplasmic reticulum subcompartment	137	4.2917e-7	5.3066e-6	2.2570	48	0.31%	248	4.4971e-2	1.4560	20	24	0.16%
Wnt signalosome	212	8.7208e-5	6.9684e-4	2.4004	25	0.16%	245	4.5501e-2	1.7472	9	9	0.07%

The test set of 15,425 genomic regions picked 12,245 (57%) of all 21,395 genes.

GO Cellular Component has 1,694 terms covering 19,074 (89%) of all 21,395 genes, and 371,380 term - gene associations.

1,694 ontology terms (100%) were tested using an annotation count range of [1, Inf].

• GO Molecular Function (4 terms)

Global controls

• Human Phenotype (7 terms)

Global controls

• Mouse Phenotype Single KO (18 terms)

Global controls

• Mouse Phenotype (20+ terms)

Global controls

... further terms and phenotypes associated with the input regions one might assess using GREAT

Alpha Fold

- online tool to visualize predicted protein structure

The screenshot shows the homepage of the AlphaFold Protein Structure Database. The background is a dark blue gradient with a faint, stylized protein structure outline. The title "AlphaFold Protein Structure Database" is prominently displayed in large white font in the center. Below it, the text "Developed by DeepMind and EMBL-EBI" is visible. At the top, there is a navigation bar with links for Home, About, FAQs, and Downloads. A search bar at the bottom left contains the text "ARNTL". To the right of the search bar are buttons for "BETA" and "Search". Below the search bar, there is a row of example protein IDs: Free fatty acid receptor 2, At1g58602, Q5VSL9, E. coli, and Help: AlphaFold DB search help. At the very bottom, there is a link for Feedback on structure: Contact DeepMind.

AlphaFold Protein Structure Database

Home About FAQs Downloads

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

ARNTL

BETA

Search

Examples: Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli Help: AlphaFold DB search help

Feedback on structure: Contact DeepMind

Alpha Fold

