



3000788 Intro to Comp Molec Biol

Week 9: Biological networks & chromatin organization

Fall 2024



Sira Sriswasdi, PhD

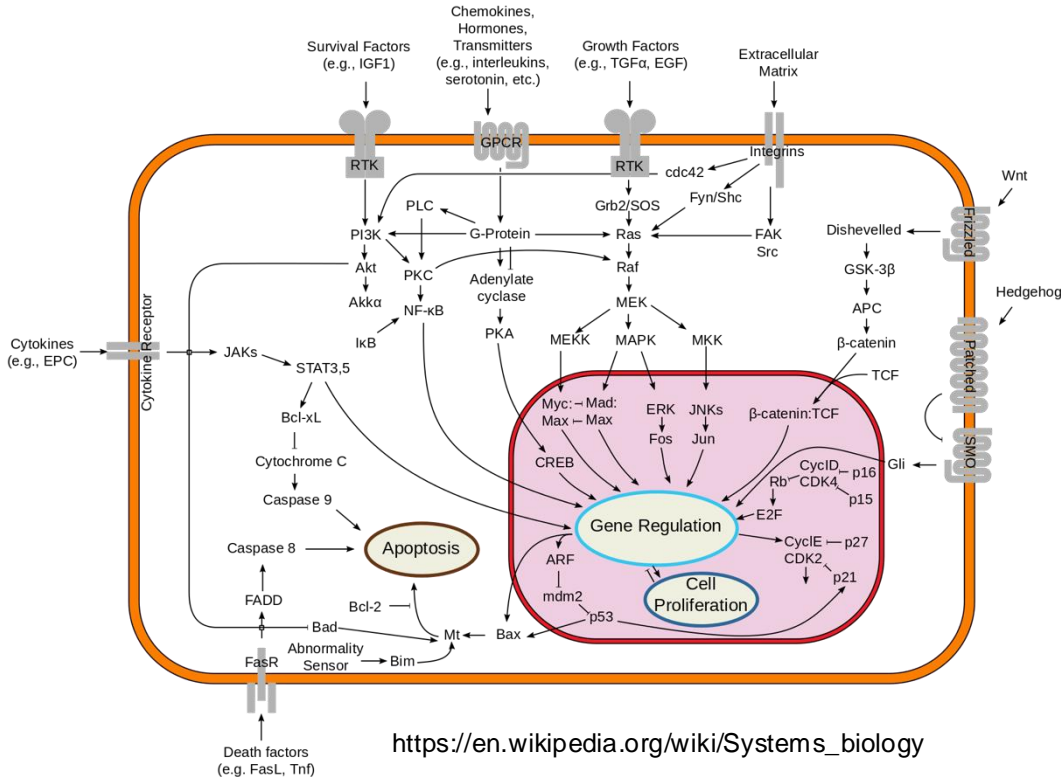
- Research Affairs
- Center of Excellence in Computational Molecular Biology (CMB)
- Center for Artificial Intelligence in Medicine (CU-AIM)

Part 1: Systems biology



- Biological systems consists of components and interactions
- Gene expression = snapshot of gene activity
- Interaction network = mechanistic diagram
- Systems biology = understanding the dynamics of the systems
 - Multi-omics
 - Model the time-dependent dynamics

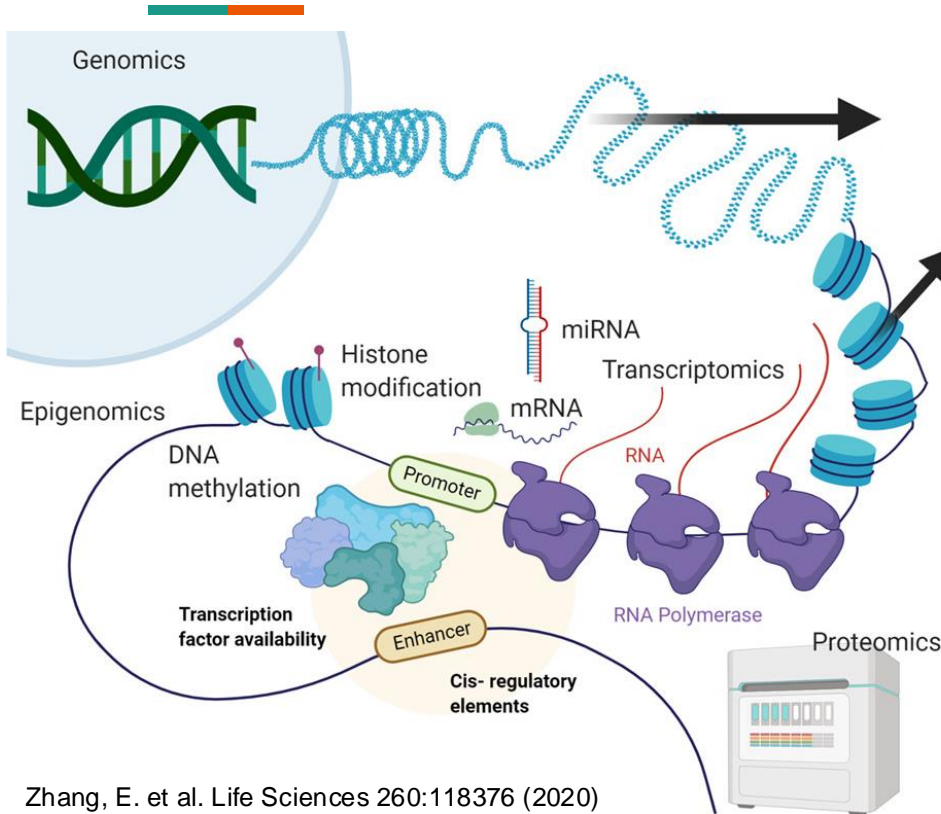
Systems biology



https://en.wikipedia.org/wiki/Systems_biology

- A (**biological**) system consists of components (**genes**) and rules (**gene expression regulations**) that control its characteristics (**phenotypes**)
- **Systems biology** = integration of data and model to fully understand the system

Multi-omics approach

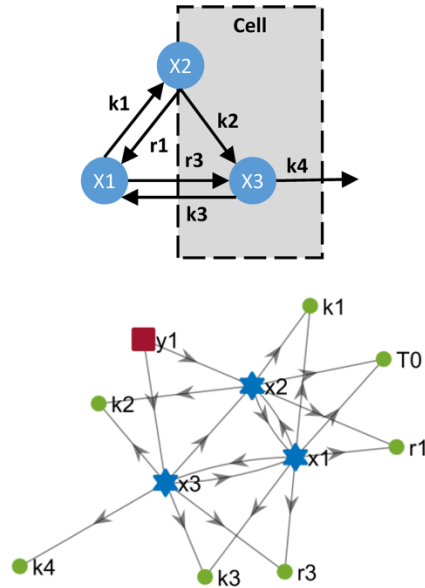


- Genomics → gene state
- Epigenomics → chromatin state
- Transcriptomics → gene expression
- Proteomics → protein expression
- Other assays → protein function
- **Provide mechanistic understanding**

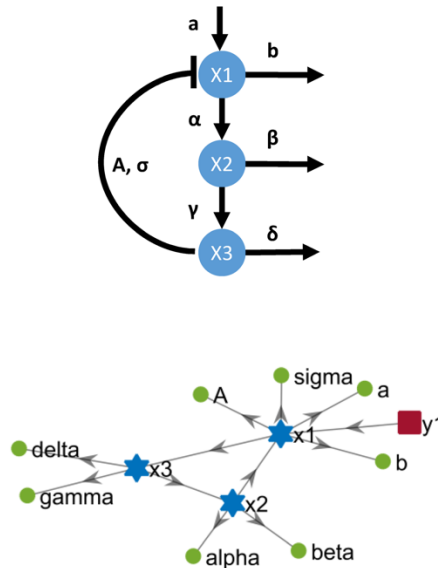
Modeling approach

Villaverde, A.F. PLoS Comp Biol 12:e1005153 (2016)

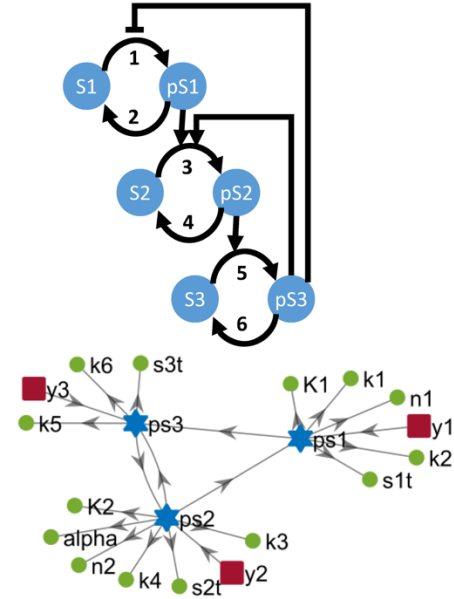
A Pitavastatin uptake



B Goodwin oscillator

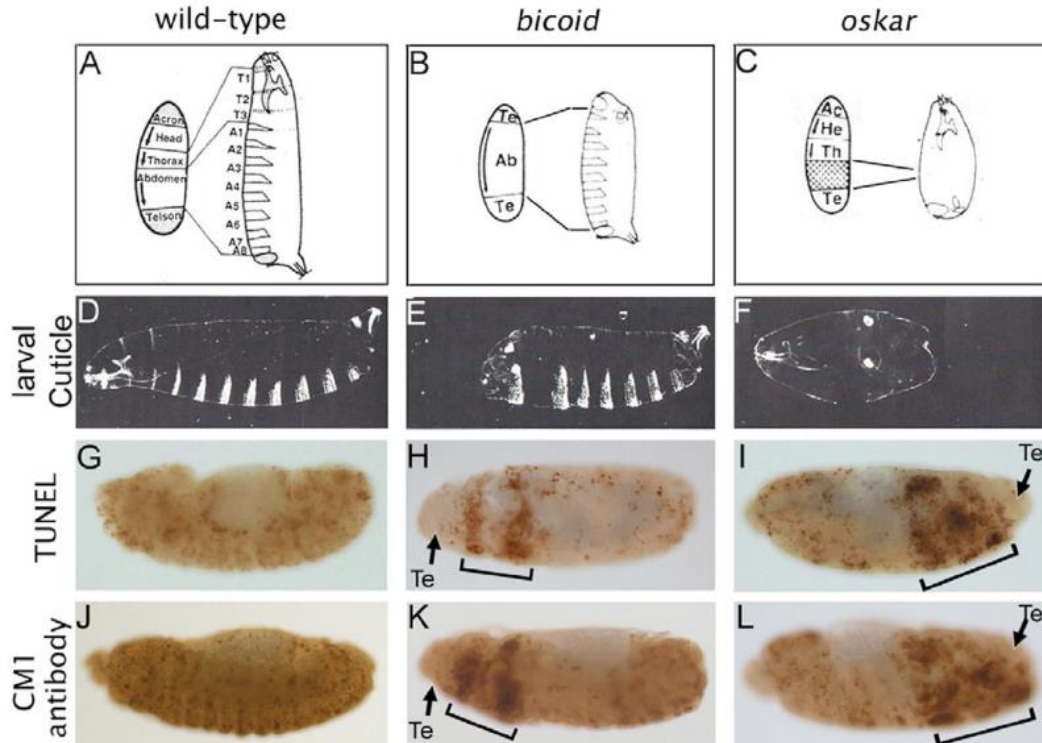


C MAPK cascade



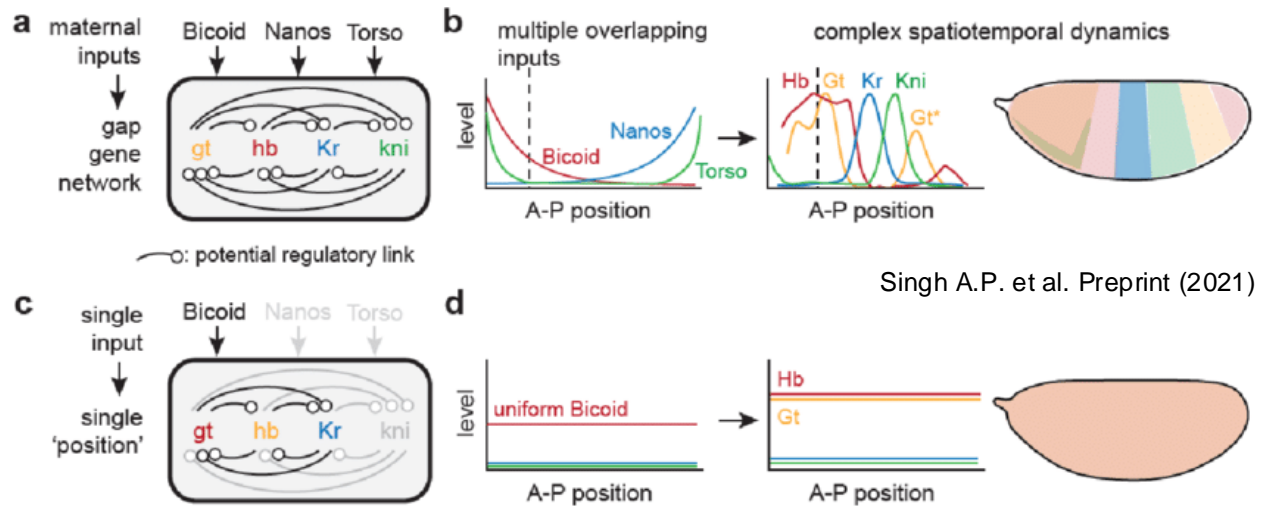
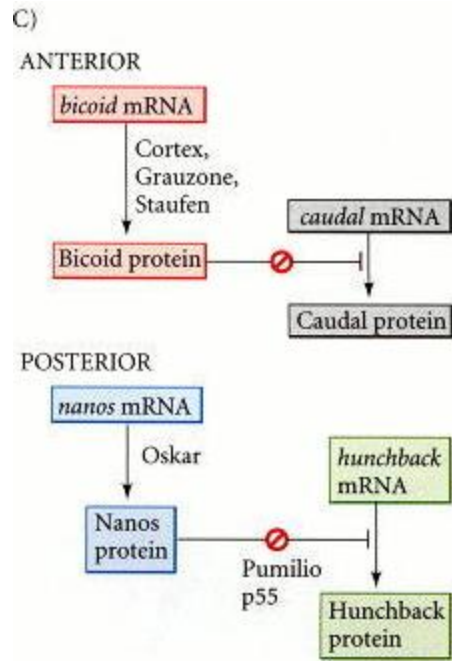
- Represent molecular interaction as **network of physical-chemical reactions**

Why is systems biology interesting?



- How does gene and protein know where to activate in the body during development?
- Each cell starts in identical state!

Why is systems biology interesting?



Singh A.P. et al. Preprint (2021)

- Gene regulatory dynamics lead to specific expression pattern over time

A simple enzymatic reaction

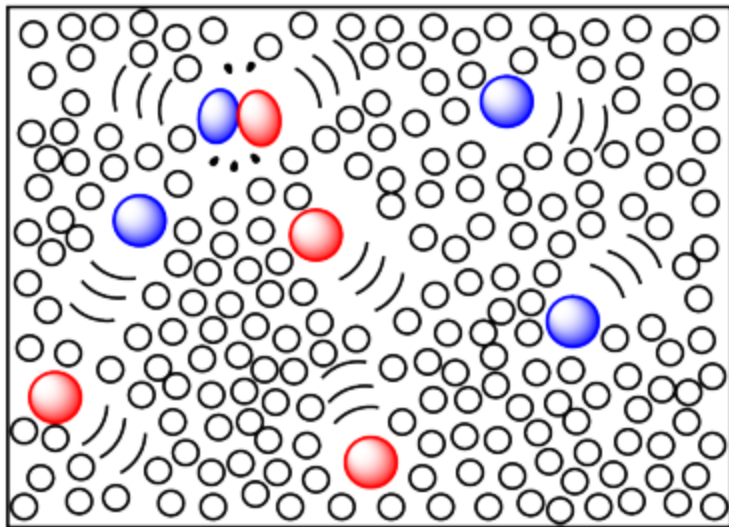


- $\frac{d[S]}{dt}$ = Rate of change in S = $-k_{\text{on}}[E][S] + k_{\text{off}}[ES]$
- $\frac{d[E]}{dt}$ = Rate of change in E = $-k_{\text{on}}[E][S] + k_{\text{off}}[ES] + k_{\text{react}}[ES]$
- $\frac{d[P]}{dt}$ = Rate of change in P = $k_{\text{react}}[ES]$

A key mental image

Rate of $E + S \rightarrow ES$ =
Rate of E meeting S x
Rate of binding

Rate of $E + S \rightarrow ES$ =
[E][S] x k



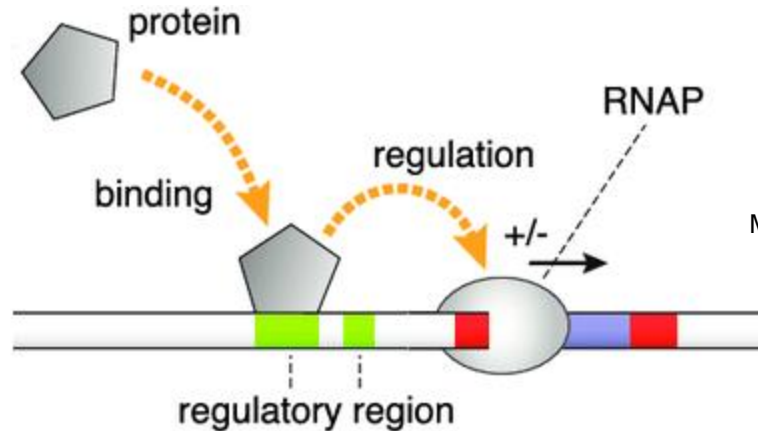
Rate of E meeting S
scales linearly with [E]
and [S]

Rate of binding is a
constant for E and S

<https://employees.csbsju.edu/cschaller/Reactivity/kinetics/rkphase.htm>

- Molecules must find each other in 3D to bind and interact
- Reaction takes time (and typically energy)

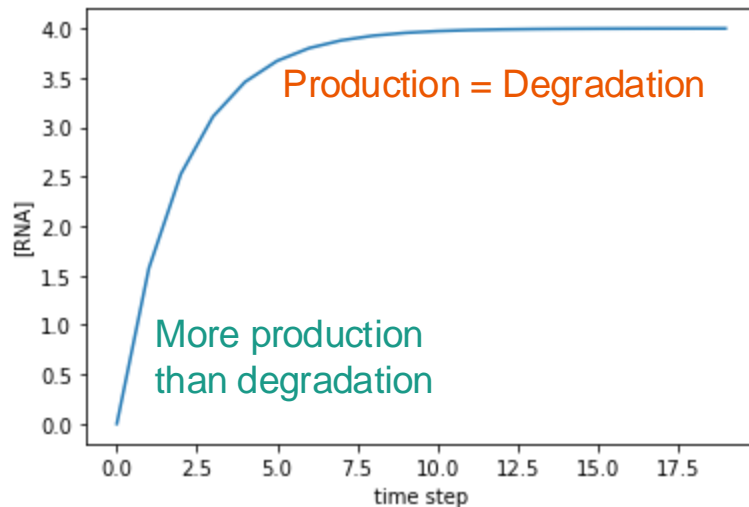
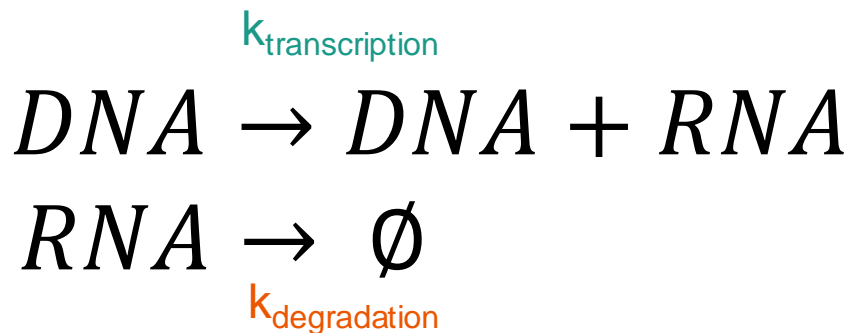
From enzymatic reaction to transcription



Marbach, D. et al. EvoBIO 2007

- $\text{TF} + \text{DNA}_{\text{inactive}} \leftrightarrow \text{TF-DNA}$
- $\text{RNAP} + \text{TF-DNA} \rightarrow \text{RNAP} + \text{TF-DNA} + \text{RNA}$
- $\text{RNA} \rightarrow \text{degraded RNA}$

A differential equation for gene expression



- $\frac{d[RNA]}{dt}$ = rate of change of RNA = $k_{\text{transcription}}$ - $k_{\text{degradation}}[RNA]$
- What would the graph of [RNA] look like?

Simulating differential equation

- $\frac{d[RNA]}{dt} = k_{transcription} - k_{degradation}[RNA]$
- Start at an initial condition $x(0) = x_0$,
 - Calculate $x(1) = x(0) + x'(0) = x_0 + k_{transcription} - k_{degradation}x_0$
 - Calculate $x(2) = x(1) + x'(1) = x(1) + k_{transcription} - k_{degradation}x(1)$
 - $= x_0 + 2 \cdot k_{transcription} - k_{degradation}(x_0 + x(1))$
 - [RNA] after 2 timesteps = initial [RNA] + 2 units of transcription - combined degradation of [RNA] at both timesteps

A toy example



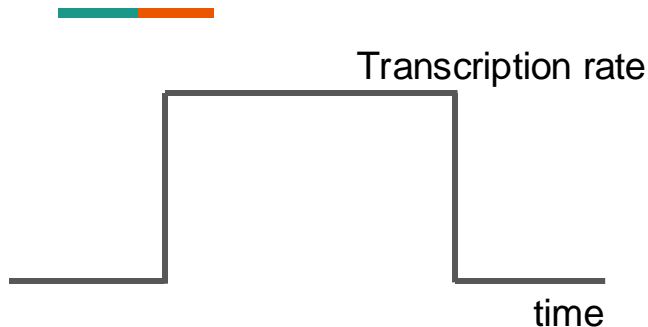
- $\frac{d[RNA]}{dt} = 4 - 0.2 [RNA]$
- Start at an initial condition $x(0) = 100$,
 - $x(1) = 100 + 4 - 20 = 84$
 - $x(2) = 84 + 4 - 17 = 71$
 - $x(3) = 71 + 4 - 14 = 61$
 - $x(4) = 61 + 4 - 12 = 53$
 - $x(5) = 53 + 4 - 10 = 47$

Another toy example



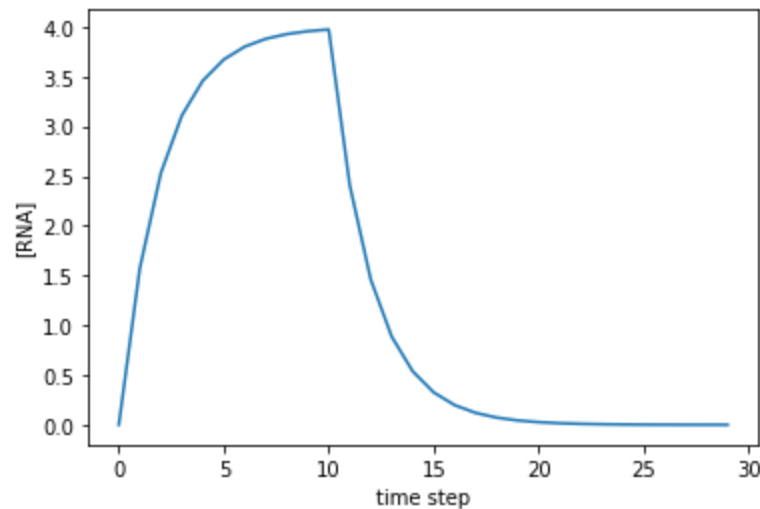
- $\frac{d[RNA]}{dt} = 4 - 0.2 [RNA]$
- Start at an initial condition $x(0) = 0$,
 - $x(1) = 0 + 4 - 0 = 4$
 - $x(2) = 4 + 4 - 1 = 7$
 - $x(3) = 7 + 4 - 1 = 10$
 - $x(4) = 10 + 4 - 2 = 12$
 - $x(5) = 12 + 4 - 2 = 14$

Time-dependent activation

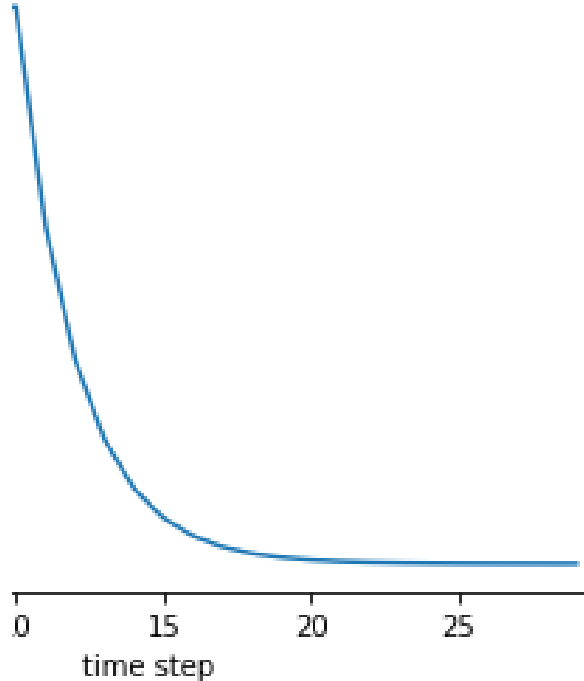


- First half of the dynamics is the same as before
- Second half is when RNA degrades until the [RNA] falls back to the basal level

```
1 def time_transcription(rna, time, k_trans, k_deg):  
2     if time < 10:  
3         return k_trans - k_deg * rna  
4     else:  
5         return - k_deg * rna
```

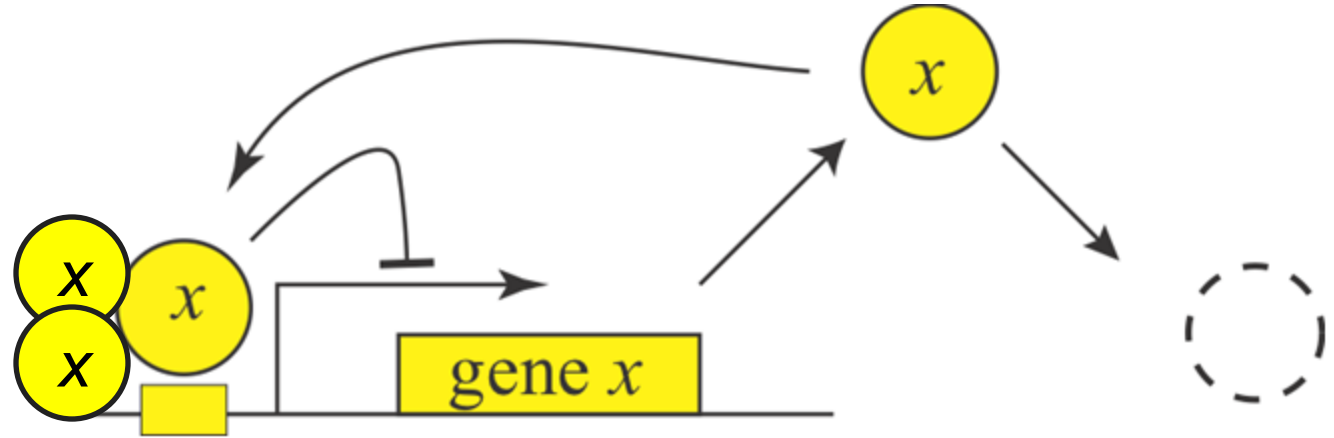


Exponential decay



- $\frac{d[RNA]}{dt} = -k_{\text{degradation}}[RNA]$
- Fast decay in the beginning because there are a lot of RNA molecules
- Slower decay at the end
- $\frac{de^{-kt}}{dt} = -ke^{-kt}$

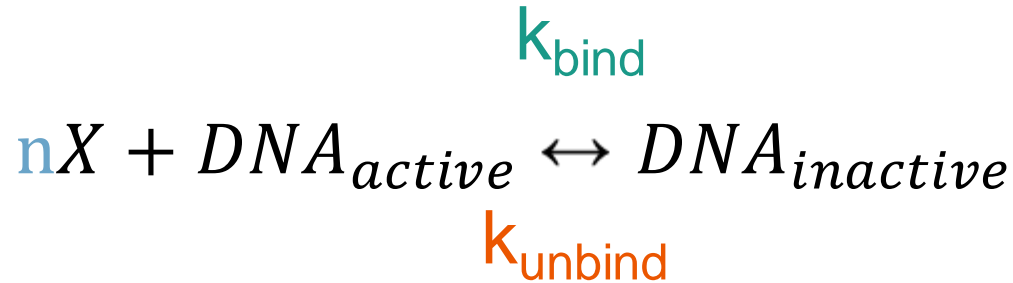
Negative auto-regulation



http://be150.caltech.edu/2019/handouts/03_small_circuits.html

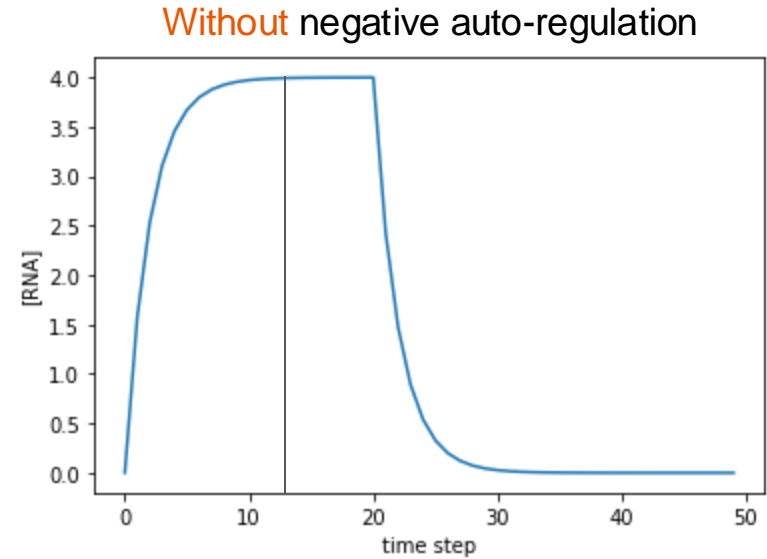
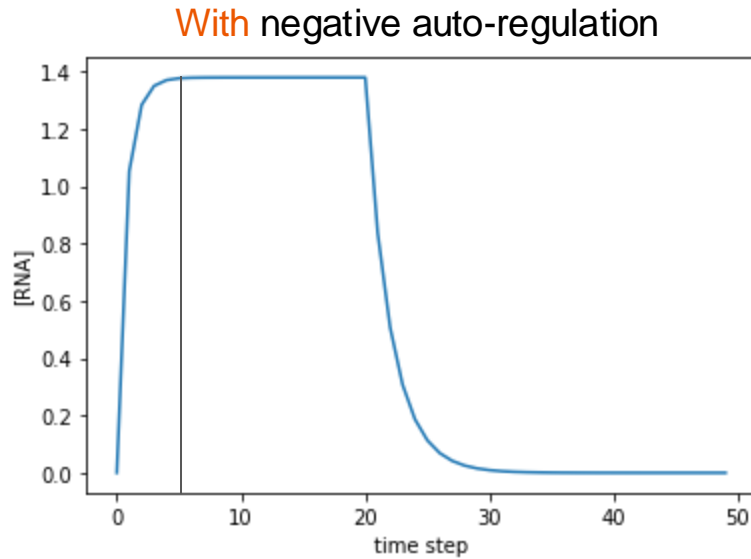
- $\frac{d[X]}{dt} = k_{transcription}[DNA_{active}] - k_{degradation}[X]$
- $[DNA_{active}]$ also depends on $[X]$ through binding dynamics
- Two variables!

Negative auto-regulation



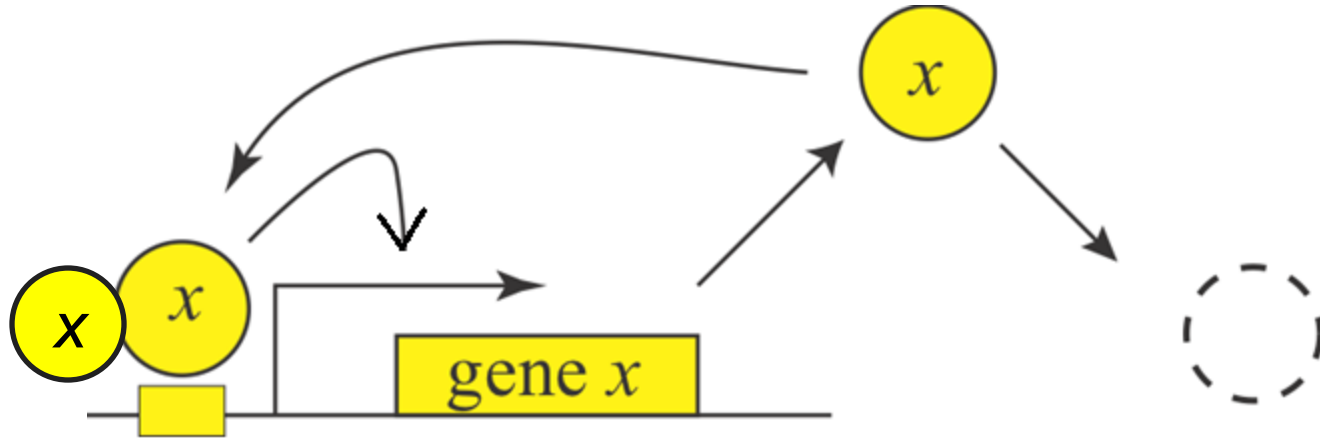
- $\frac{dDNA_{inactive}}{dt} = k_{bind}[X]^{\textcolor{teal}{n}}[DNA_{active}] - k_{unbind}[DNA_{inactive}]$
- At binding equilibrium, fraction of unbound (active) DNA = $\frac{k_{bind}[X]^n}{k_{unbind} + k_{bind}[X]^n}$
- Also known as **Hill function**

Faster response time with negative auto-regulation



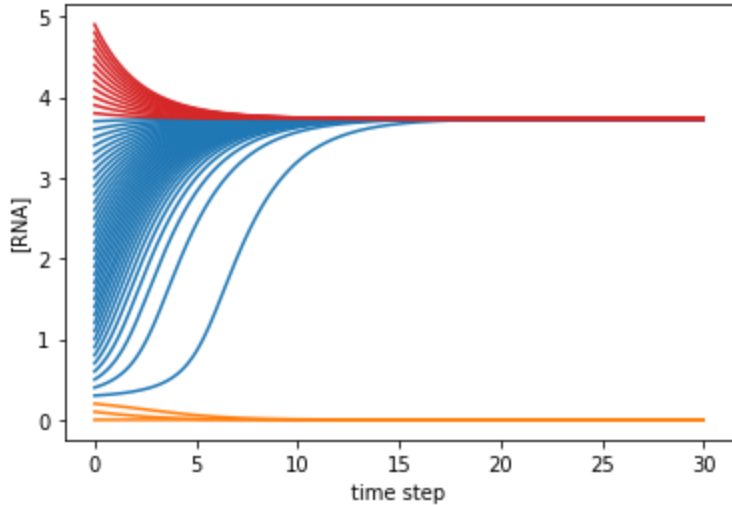
$$\frac{d[X]}{dt} = \frac{k_{\text{transcription}}}{1 + (k[X])^n} - k_{\text{degradation}}[X]$$

Positive auto-regulation



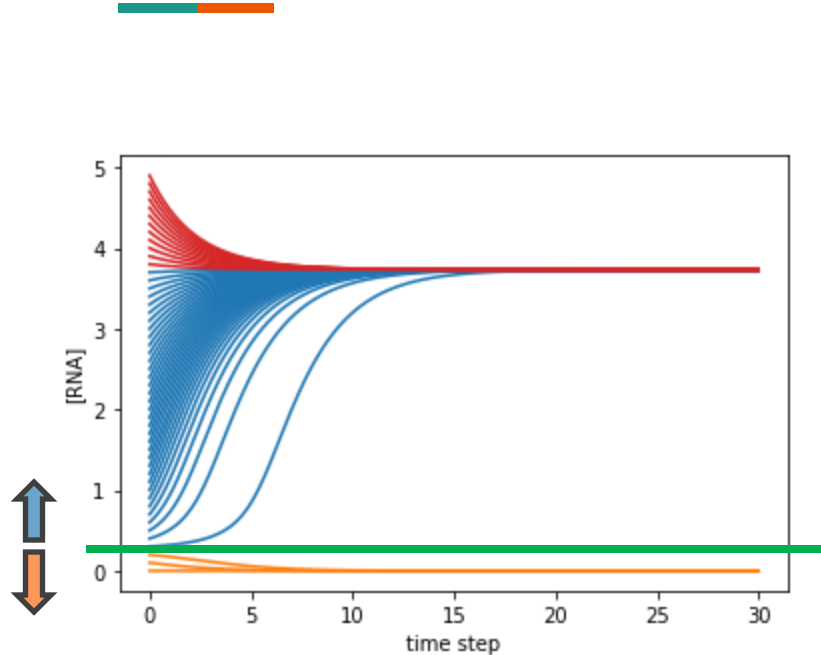
- $$\frac{d[X]}{dt} = \frac{k_{transcription}(k[X])^2}{1 + (k[X])^2} - k_{degradation}[X]$$
- Can you guess how this system will behave?
 - How many equilibria are there?

Bistability



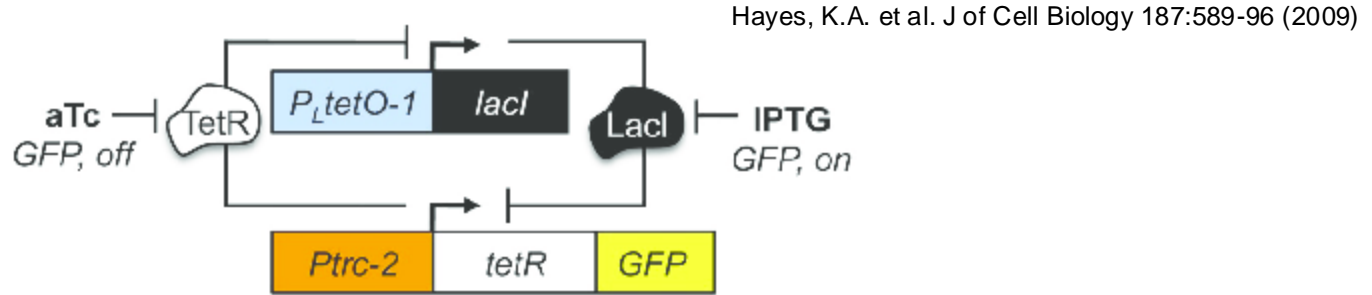
- For **low** $[X]$, the **degradation term wins over the transcription** and $[X]$ decays down to zero
- For **intermediate** $[X]$, the **transcription term wins over degradation** and $[X]$ increases until reaching the other equilibrium
- For **high** $[X]$, the **degradation term wins over the transcription** and $[X]$ decays down to the nearest equilibrium

Bistability as memory



- Given a cell in **low** expression state
- A drug treatment raises the expression level in the cell pass the **green** threshold
- After stopping the treatment, the cell will “**memorize**” that it is now in the **high** expression state

Memory from gene toggle switch



- Two genes repressing each other
- **Two equilibria:** High LacI or high GFP
- If GFP is ON, how can we turn it off? Adding **aTc** or **IPTG**?
- Once GFP is OFF, can we stop supplying those molecules to the cell?

Two-gene linear system



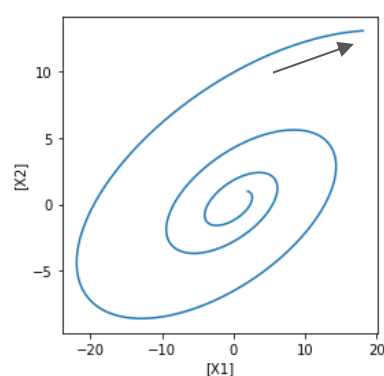
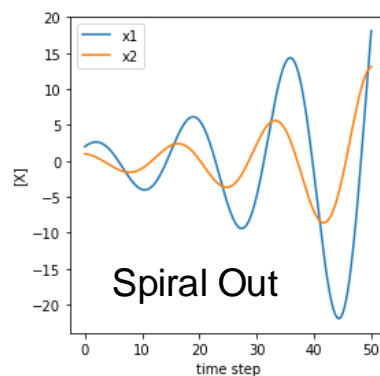
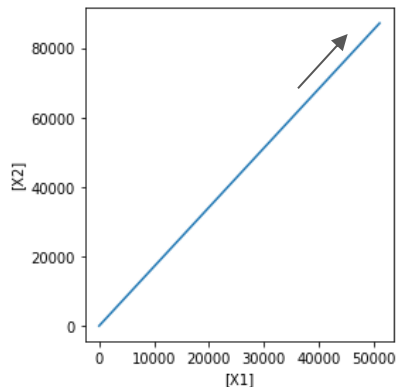
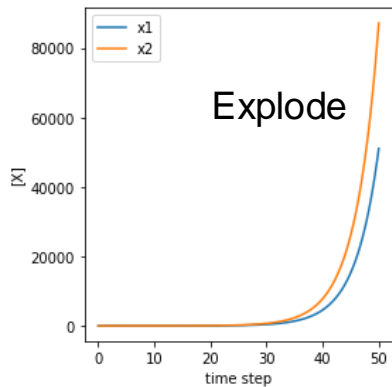
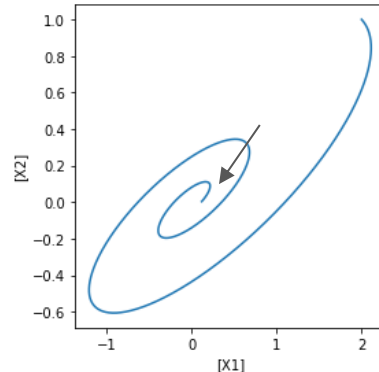
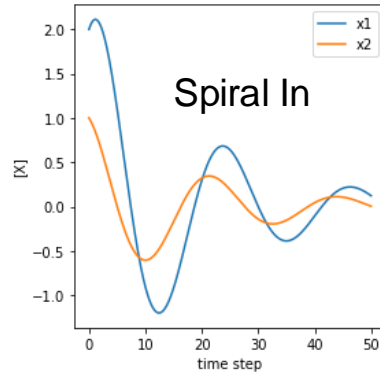
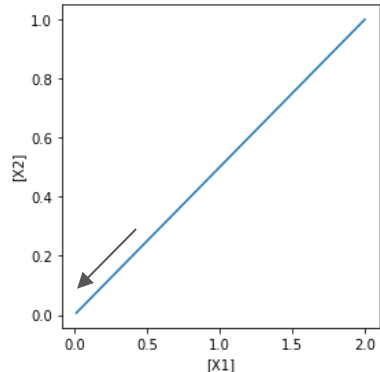
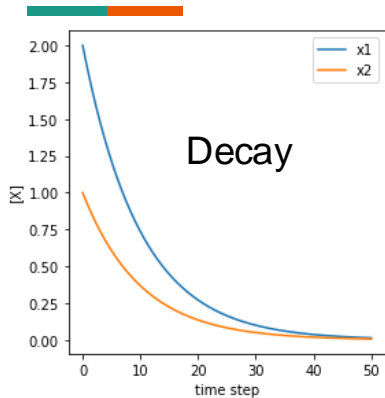
$$\frac{d[X1]}{dt} = k_{1,1}[X1] + k_{1,2}[X2]$$

$$\frac{d[X2]}{dt} = k_{2,1}[X1] + k_{2,2}[X2]$$

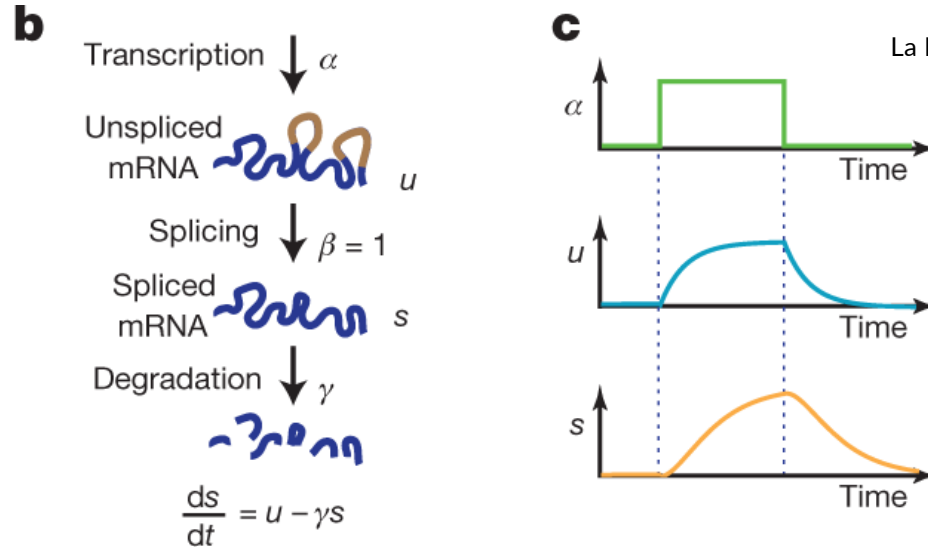
```
1 def two_loci_linear(rna, time, k11, k12, k21, k22):  
2     dx1_dt = k11 * rna[0] + k12 * rna[1]  
3     dx2_dt = k21 * rna[0] + k22 * rna[1]  
4  
5     return [dx1_dt, dx2_dt]
```

- A simple two-gene system with only linear effects
- Depending on $k_{i,j}$, interesting dynamics can be derived
 - Actually, depending on **Trace** and **Determinant** of the matrix $\{k_{i,j}\}$

Some possible dynamics



RNA velocity model



- **Two-step process:** From DNA to unspliced RNA to spliced RNA
- $\frac{d[U]}{dt} = \alpha - \beta[U], \frac{d[S]}{dt} = \beta[U] - \gamma[S]$
- Not the same dynamics as simplifying $\frac{d[S]}{dt} = \alpha - \gamma[S]$

Part 2: Useful online resources



- Use public datasets to develop hypothesis
- New findings can still be discovered in published data
- Be careful of hidden confounding and design factors
- Use multiple tools to confirm

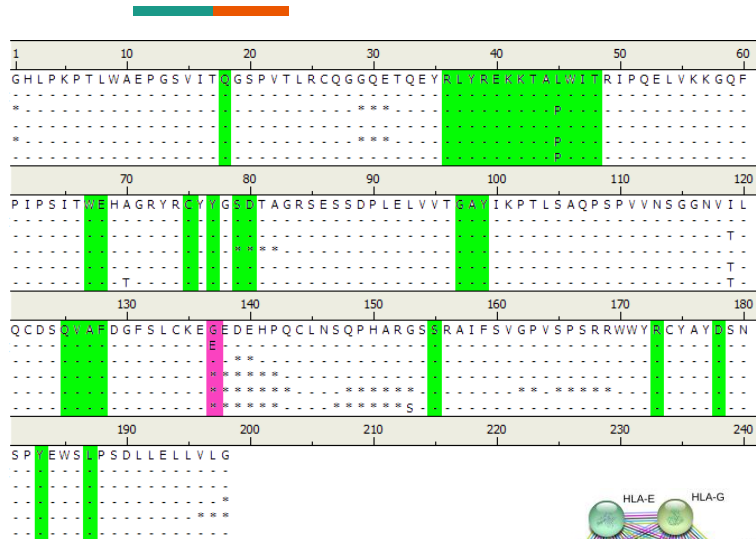
Some motivations

- Quick answers to quick questions
 - Impact of a newly identified mutation?
 - Where does a TF bind on genome?
- Free hypothesis generation
 - What genes are consistently up-regulated in a certain disease?
- Combine data from multiple modalities
 - Omics
 - Clinical
 - Imaging

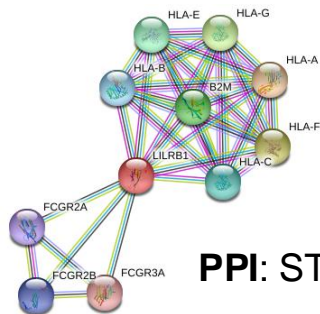


Image from boston.lti.cs.cmu.edu

Example: G137E mutation on ILT2



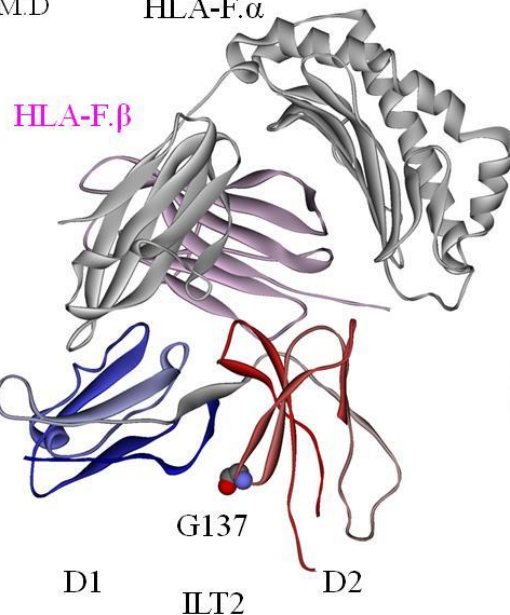
Sequence: Look up variant on ClinVar and alignment to reported sequences



PPI: STRING

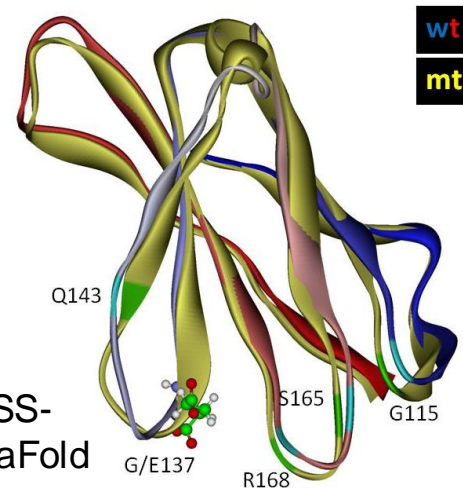
5KNM.D

HLA-F.α



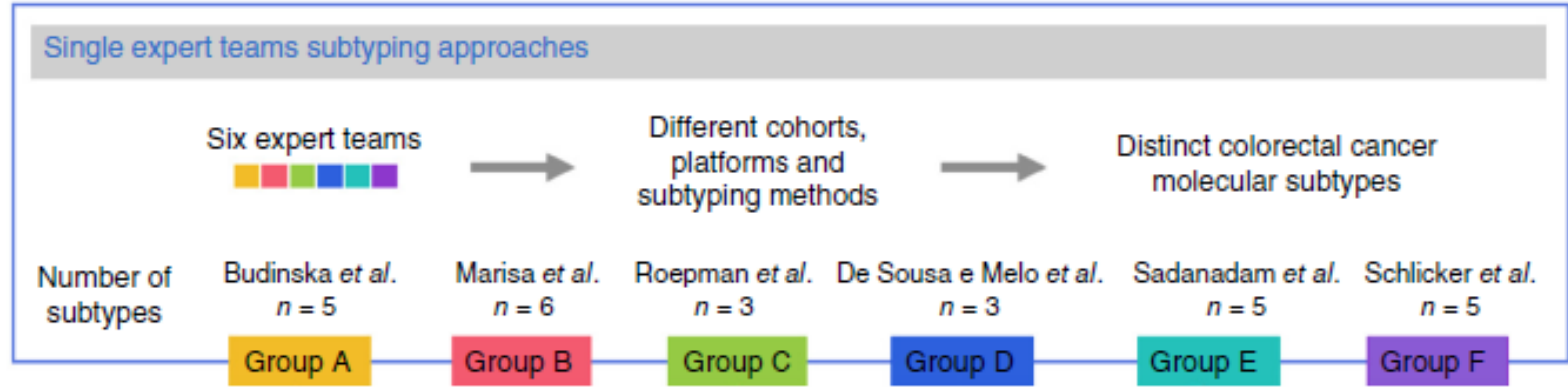
Structure: Highlight mutated residue on structure from PDB.

Modeling: SWISS-MODEL or AlphaFold



wt
mt

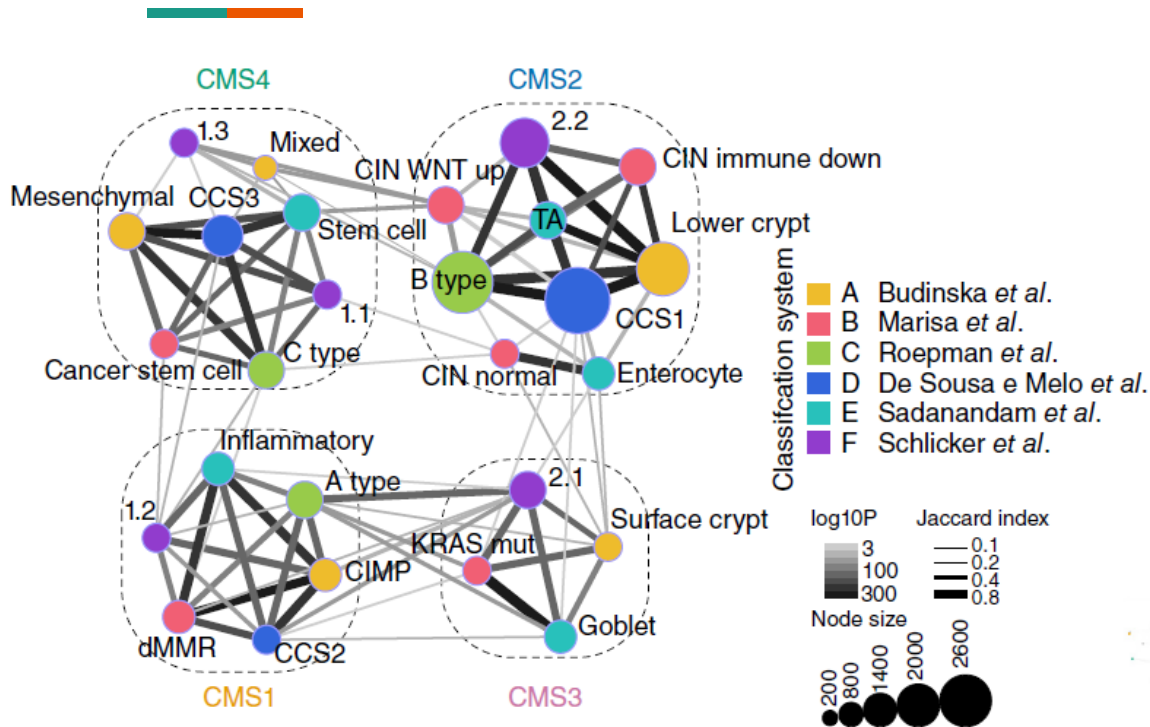
Example: Consensus cancer subtyping



Guinney et al. Nat Medicine 21: 1350-1356 (2015)

- Different clinicians characterize cancer patients differently
- But there should be a common molecular basis!

Example: Consensus cancer subtyping



- Compare clinician's decisions on a common group of patients
- Identify 4 major groups: Consensus Molecular Subtype (CMS)

Knowledgebase

- NCBI/GenBank
- Ensembl
- Uniprot
- GeneCard
- ENCODE
- Human Protein Atlas



UniProtKB - O75473 (LGR5_HUMAN)

Display

[Help video](#)[BLAST](#)[Align](#)[Format](#)[Add to basket](#)[History](#)[Entry](#)[Publications](#)[Feature viewer](#)[Feature table](#)[None](#)☒ Function☒ Names & Taxonomy☒ Subcellular location☒ Pathology & Biotech☒ PTM / Processing☒ Expression☒ Interaction☒ Structure☒ Family & Domains☒ Sequences (3)☒ Similar proteins☒ Cross-references**Protein****Leucine-rich repeat-containing G-protein coupled receptor 5****Gene****LGR5****Organism***Homo sapiens (Human)***Status**Reviewed - Annotation score: ●●●●● - Experimental evidence at protein levelⁱ

Functionⁱ

Receptor for R-spondins that potentiates the canonical Wnt signaling pathway and acts as a stem cell marker of the intestinal epithelium; binding to R-spondins (RSPO1, RSPO2, RSPO3 or RSPO4), associates with phosphorylated LRP6 and frizzled receptors that are activated triggering the canonical Wnt signaling pathway to increase expression of target genes. In contrast to classical G-protein coupled receptors heterotrimeric G-proteins to transduce the signal. Involved in the development and/or maintenance of the adult intestinal stem cells during

[5 Publications](#)

Miscellaneous

LGR5 is used as a marker of adult tissue stem cells in the intestine, stomach, hair follicle, and mammary epithelium.

[1 Publication](#)

GO - Molecular functionⁱ

- G protein-coupled peptide receptor activity [Source: GO_Central](#)
- G protein-coupled receptor activity [Source: ProtInc](#)
- protein-hormone receptor activity [Source: InterPro](#)
- transmembrane signaling receptor activity [Source: UniProtKB](#)

[Complete GO annotation on QuickGO ...](#)

GO - Biological processⁱ

Accession IDs



ID system	Accession
Gene Symbol	LGR5
HGNC gene	4504
Entrez (NCBI)	8549
RefSeq transcript	NM_00367
Ensembl	ENSG00000139292
Uniprot protein	O75473
OMIM	606667

Data repository

454 sequencing of Human HapMap individual NA18505 genomic paired-end library (SRR000001)

[Metadata](#) [Analysis](#) [Reads](#) [Data access](#)

Run	Spots	Bases	Size	GC content	Published	Access Type
SRR000001	471.0k	129.5Mbp	312.5M	41.3%	2008-04-04	public

Quality graph ([bigger](#))

This run has 4 reads per spot:

L=4, 100% L=187, σ =95.9, 100% L=44, 50% L=123, σ =65.5, 50%

[Legend](#)

Experiment	Library Name	Platform	Strategy	Source	Selection	Layout	Action
SRX000007	SID2748	LS454	WGS	GENOMIC	RANDOM	PAIRED	BLAST

Biosample	Sample Description	Organism	Links
SAMN00001583 (SRS000100)	Human HapMap individual Coriell catalog ID NA18505	Homo sapiens	<ul style="list-style-type: none">dbSNP Batch ID 1061891Individual record in dbSNP

Bioproject	SRA Study	Title
PRJNA33627	SRP000001	Paired-end mapping reveals extensive structural variation in the human genome
Show abstract		



[Full experiment listing](#)

PXD027487

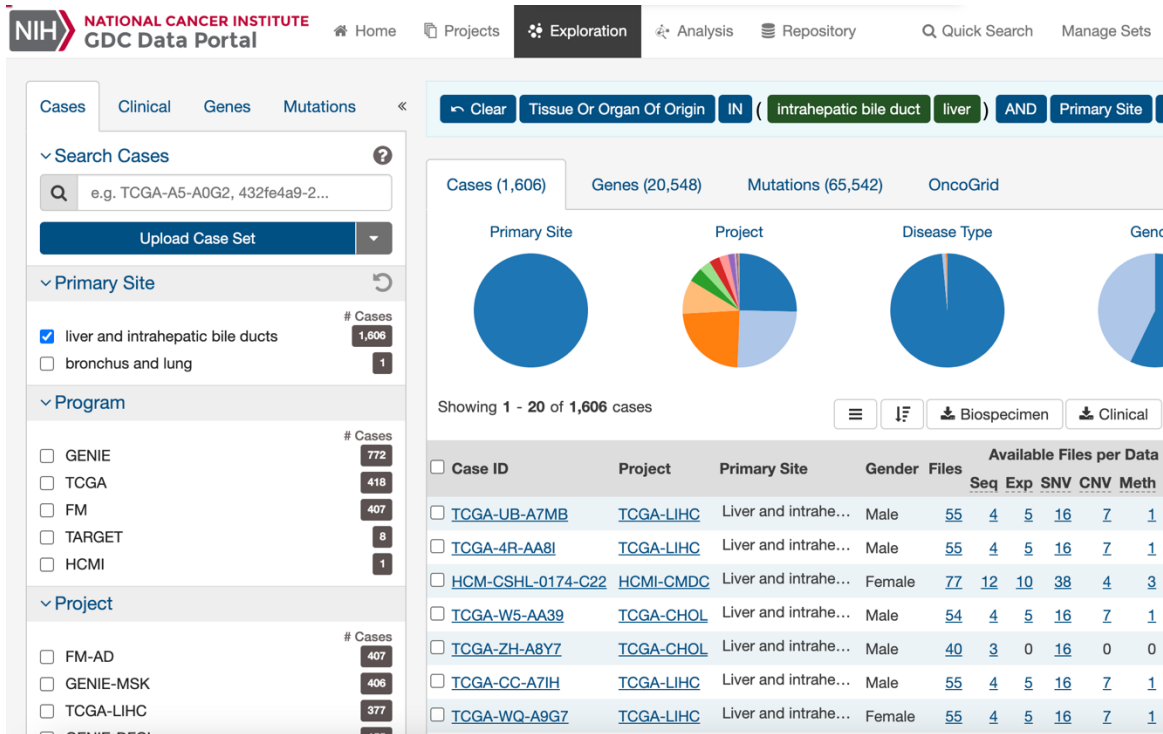
PXD027487 is an **original dataset** announced via ProteomeXchange.

Dataset Summary

Title	Interactome of GIGYF2 and EIF4E2 with and without SMG11 treatment
Description	Translation of messenger RNAs (mRNAs) with premature translation termination codons produces truncated proteins with potentially deleterious effects. This is prevented by nonsense-mediated mRNA decay (NMD) of these mRNAs. NMD is triggered by ribosomes terminating upstream of a splice site marked by an exon-junction complex (EJC), but also acts on many mRNAs lacking a splice junction after their termination codon. We developed a genome-wide CRISPR flow cytometry screen to identify regulators of mRNAs with premature termination codons in K562 cells. This screen recovered essentially all core NMD factors and suggested a role for EJC factors in degradation of PTCs without downstream splicing. Among the strongest hits were the translational repressors GIGYF2 and EIF4E2. GIGYF2 and EIF4E2 mediate translational repression but not mRNA decay of a subset of NMD targets and interact with NMD factors genetically and physically. Our results suggest a model wherein recognition of a stop codon as premature can lead to its translational repression through GIGYF2 and EIF4E2.
HostingRepository	PRIDE
AnnounceDate	2021-10-12
AnnouncementXML	Submission_2021-10-12_00:39:40.790.xml
DigitalObjectIdentifier	
ReviewLevel	Peer-reviewed dataset
DatasetOrigin	Original dataset

- Reanalysis with a common pipeline


Genomic Data Commons



- Cancer multi-omics
 - Exome
 - RNA-seq
 - Methylation
- Clinical and demographic data
- Some with histopathological images and radiographic images

PDB: 3D structures

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB

Refinements ? 

Summary Gallery Compact -- Tabular Report -- ↓ Score Download Files ☒ All ☐ Selected

Displaying 1 to 25 of 33 Structures Page 1 of 2 ← Previous Next → Display 25 per page

SCIENTIFIC NAME OF SOURCE ORGANISM

- ☐ Homo sapiens (13)
- ☐ Xenopus tropicalis (11)
- ☐ Mus musculus (10)
- ☐ Danio rerio (2)
- ☐ unidentified (1)

TAXONOMY

- ☐ Eukaryota (32)
- ☐ unclassified sequences (1)

EXPERIMENTAL METHOD

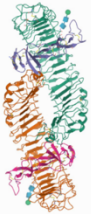

- ☐ X-RAY DIFFRACTION (32)
- ☐ SOLUTION NMR (1)

POLYMER ENTITY TYPE

- ☐ Protein (33)

REFINEMENT RESOLUTION (Å)

- ☐ 1.5 - 2.0 (2)
- ☐ 2.0 - 2.5 (11)
- ☐ 2.5 - 3.0 (9)
- ☐ 3.0 - 3.5 (8)
- ☐ 4.0 - 4.5 (1)
- ☐ > 4.5 (1)

4BSU   Download File View File ☒

Structure of the ectodomain of LGR5 in complex with R-spondin-1 (Fu1Fu2) in C2 crystal form

Peng, W.C., de Lau, W., Forneris, F., Granneman, J.C.M., Huch, M., Clevers, H., Gros, P.

(2013) Cell Rep 3: 1885

Released 2013-06-26

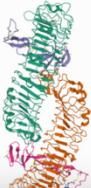

Method X-RAY DIFFRACTION 3.2 Å

Organisms [Homo sapiens](#)

Macromolecule [LEUCINE-RICH REPEAT-CONTAINING G-PROTEIN COUPLED RECEPTOR 5 \(protein\)](#)
[R-SPONDIN-1 \(protein\)](#)

Unique Ligands [NAG](#)

Unique branched monosaccharides BMA, NAG

4BST   Download File View File ☒

Structure of the ectodomain of LGR5 in complex with R-spondin-1 (Fu1Fu2) in P6122 crystal form

Peng, W.C., de Lau, W., Forneris, F., Granneman, J.C.M., Huch, M., Clevers, H., Gros, P.

(2013) Cell Rep 3: 1885

Released 2013-06-19

Method X-RAY DIFFRACTION 4.3 Å

Gene Expression Omnibus / ArrayExpress

Study type

TOP 10

see all

- ☐ chip-chip by tiling array 706
- ☐ chip-seq 1,467
- ☐ comparative genomic hybridization by array 883
- ☐ genotyping by array 254
- ☐ methylation profiling by array 617
- ☐ other 222
- ☐ rna-seq of coding rna 1,437
- ☐ rna-seq of non coding rna 241
- ☐ transcription profiling by array 9,339
- ☐ unknown experiment type 218

You query contains a term which has too many synonyms and more specific phrases in EFO. The results shown below do not include those expanded terms.

Search results for breast cancer metastasis

1 – 20 of 1,417+ results

Sort by:

Relevance ▼

▼

▲

E-MTAB-4801 · 9 January 2019 · 3 links · 3 files

Variation in RNA expression in a panel of 30 breast cancer cell lines

... type comparison design EFO_0001745 cell line cell line BT20 BT474 BT549 ... 27 other values Homo sapiens female mammary gland invasive ductal carcinoma breast ductal adenocarcinoma metaplastic breast carcinoma squamous cell breast carcinoma, acantholytic variant breast adenocarcinoma breast...

E-MTAB-8807 · 4 April 2020 · 1 link · 2 files

Estrogen receptor beta inhibits cholesterol biosynthesis through overexpression of mir-181a-5p in Triple Negative Breast Cancer

... carcinoma squamous cell breast carcinoma, acantholytic variant not specified invasive breast ductal carcinoma invasive lobular carcinoma adenocarcinoma atypical carcinoma carcinoma ER beta negative ER beta positive HCC1806 null null null epithelial cell squamous cell breast carcinoma, acantholytic...

- Search for omics datasets

Google Dataset / FigShare

Dataset Search

breast cancer metastasis



- Whole-exome sequencing of **breast cancer metastasis** and corresponding blood samples
- Genomic Evolution of **Breast Cancer Metastasis** and Relapse
- A clinical decision support system learned from data to personalize treatment recommendations towards preventing **breast cancer metastasis**
- Cooperativity between EMT and non-EMT cells promotes **breast cancer metastasis** via paracrine GLI activation
- Primary tumor grafts as advanced models for breast cancer that authentically reflect tumor histopathology, growth, metastasis, and patient outcomes (copy number)
- Imipramine Blue: A potent therapeutic regimen that suppresses breast cancer growth and metastasis

1,494,669 results found



Investigating novel mechanisms of metastasis in ...



Stratification of radiosensitive brain metastases based o...



Identification of Molecular Mediators of Endocrine ...

- Google Dataset compiles links from major public data repositories
- Figshare is a free repository that scientists often use for non-omics data

MSigDB: Curated gene sets

Human MSigDB Collections



The 33196 gene sets in the Human Molecular Signatures Database (MSigDB) are divided into 9 major collections, and several sub-collections. See the table below for a brief description of each, and the [Human MSigDB Collections: Details and Acknowledgments](#) page for more detailed descriptions. See also the [MSigDB Release Notes](#).

Click on the "browse gene sets" links in the table below to view the gene sets in a collection. Or download the gene sets in a collection by clicking on the links below the "Download Files" headings. For a description of the [GMT file format](#) see the [Data Formats](#) in the [Documentation](#) section. The gene sets can be downloaded as NCBI (Entrez) Gene Identifiers or HUGO (HGNC) Gene Symbols. There are also JSON bundles containing the HUGO (HGNC) Gene Symbols along with some useful metadata. An XML file containing all the Human MSigDB gene sets is available as well.

Collection	Description	Download Options
H: hallmark gene sets (browse 50 gene sets)	Hallmark gene sets summarize and represent specific well-defined biological states or processes and display coherent expression. These gene sets were generated by a computational methodology based on identifying overlaps between gene sets in other MSigDB collections and retaining genes that display coordinate expression. details	Download GMT Files Gene Symbols NCBI (Entrez) Gene IDs JSON bundle
C1: positional gene sets (browse 299 gene sets)	Gene sets corresponding to human chromosome cytogenetic bands. details	Download GMT Files Gene Symbols NCBI (Entrez) Gene IDs JSON bundle
C2: curated gene sets (browse 6449 gene sets)	Gene sets in this collection are curated from various sources, including online pathway databases and the biomedical literature. Many sets are also contributed by individual domain experts. The gene set page for each gene set lists its source. The C2 collection is divided into the following two sub-collections: Chemical and genetic perturbations (CGP) and Canonical pathways (CP). details	Download GMT Files Gene Symbols NCBI (Entrez) Gene IDs JSON bundle

- Function-based / Disease-based / Publication-based / Genomic location-based
- For gene panel, enrichment test, etc.

miRNA target



Release 7.1: June 2016

Agarwal *et al.*, 2015

Search for predicted microRNA targets in mammals

[\[Go to TargetScanMouse\]](#)

[\[Go to TargetScanWorm\]](#)

[\[Go to TargetScanFly\]](#)

[\[Go to TargetScanFish\]](#)

1. Select a species

AND

2. Enter a human gene symbol (e.g. "Hmga2")
or an Ensembl gene (ENSG00000149948) or transcript (ENST00000403681) ID

AND/OR


3. Do one of the following:

- Select a broadly conserved* microRNA family
- Select a conserved* microRNA family
- Select a poorly conserved but confidently annotated microRNA family
- Select another miRBase annotation

Other miRBase annotations

Note that most of these families are star miRNAs or RNA fragments misannotated as miRNAs.

Enter a microRNA name (e.g. "miR-9-5p")

miRBase


[Home](#) [Search](#) [Browse](#) [Help](#) [Download](#) [Blog](#) [Submit](#)

Latest miRBase blog posts

High confidence miRNA set available for miRBase 21 By sam (July 3, 2014)
As mentioned previously, we briefly held off from releasing the set of "high confidence" miRNAs for miRBase 21, because of a last-gasp bug. Those data are now available, tagged with the label "high confidence" on the entry pages, and for download on the FTP site. The total number of miRNAs labelled "high confidence" has increased [...]

miRBase 21 finally arrives By sam (June 26, 2014)
Apologies for the longer-than-usual wait. miRBase 21 is now available on the website, and all data available for download on the FTP site. As usual, the release notes describe the major changes. Of particular note this time, the Genome Reference Consortium have released a new human genome assembly, GRCh38. We have therefore remapped the human [...]

miRBase: the microRNA database



Target Search

Target Mining

Custom Prediction

FuncMir Collection

Data Download

Statistics

Help | FAQ

Comments

Citation | Policy

Choose one of the following search options:

Search by miRNA name

Human

Search by gene target

Human Gene Symbol

miRDB is an online database for miRNA target prediction and functional annotations. All bioinformatics tool, MirTarget, which was developed by analyzing thousands of miRNA-target sequencing experiments. Common features associated with miRNA target binding have targets with machine learning methods. miRDB hosts predicted miRNA targets in five s. As a recent update, users may provide their own sequences for customized target prediction computational analyses and literature mining, functionally active miRNAs in humans and as well as associated functional annotations, are presented in the FuncMir Collection in mi

SWISS-MODEL: Homology modeling



BIOZENTRUM

University of Basel
The Center for Molecular Life Sciences

SWISS-MODEL

Modelling

Repository

Start a New Modelling Project

Target

⚙ Target MAAHKGAEEHHKAAEEHHEQAAKHHHAAAEHHEKGEHEQAHHADTAYAHKKHAEHAAQAAKHDAEEHHAPKH 73

Sequence(s):

(Format must be FASTA,

Clustal,

plain string, or a valid

UniProtKB AC)

Add Hetero Target

Reset

Project Title:

Class example

Email:

Optional

Search For Templates

Build Model

AI-assisted protein folding (ColabFold)

3. Enter the amino acid sequence(s) to fold

Enter the amino acid sequence(s) to fold:

- If you enter only a single sequence, the monomer model will be used.
- If you enter multiple sequences, the multimer model will be used.

sequence_1: " MAAHKGAEH HHKAAEHHEQA AKHHHAAA EHHEKGEHEQA

sequence_2: " Insert text here

<https://github.com/sokrypton/ColabFold>

5. Run AlphaFold and download prediction

Once this cell has been executed, a zip-archive will be downloaded. In case you are having issues with the relaxation violations.

run_relax: ☒

Relaxation is faster with a GPU, but we have four reverts to using without GPU.

relax_use_gpu: ☐

[Show code](#)

- Accessible, simplified Python interface

GenePattern / Galaxy



Features

Powerful genomics tools in a user-friendly interface



GenePattern provides hundreds of analytical tools for the analysis of gene expression (RNA-seq and microarray), sequence variation and copy number, proteomic, flow cytometry, and network analysis. These tools are all available through a Web interface with no programming experience required.

GenePattern Notebook



The GenePattern Notebook environment extends the Jupyter Notebook system, allowing researchers to create documents that interleave formatted text, graphics and other multimedia, executable code, and GenePattern analyses, creating a single "research narrative" that puts scientific discussion and analyses in the same place.

Blog > GP updates

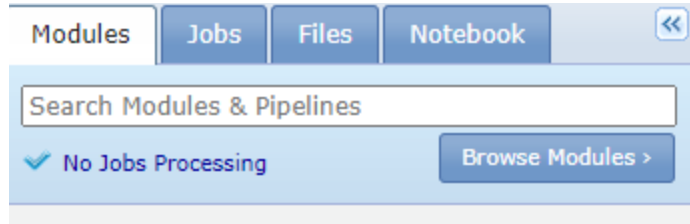
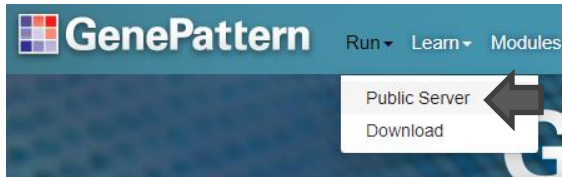
- GenePattern Coverage and Support December 23, 2021 - January 2, 2022
- End of Support for GParc
- End of support for modules using the deprecated GenePattern patch mechanism.
- End of support for Cufflinks suite of modules
- End of support for MAGeCK

[view more >](#)

Papers > Related Publications

- Comparability and reproducibility of biomedical data
- TopCluster: a multiple gene list feature analyzer for

Content



- Modules = standard bioinformatics tools
- Notebook = Python environment (similar to Google Colab) with specific bioinformatics library

Browse Modules & Pipelines

clustering

cnv analysis

conumee

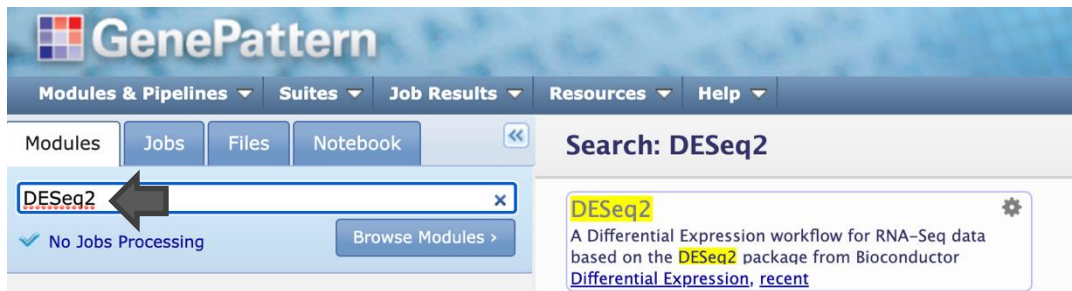
data format conversion

differential expression

dimension reduction

flow cytometry

DESeq2 on GenePattern



- Input expression data (RNA-seq read count) and sample label
- .gct and .cls are text files

input file*

▼Hide Files...(Selected 1 files)

htseq_count.gct

A GCT file containing raw RNA-Seq counts, such as is produced by MergeHTSeqCounts

cls file*

▼Hide Files...(Selected 1 files)

sample_info.cls

A categorical CLS file specifying the phenotype classes for the samples in the GCT file. This should contain exactly two classes with the control specified first.

confounding variable cls file

Upload File... Add Path or URL... Drag Files Here

2GB file upload limit using the Upload File... button. For files > 2GB upload from the Files tab.

A categorical CLS file specifying an additional confounding variable, mapped to the input file samples. Use this for a two-factor comparison.

#1.2					
60488	12				
Name	Description	sample1	sample2	sample3	sample4
ENSG00000242268.2	ENSG00000242268.2	2	10	0	2
ENSG00000270112.3	ENSG00000270112.3	8	6	0	0
ENSG00000167578.15	ENSG00000167578.15	102	633	468	1200
ENSG00000273842.1	ENSG00000273842.1	0	0	0	0
ENSG00000078237.5	ENSG00000078237.5	370	364	1220	692

```
12 2 1
# alive dead
1 1 0 0 1 0 0 0 1 0 1 1
```


DESeq2 output


471244. DESeq2


Source:  GenePattern production (new)

submitted: Oct 25 01:06:41 AM, completed: Oct 25 01:10:38 AM, size: 30 MB

[Show details](#)

 **Comments (0)**

 **Tags (0)**


 [input.file: expression.gct](#)


 [cls.file: label.cls](#)

 [class_example.Positive.vs.Negative.DESeq2_results_report.txt](#) (670.0 KB) (Last modified: 2022-10-25 01:10:26.0)

 [class_example.Positive.vs.Negative.QC.DispEsts.png](#) (78.0 KB) (Last modified: 2022-10-25 01:10:26.0)

 [class_example.Positive.vs.Negative.QC.MAplot.png](#) (9.0 KB) (Last modified: 2022-10-25 01:10:26.0)


 [class_example.Positive.vs.Negative.mean_values_by_class.txt](#) (230.0 KB) (Last modified: 2022-10-25 01:10:26.0)

 [class_example.Positive.vs.Negative.normalized_counts.gct](#) (14.5 MB) (Last modified: 2022-10-25 01:10:26.0)

 [class_example.Positive.vs.Negative.normalized_counts.txt](#) (14.5 MB) (Last modified: 2022-10-25 01:10:26.0)

 [class_example.Positive.vs.Negative....0_downregulated_genes_report.txt](#) (3.0 KB) (Last modified: 2022-10-25 01:10:26.0)

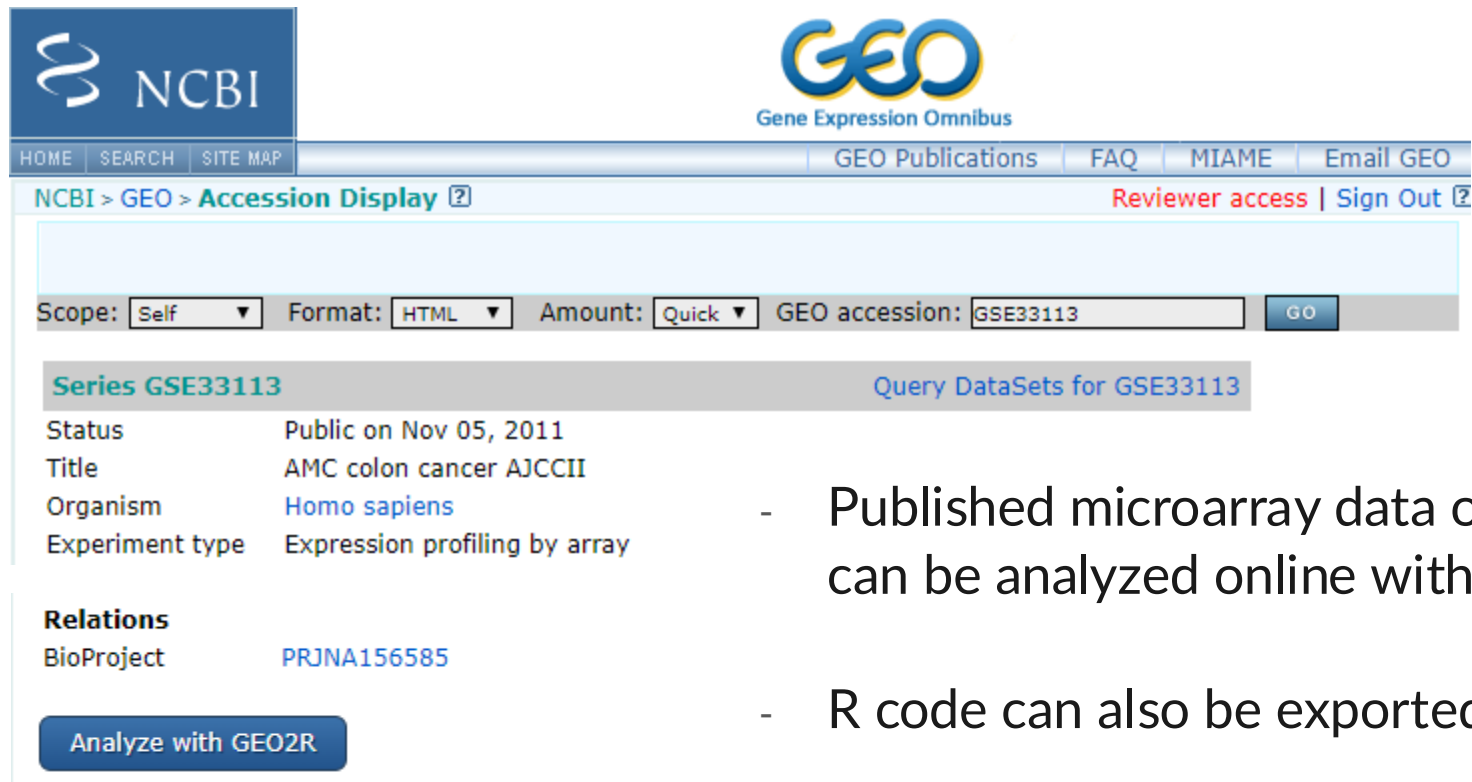
 [class_example.Positive.vs.Negative.top20_upregulated_genes_report.txt](#) (3.0 KB) (Last modified: 2022-10-25 01:10:26.0)

 [stderr.txt](#) (1.0 KB) (Last modified: Tue Oct 25 01:10:37 UTC 2022)

 [stdout.txt](#) (4.0 KB) (Last modified: Tue Oct 25 01:10:37 UTC 2022)

 [gp_execution_log.txt](#) (1.0 KB) (Last modified: Tue Oct 25 01:10:38 UTC 2022)

GEO2R: Differential expression



The screenshot shows the NCBI GEO Accession Display page for the series GSE33113. The page includes the NCBI logo, the GEO logo (Gene Expression Omnibus), and navigation links. The main content area displays the series title, status, organism, and experiment type. A 'Relations' section shows the associated BioProject (PRJNA156585). A button at the bottom left is labeled 'Analyze with GEO2R'.

NCBI > GEO > **Accession Display** [?](#) [Reviewer access](#) | [Sign Out](#) [?](#)

Scope: Format: Amount: GEO accession:

Series GSE33113 [Query DataSets for GSE33113](#)

Status	Public on Nov 05, 2011
Title	AMC colon cancer AJCCII
Organism	Homo sapiens
Experiment type	Expression profiling by array

Relations

BioProject	PRJNA156585
------------	-----------------------------

- Published microarray data on GEO can be analyzed online with GEO2R
- R code can also be exported

Demo time!



- Let's try some online tools:
 - ColabFold interface for AlphaFold2
 - STRING for viewing protein-protein interaction
 - GEO2R for differential expression analysis of microarray data

Any question?

