

Bioinformatics analysis of differentially expressed genes in osteoblasts during spaceflight

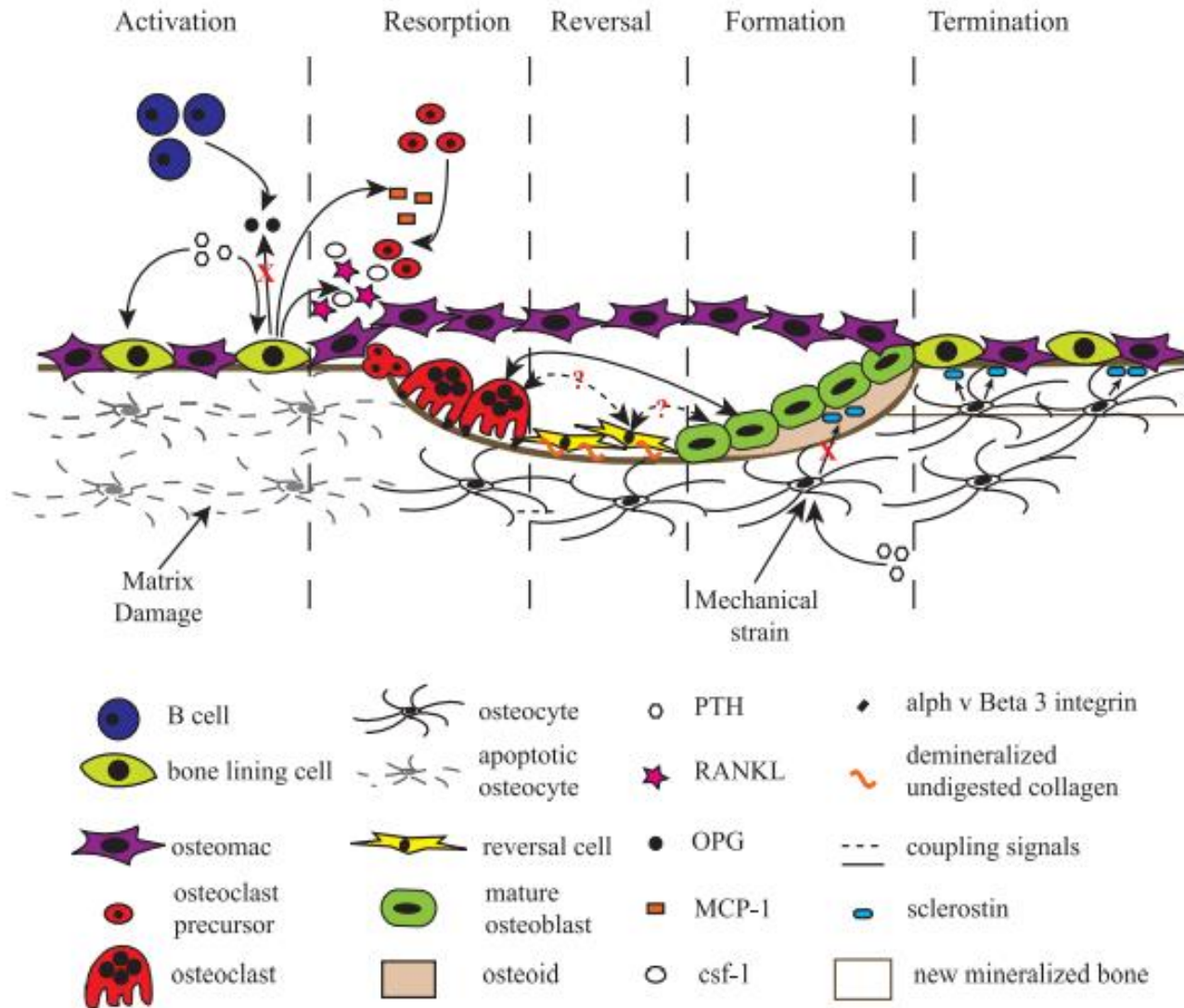
Group 9:崔轶男
赵璐璐
李晓宇
李铠

Bone loss in spaceflight



- Microgravity cause bone loss of astronauts
- Loss 1-2% bone mass per month in space

Bone remodeling



RNA-seq in Osteoblasts



- Osteoblasts carried by spacecraft
- RNA extraction
- Illumina Hiseq™ 2000 RNA-seq

Bioinformatics Analysis

- Mapping to Genome
- Find Differentially Expressed Genes
- GO Annotation
- Pathway Analysis
- WDR36 and F11R

Bioinformatics Analysis

- **Mapping to Genome**

- Find Differentially Expressed Genes

- GO Annotation

- Pathway Analysis

- WDR36 and F11R

Mapping to Genome

- Filtration
- Mapping
- Calculate Expression

Mapping to Genome

Filtration

- Filtrate the sequencing results Fligh1、 Flight2 and Control1、 Control2
- Take human's genomics data as the reference genome by software **BLASTP** and set the parameter as E-value<1.0E-20
- Get filter_Flight1、 filter_Flight2 and filter_Control1、 filter_Control2;

Mapping to Genome

Mapping

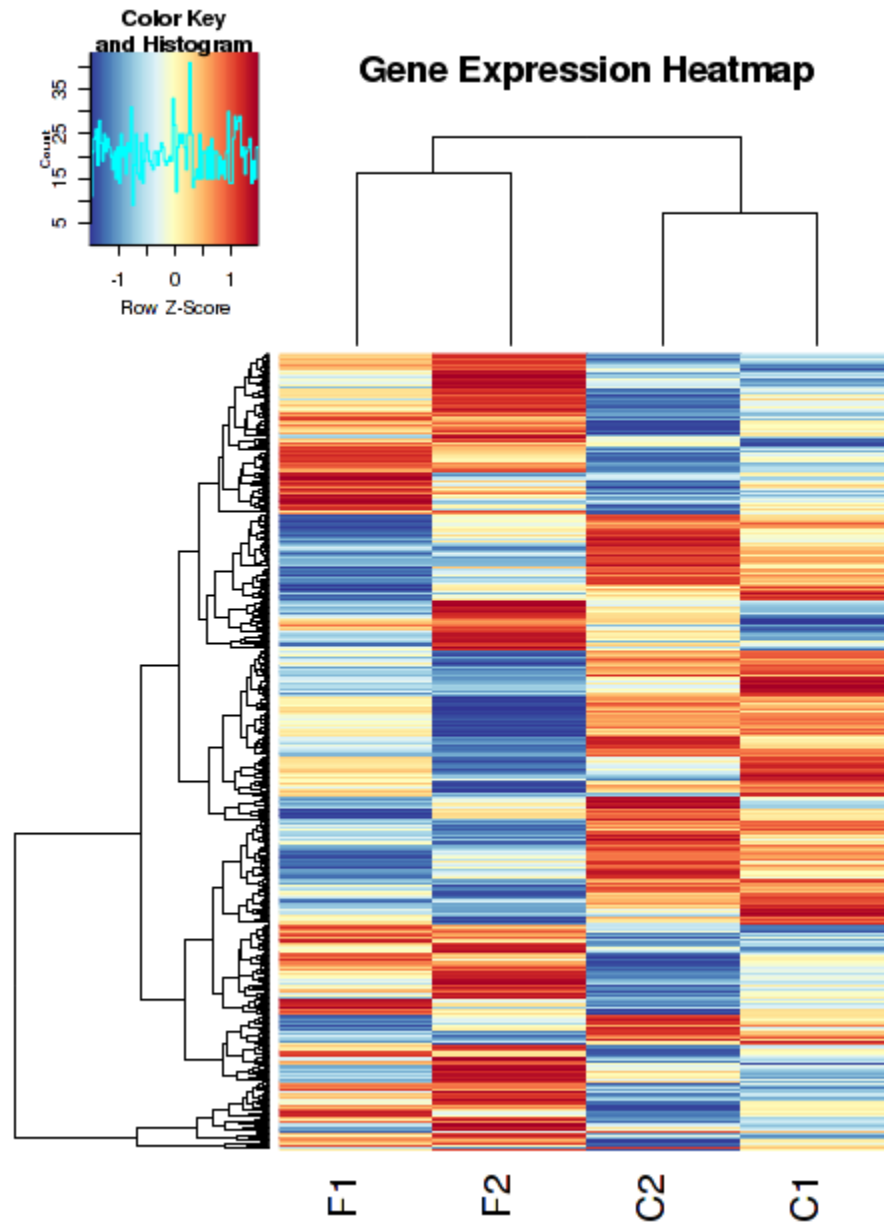
- Compare the filter_Flight1, filter_Flight2 and filter_Control1, filter_Control2 with the reference genome by software **GSMAP**
- Dispose them by **SAMTOOLS** and get filter_Flight1_out, filter_data1-2_out.fasta and filter_data2-1_out.fasta, filter_data2-1_out.fasta;

Mapping to Genome

Calculate expression

- Use software CUFFLINKS to calculate RPKM, RPKM(Reads Per Kilobases per Million reads.
- The formula is

$$\text{RPKM} = \frac{\text{total exon reads}}{\text{mapped reads (millions)} * \text{exon length (KB)}}$$



Bioinformatics Analysis

- Mapping to Genome
- **Find Differentially Expressed Genes**
- GO Annotation
- Pathway Analysis
- WDR36 and F11R

Find Differentially expressed genes

R—DESeq

Cutoff

➤ Foldchange > 1.5

➤ FDR < 0.05

```
1 library(DESeq)
2 MG <- read.delim("mRNA count all.txt",row.names=1)
3 MGconditions <- factor(c("Flight","Flight","Control","Control"))
4 MGcds <- newCountDataSet(MG,MGconditions)
5 MGcds <- estimateSizeFactors(MGcds)
6 MGcds <- estimateDispersions(MGcds)
7 MGres <- nbinomTest(MGcds,"Flight","Control")
8 write.csv( MGres,file="MGTop.csv")
9 MGG <- counts(MGcds,normalized=TRUE)[which(MGres$padj<=0.05 & (MGres$foldChange>1.5
10 write.csv(MGG,file = "MGGs.csv")
11 read.csv(file = "MGGs.csv",header = TRUE)
```

Differentially expressed genes

➤ UP-expression: 41 genes

ACVRL1, PLIN2, ANG, SERPING1, C3, CA9

➤ Down-Expression: 459 genes

HSPA1B, IDI1, IFRD1, IGFBP5, IL13RA2, KCNJ12

Bioinformatics Analysis

- Mapping to Genome
- Find Differentially Expressed Genes
- **GO Annotation**
- Pathway Analysis
- WDR36 and F11R

GO (Gene Ontology) Annotation

➤ Molecular Function

➤ Biological Process

➤ Cellular Component

DAVID GO annotation

The screenshot shows the DAVID Functional Annotation Tool website. At the top, there is a blue header with the DAVID logo and the text "Functional Annotation Tool" and "DAVID Bioinformatics Resources 6.7, NIAID/NIH". Below the header is a navigation menu with links: Home, Start Analysis, Shortcut to DAVID Tools, Technical Center, Downloads & APIs, Term of Service, Why DAVID?, and About Us.

Below the navigation menu, there are two red announcements:

- *** Announcing DAVID 6.8 Beta with updated Knowledgebase ([more info](#)). You may explore the new version at david-d.ncifcrf.gov.
- *** Please read: Due to data center maintenance, DAVID will be offline from Friday, June 17th @ 4pm EST through Sunday, June 19th with the possibility of being back online sooner. ***

The main content area is titled "Functional Annotation Tool" and has a left sidebar with tabs: Upload, List, and Background. The "Upload" tab is selected.

Under the "Upload" tab, there is a section "Upload Gene List" with links for "Demolist 1", "Demolist 2", and "Upload Help".

Below this is "Step 1: Enter Gene List" with two options:

- A: Paste a list: A text input field with a "Clear" button.
- Or
- B: Choose From a File: A file selection button labeled "选择文件" and "未选择任何文件". Below it is a checkbox for "Multi-List File" with a help icon.

On the right side of the main content area, there is a blue arrow pointing left with the text "Submit your gene list to start the tool!". To the right of this are three links: "Tell us how you like the tool", "Read technical notes of the tool", and "Contact us for questions".

Below this is a section titled "Key Concepts:" with two sub-sections:

- The DAVID Gene Concept**: DAVID 6.7 is designed around the "DAVID Gene Concept", a graph theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genomic resources including NCBI, PIR and Uniprot/SwissProt. The DAVID Gene Concept method groups tens of million of identifiers from over 65,000 species into 1.5 million unique protein/gene records. [More](#)
- Term/Gene Co-Occurrence Probability**: Ranking functional categories based on co-occurrence with sets of genes in a gene list can rapidly aid in unraveling new biological processes associated with cellular functions and pathways. DAVID 6.7 allows investigators to sort gene categories from dozens of annotation systems. Sorting can be based either the number of genes within each category or by the FASE score. [More](#)









<https://david.ncifcrf.gov/summary.jsp>

Biology Process Clustering

Annotation Cluster 1		Enrichment Score: 1.41	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_DIRECT	tRNA modification	RT		5	1.6E-2	9.9E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	tRNA processing	RT		7	5.3E-2	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	tRNA methylation	RT		3	7.4E-2	1.0E0
Annotation Cluster 2		Enrichment Score: 1.23	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_DIRECT	double-strand break repair via homologous recombination	RT		8	4.1E-2	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	double-strand break repair	RT		8	5.5E-2	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	DNA repair	RT		14	8.9E-2	1.0E0
Annotation Cluster 3		Enrichment Score: 0.11	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_DIRECT	phosphatidylinositol-mediated signaling	RT		4	5.2E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	Fc-epsilon receptor signaling pathway	RT		8	5.7E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	Ras protein signal transduction	RT		6	6.4E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	epidermal growth factor receptor signaling pathway	RT		7	7.4E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	activation of MAPKK activity	RT		5	7.7E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	fibroblast growth factor receptor signaling pathway	RT		6	7.9E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	insulin receptor signaling pathway	RT		6	8.2E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	MAPK cascade	RT		4	9.2E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	axon guidance	RT		9	9.2E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	neurotrophin TRK receptor signaling pathway	RT		6	9.5E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	vascular endothelial growth factor receptor signaling pathway	RT		4	9.5E-1	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	small GTPase mediated signal transduction	RT		10	9.8E-1	1.0E0

RNA metabolism, DNA repair, Signal transduction

Molecular Function Clustering

Annotation Cluster 1		Enrichment Score: 8.67	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_MF_DIRECT	metal ion binding	RT		90	1.7E-12	7.8E-10
<input type="checkbox"/>	GOTERM_MF_DIRECT	DNA binding	RT		69	2.2E-9	5.0E-7
<input type="checkbox"/>	GOTERM_MF_DIRECT	transcription factor activity, sequence-specific DNA binding	RT		42	2.6E-6	4.0E-4
Annotation Cluster 2		Enrichment Score: 1.06	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_MF_DIRECT	RNA polymerase II core promoter proximal region sequence-specific DNA binding	RT		18	1.9E-3	2.0E-1
<input type="checkbox"/>	GOTERM_MF_DIRECT	transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	RT		6	5.8E-1	1.0E0
<input type="checkbox"/>	GOTERM_MF_DIRECT	chromatin binding	RT		9	5.9E-1	1.0E0

Cellular Component Clustering

Annotation Cluster 1		Enrichment Score: 0	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_CC_DIRECT	cell surface	RT		7	9.8E-1	1.0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	integral component of plasma membrane	RT		19	9.9E-1	1.0E0
<input type="checkbox"/>	GOTERM_CC_DIRECT	plasma membrane	RT		46	1.0E0	1.0E0

Bioinformatics Analysis

- Mapping to Genome
- Find Differentially Expressed Genes
- GO Annotation
- **Pathway Analysis**
- WDR36 and F11R

Panther Pathway Analysis



Home About PANTHER Data PANTHER Tools Workspace Downloads Help/Tutorial

Now includes comprehensive GO annotations directly imported from the GO database

Search

All

Quick links

[Whole genome function views](#)

[Genome statistics](#)

[Data Version](#)

[How to cite PANTHER](#)

NEW! [Recent publication describing PANTHER](#)

News

Release of PANTHER-PSEP tool for scoring non synonymous genetic variants.

[Click for additional info.](#)

Gene List Analysis

Browse

Sequence Search

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

Help Tips

Steps:

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

1.

Enter ids and or select file for batch upl or list from workspace for comparing to

Enter IDs: [Supported IDs](#)

Upload IDs: [File format](#)

未选择任何文件

Please [login](#) to be able to select lists from

Select List Type:

- ID List
- Previously exported text search result
- Workspace list
- PANTHER Generic Mapping File

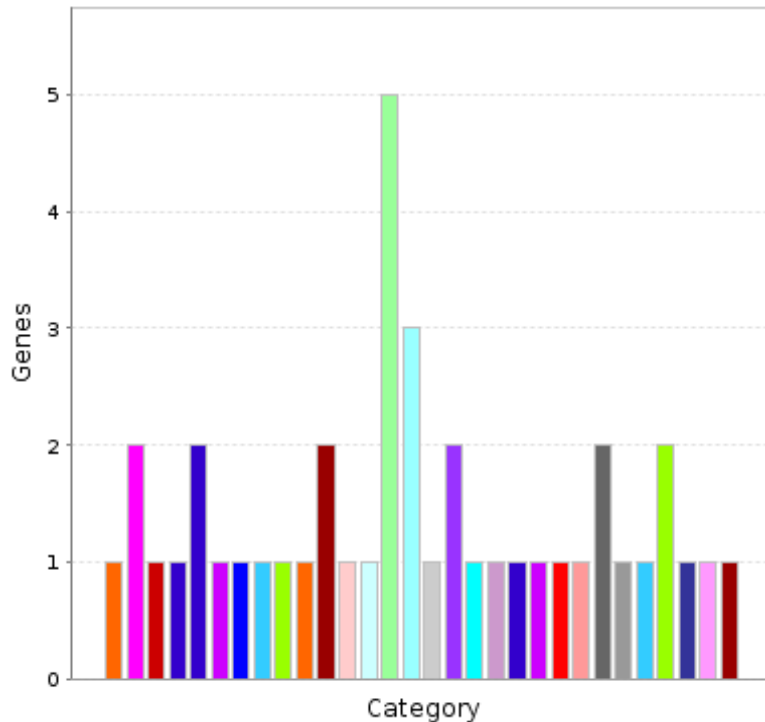
2. Select organism.

Homo sapiens
Mus musculus
Rattus norvegicus

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

PANTHER Pathway

Total # Genes: 184 Total # pathway hits: 42



Click to get gene list for a category:

- [Angiogenesis \(P00005\)](#)
- [Apoptosis signaling pathway \(P00006\)](#)
- [B cell activation \(P00010\)](#)
- [Blood coagulation \(P00011\)](#)
- [CCKR signaling map \(P06959\)](#)
- [Cholesterol biosynthesis \(P00014\)](#)
- [De novo purine biosynthesis \(P02738\)](#)
- [EGF receptor signaling pathway \(P00018\)](#)
- [Endothelin signaling pathway \(P00019\)](#)
- [Flavin biosynthesis \(P02741\)](#)
- [General transcription regulation \(P00023\)](#)
- [Huntington disease \(P00029\)](#)
- [Hypoxia response via HIF activation \(P00030\)](#)
- [Inflammation mediated by chemokine and cytokine signaling pathway \(P00031\)](#)
- [Integrin signalling pathway \(P00034\)](#)
- [Interferon-gamma signaling pathway \(P00035\)](#)
- [Interleukin signaling pathway \(P00036\)](#)
- [Muscarinic acetylcholine receptor 2 and 4 signaling pathway \(P00043\)](#)

Pathway:

- Inflammation mediated by chemokine and cytokine signaling pathway
- Integrin signaling pathway

KOBAS Pathway Analysis

KOBAS 2.0

Run KOBAS 2.0

Annotate

Identify

Annotate + Identify

Advanced KOBAS

Login (free registration)

User Space

Analysis History

Download

Download

Help

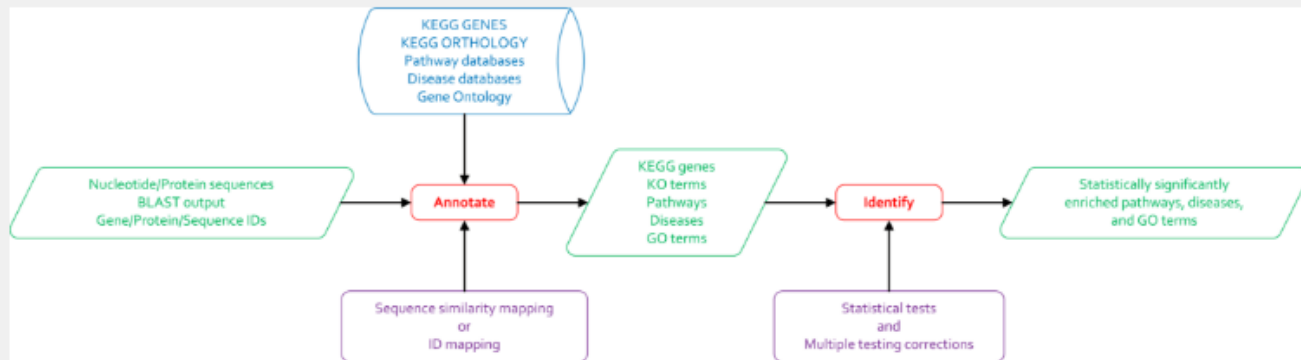
Tutorial

Feedback

Contact

Welcome to KOBAS 2.0

✦ KOBAS 2.0 is an update of KOBAS (**KEGG Orthology Based Annotation System**). It can identify statistically significantly enriched pathways, human diseases, and functional terms for an input set of genes using biological knowledge from well-known pathway databases, disease databases, and Gene Ontology.



Updates of KOBAS 2.0

- ✦ [Annotate + Identify](#) is a new program of KOBAS 2.0. It combines [Annotate](#) and [Identify](#).
- ✦ KOBAS 2.0 integrates [Gene Ontology](#).
- ✦ KOBAS 2.0 integrates [OMIM](#), [KEGG DISEASE](#), [FunDO](#), [GAD](#), and [NHGRI GWAS Catalog](#) disease databases.
- ✦ KOBAS 2.0 integrates [KEGG PATHWAY](#), [PID](#), [BioCarta](#) (from PID), [Reactome](#), [BioCyc](#), and [PANTHER](#) pathway databases.
- ✦ KOBAS 2.0 can annotate queries to either KEGG genes or KEGG Orthology (KO) terms.

<http://kobas.cbi.pku.edu.cn/home.do>

Download... HELP

RAW CONTENT

TABLE VIEW (FOR PATHWAY IDENTIFICATION RESULT)

TABLE VIEW (FOR DISEASE IDENTIFICATION RESULT)

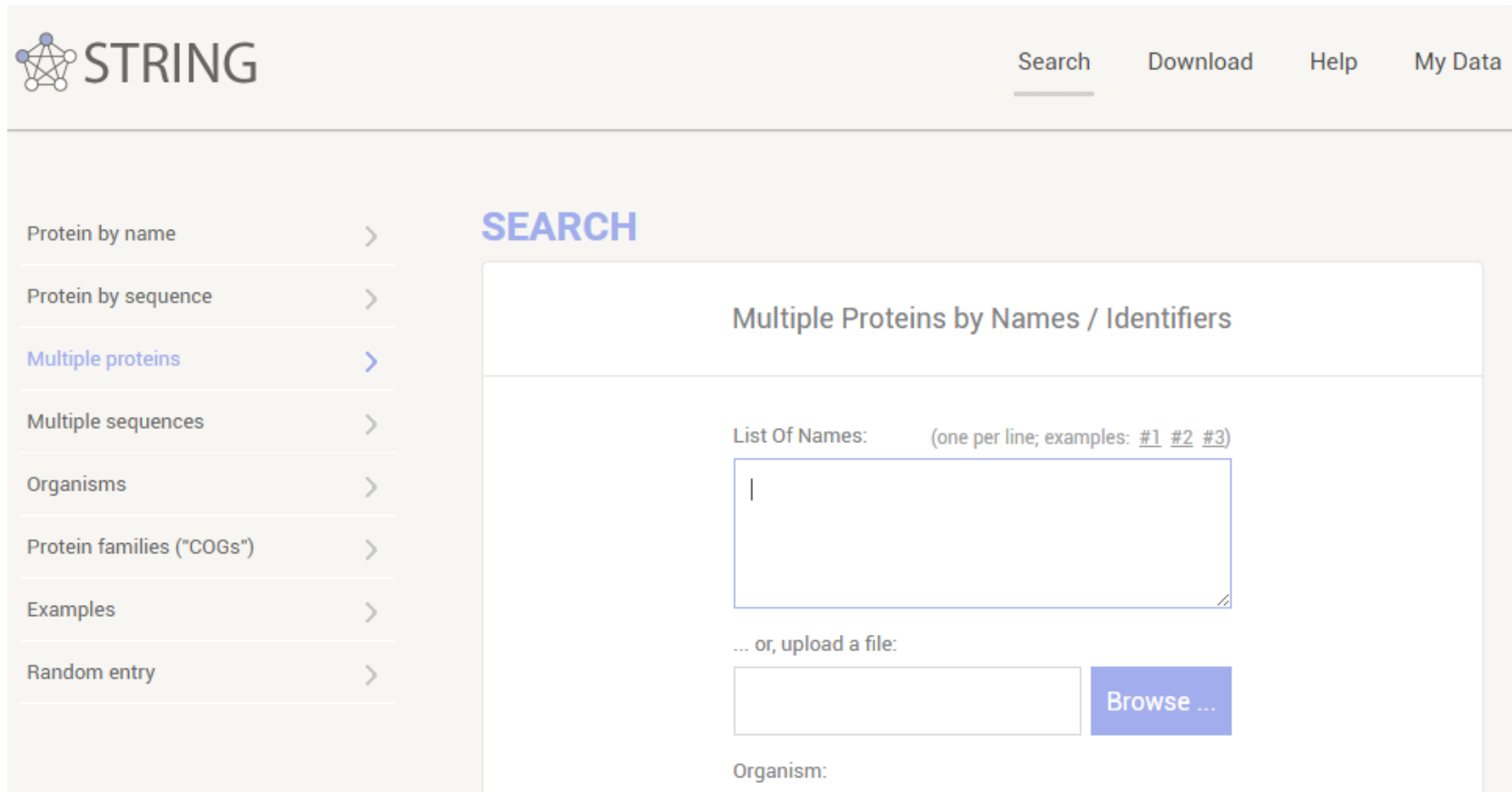
Term	Database	ID	Input number	Background n...	P-Value	Corrected P-Value
Phagosome	KEGG PATHW...	hsa04145	8	153	0.000189071571576	0.0236339464469
Staphylococcus aureus infection	KEGG PATHW...	hsa05150	4	55	0.00263041537904	0.104974032568
Herpes simplex infection	KEGG PATHW...	hsa05168	7	184	0.00276944351486	0.104974032568
Cell adhesion molecules (CAMs)	KEGG PATHW...	hsa04514	6	142	0.00335916904216	0.104974032568
Complement and coagulation ca...	KEGG PATHW...	hsa04610	4	69	0.00565941958253	0.140893338375
Allograft rejection	KEGG PATHW...	hsa05330	3	37	0.00692096422013	0.140893338375
Antigen processing and presenta...	KEGG PATHW...	hsa04612	4	77	0.00814541616932	0.140893338375
Graft-versus-host disease	KEGG PATHW...	hsa05332	3	41	0.00901717365603	0.140893338375
Type I diabetes mellitus	KEGG PATHW...	hsa04940	3	43	0.0101895043773	0.141520894129
Autoimmune thyroid disease	KEGG PATHW...	hsa05320	3	52	0.0165224431945	0.191380787942
Viral carcinogenesis	KEGG PATHW...	hsa05203	6	205	0.0176308829716	0.191380787942
Legionellosis	KEGG PATHW...	hsa05134	3	55	0.0190253043239	0.191380787942
Starch and sucrose metabolism	KEGG PATHW...	hsa00500	3	56	0.019903601946	0.191380787942
Viral myocarditis	KEGG PATHW...	hsa05416	3	58	0.0217263523418	0.193985288766
Leishmaniasis	KEGG PATHW...	hsa05140	3	72	0.0369476397705	0.307896998088
Pertussis	KEGG PATHW...	hsa05133	3	75	0.040762855212	0.318459806344
Tight junction	KEGG PATHW...	hsa04530	4	138	0.0505091048619	0.352809471802
Epstein-Barr virus infection	KEGG PATHW...	hsa05169	5	200	0.0508045639394	0.352809471802

Displaying 1 - 20 of 125 Page 1 of 7

Pathway:

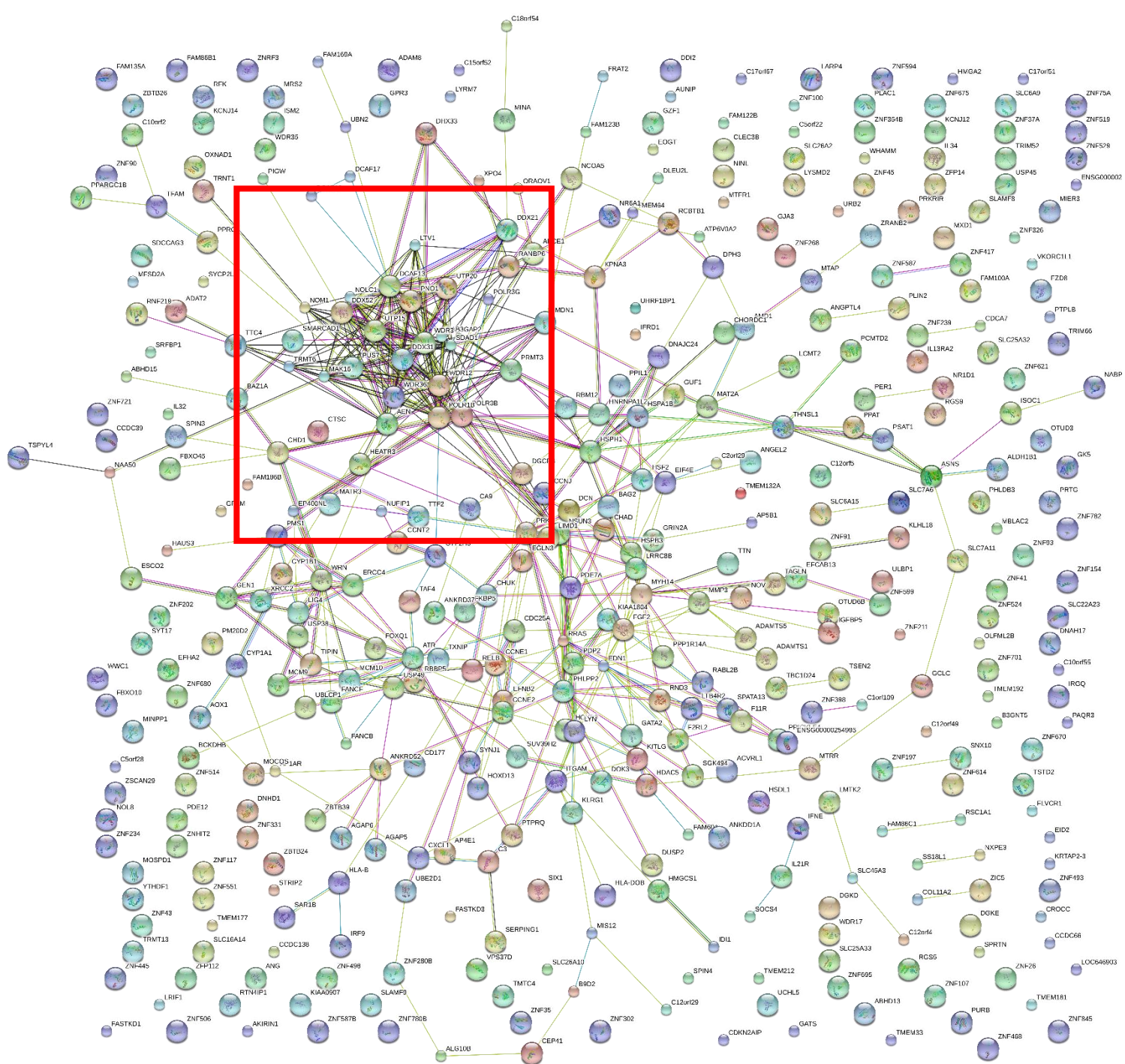
Phagosome, Staphylococcus aureus infection,
Herpes simplex infection, Cell adhesion molecules

STRING Analysis



The screenshot shows the STRING database search interface. At the top left is the STRING logo, and at the top right are navigation links for Search, Download, Help, and My Data. A left sidebar contains a menu with options: Protein by name, Protein by sequence, Multiple proteins (highlighted in blue), Multiple sequences, Organisms, Protein families ("COGs"), Examples, and Random entry. The main content area is titled "SEARCH" and "Multiple Proteins by Names / Identifiers". It features a text input field for "List Of Names" with a placeholder "(one per line; examples: #1 #2 #3)", a file upload option "... or, upload a file:" with a "Browse ..." button, and an "Organism:" label with an empty input field.

<http://string-db.org/cgi/input.pl>



Bioinformatics Analysis

- Mapping to Genome
- Find Differentially Expressed Genes
- GO Annotation
- Pathway Analysis
- **WDR36 and F11R**

➤ WDR36

RNA metabolism

Key point in STRING

➤ F11R

Cell membrane

Cell adhesion molecular

WDR 36

BLAST Align Retrieve/ID mapping Help Contact

UniProtKB - Q8NI36 (WDR36_HUMAN) Basket 23

Display BLAST Align Format Add to basket History Feedback Help video Other tutorials and videos

Entry
Feature viewer
Feature table

Protein | **WD repeat-containing protein 36**
Gene | **WDR36**
Organism | *Homo sapiens (Human)*
Status | Reviewed - Annotation score: - Experimental evidence at protein levelⁱ

Function None

Entry informationⁱ

Entry name ⁱ	WDR36_HUMAN
Accession ⁱ	Primary (citable) accession number: Q8NI36 Secondary accession number(s): A2RUS4, Q68E02, Q8N1Q2
Entry history ⁱ	Integrated into UniProtKB/Swiss-Prot: April 26, 2004 Last sequence update: October 1, 2002 Last modified: June 8, 2016 This is version 127 of the entry and version 1 of the sequence. [Complete history]
Entry status ⁱ	Reviewed (UniProtKB/Swiss-Prot)
Annotation program	Chordata Protein Annotation Program
Disclaimer	Any medical or genetic information present in this entry is provided for research, educational and informational purposes only. It is not in any way intended to be used as a substitute for professional medical advice, diagnosis, treatment or care.

Function: RNA metabolism

Functionⁱ

Involved in the nucleolar processing of SSU 18S rRNA. Involved in T-cell activation and highly coregulated with IL2.

📄 1 Publication ▾

GO - Molecular functionⁱ

- [poly\(A\) RNA binding](#) 📄 Source: UniProtKB ▾

GO - Biological processⁱ

- [regulation of axon extension](#) 📄 Source: Ensembl
- [response to stimulus](#) 📄 Source: UniProtKB-KW
- [retina homeostasis](#) 📄 Source: Ensembl
- [rRNA processing](#) 📄 Source: Reactome
- [visual perception](#) 📄 Source: UniProtKB-KW

[Complete GO annotation...](#)

Keywords - Biological processⁱ

[Ribosome biogenesis](#), [rRNA processing](#), [Sensory transduction](#), [Vision](#)

Enzyme and pathway databases

Reactome ⁱ	R-HSA-6790901 . rRNA modification in the nucleus. R-HSA-6791226 . Major pathway of rRNA processing in the nucleolus.
Signalink ⁱ	Q8NI36 .

Subcellular Location: nucleolus and nucleoplasm

Subcellular locationⁱ

- [Nucleus](#) > [nucleolus](#) 2 Publications

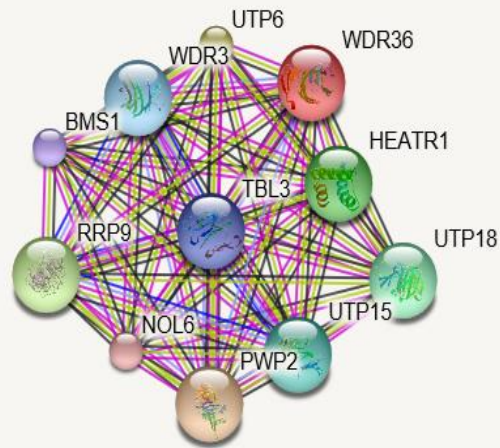
GO - Cellular componentⁱ

- [nucleolus](#) Source: HPA
- [nucleoplasm](#) Source: Reactome
- [small-subunit processome](#) Source: InterPro

[Complete GO annotation...](#)

Keywords - Cellular componentⁱ

[Nucleus](#)



[Legend](#) [Data Settings](#) [View Settings](#) [Tables / Exports](#) [Evidence](#) [Analysis](#) 

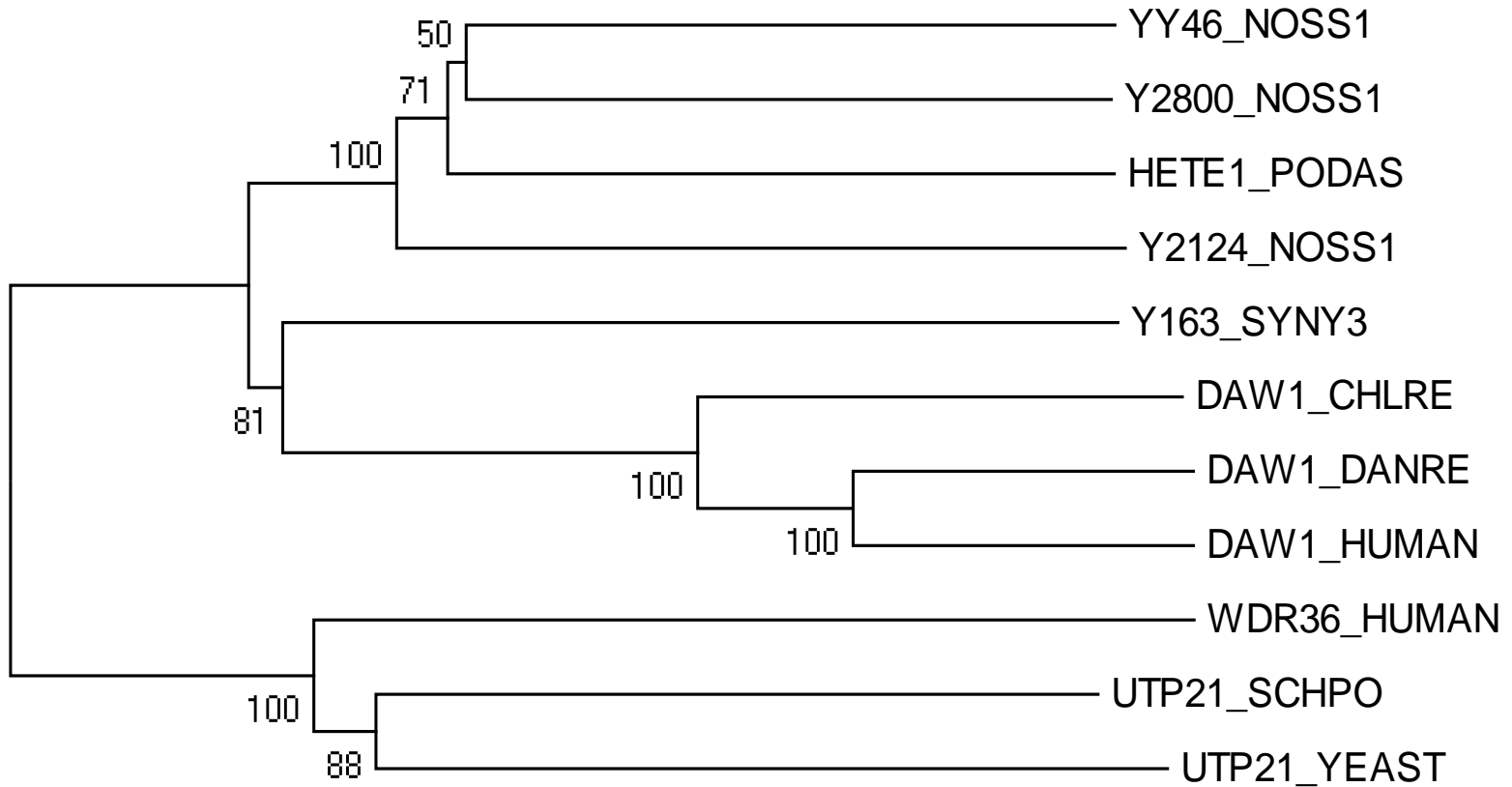
Nodes:

Network nodes represent proteins
splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the

Node Size
small nodes: protein of unknown 3D structure

Node Color
colored nodes: query proteins and first shell of interactors

Phylogenetic Tree






0.1

F11R







Subcellular Location: cell membrane

Subcellular locationⁱ








Topology

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Transmembrane ⁱ	236 - 261	26	Helical  Sequence analysis 		

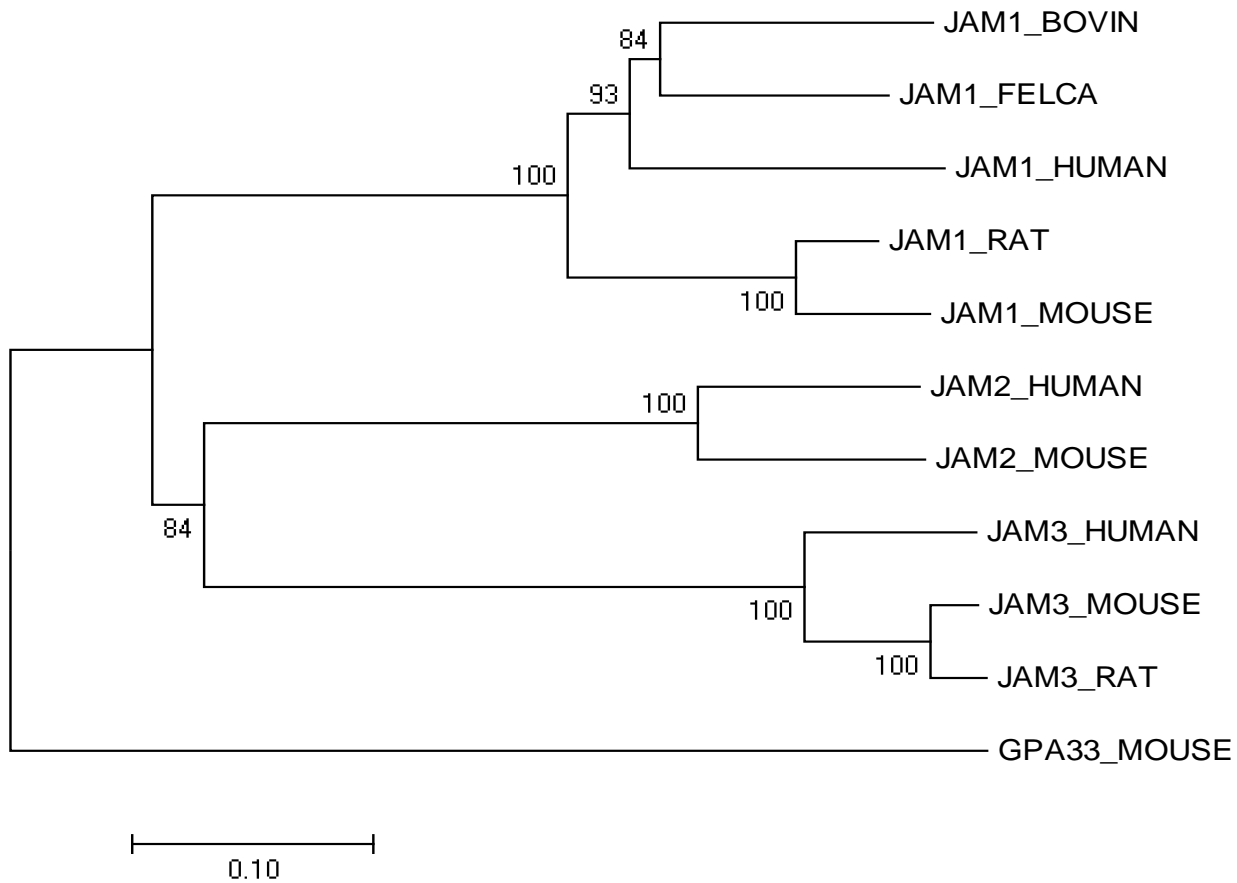
GO - Cellular componentⁱ

- bicellular tight junction  Source: Ensembl
- cell junction  Source: HPA
- cytoplasmic vesicle  Source: Ensembl
- integral component of membrane  Source: UniProtKB-KW
- microtubule cytoskeleton  Source: HPA
- slit diaphragm  Source: Ensembl

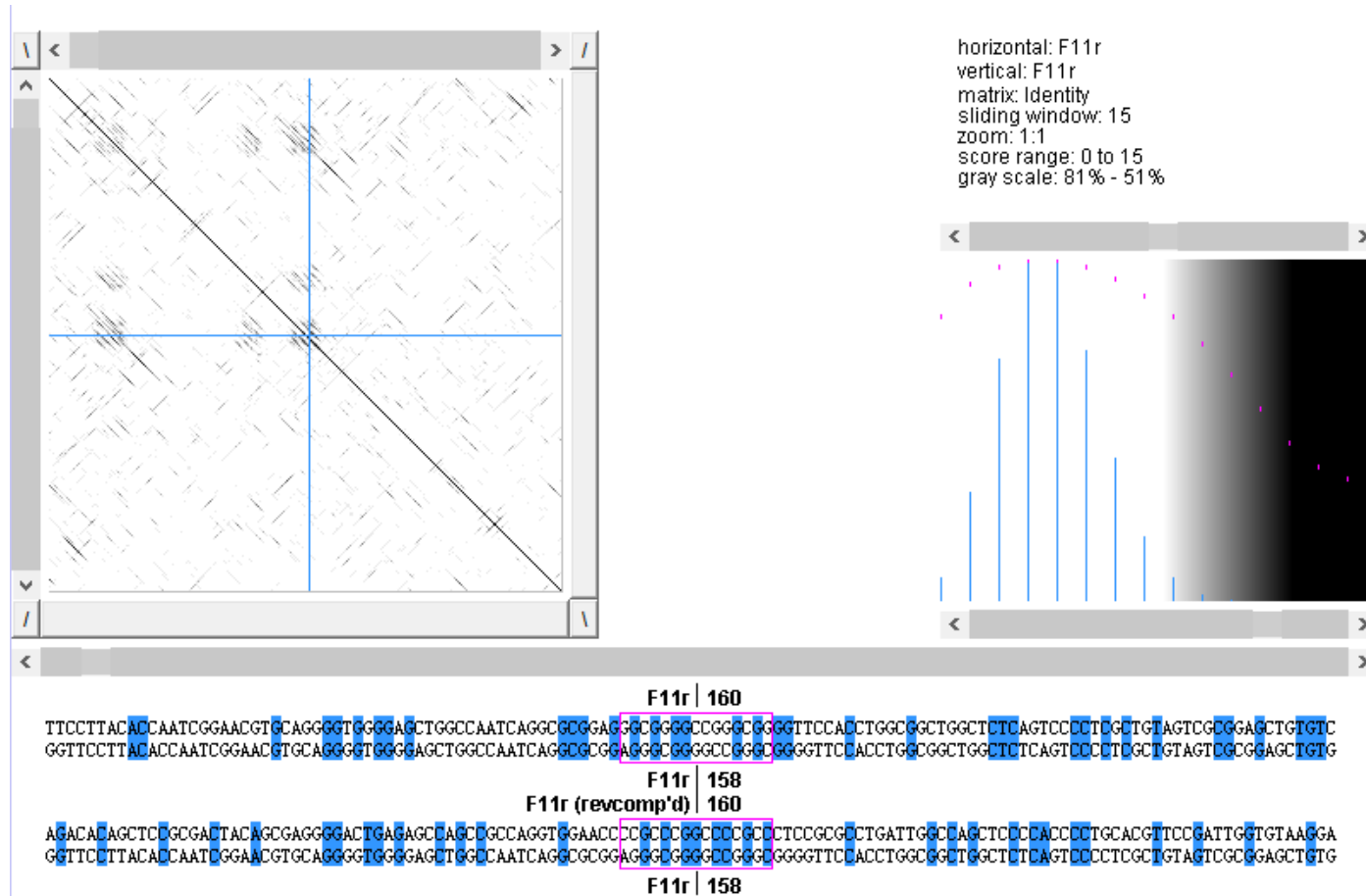
Function

- cell adhesion  Source: Ensembl
- epithelial cell differentiation  Source: Ensembl
- intestinal absorption  Source: Ensembl
- positive regulation of blood pressure  Source: Ensembl
- regulation of cytokine production  Source: Ensembl
- regulation of membrane permeability  Source: Ensembl
- response to radiation  Source: Ensembl

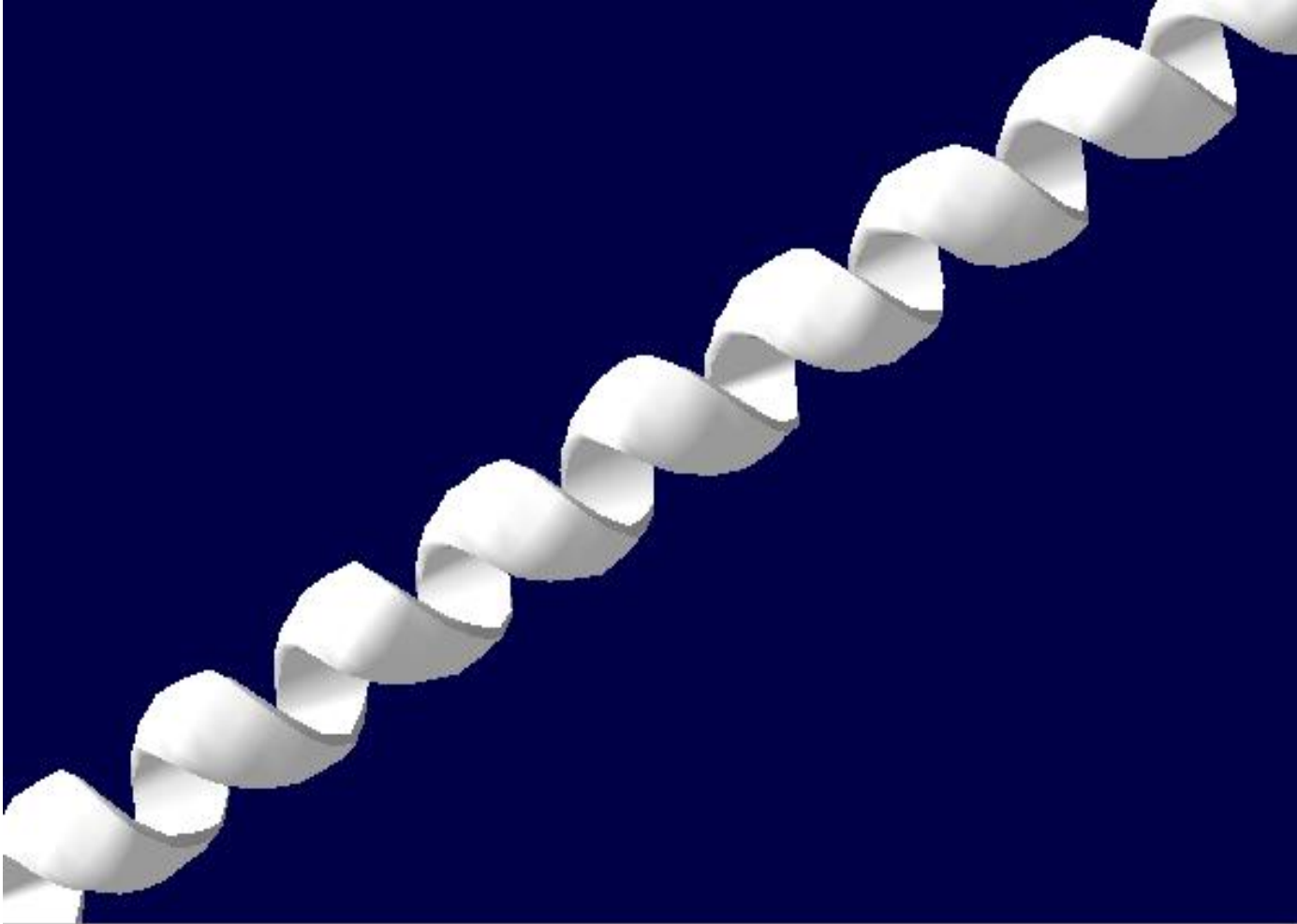
Phylogenetic Tree



Dotlet analysis



Sturcture



Contribution

崔轶男： Mapping to Genome

李铠： Find Differentially expressed genes, GO annotate, Pathway analysis

赵璐璐： WDR36 analysis

李晓宇： F11R analysis

Thank you for your attention!