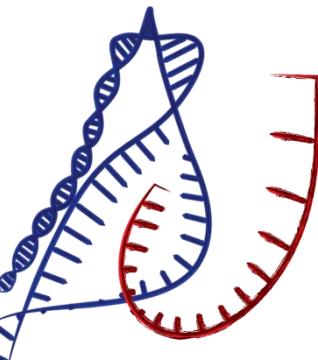


---

# Online Functional Resources

---

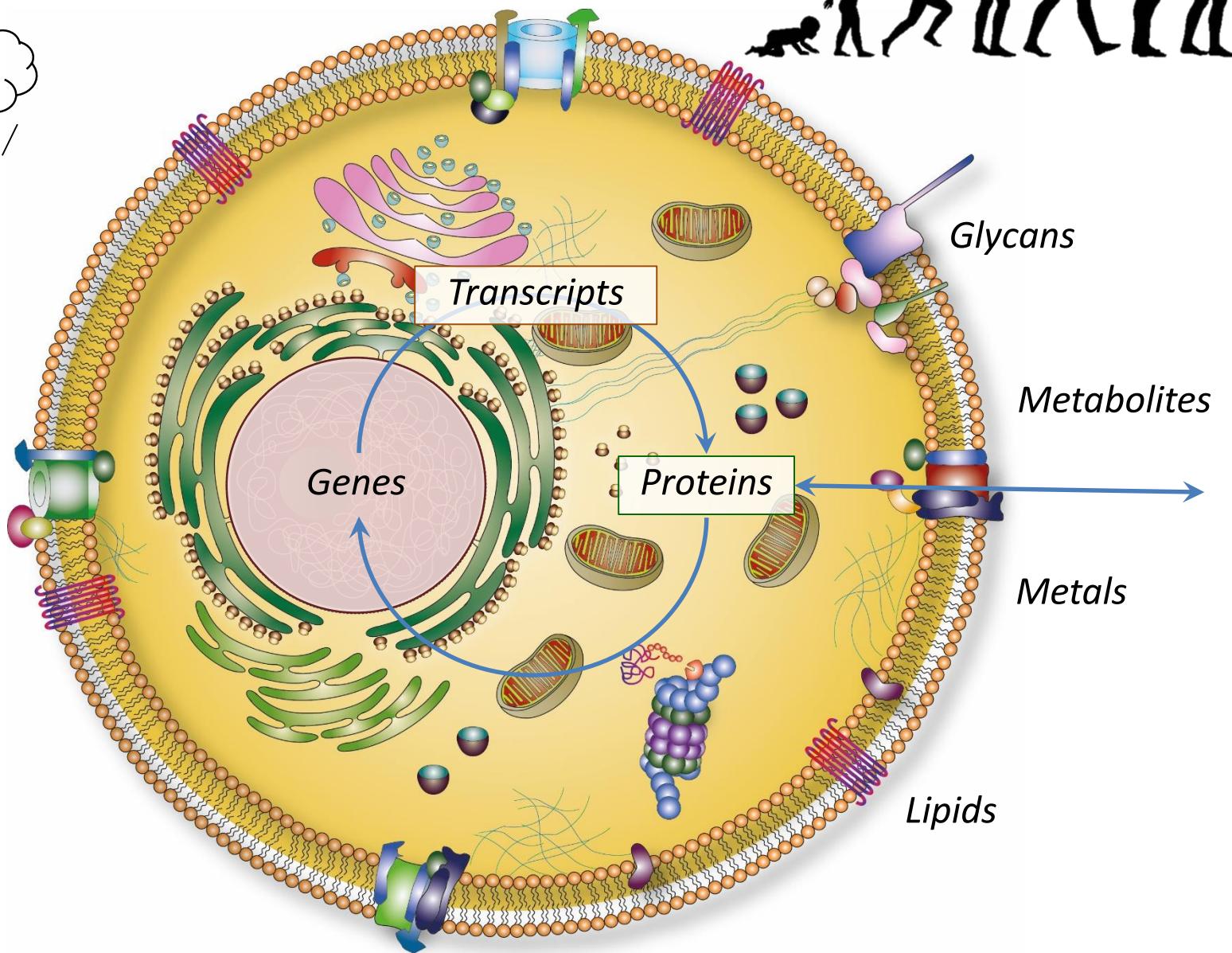
Marc Vaudel



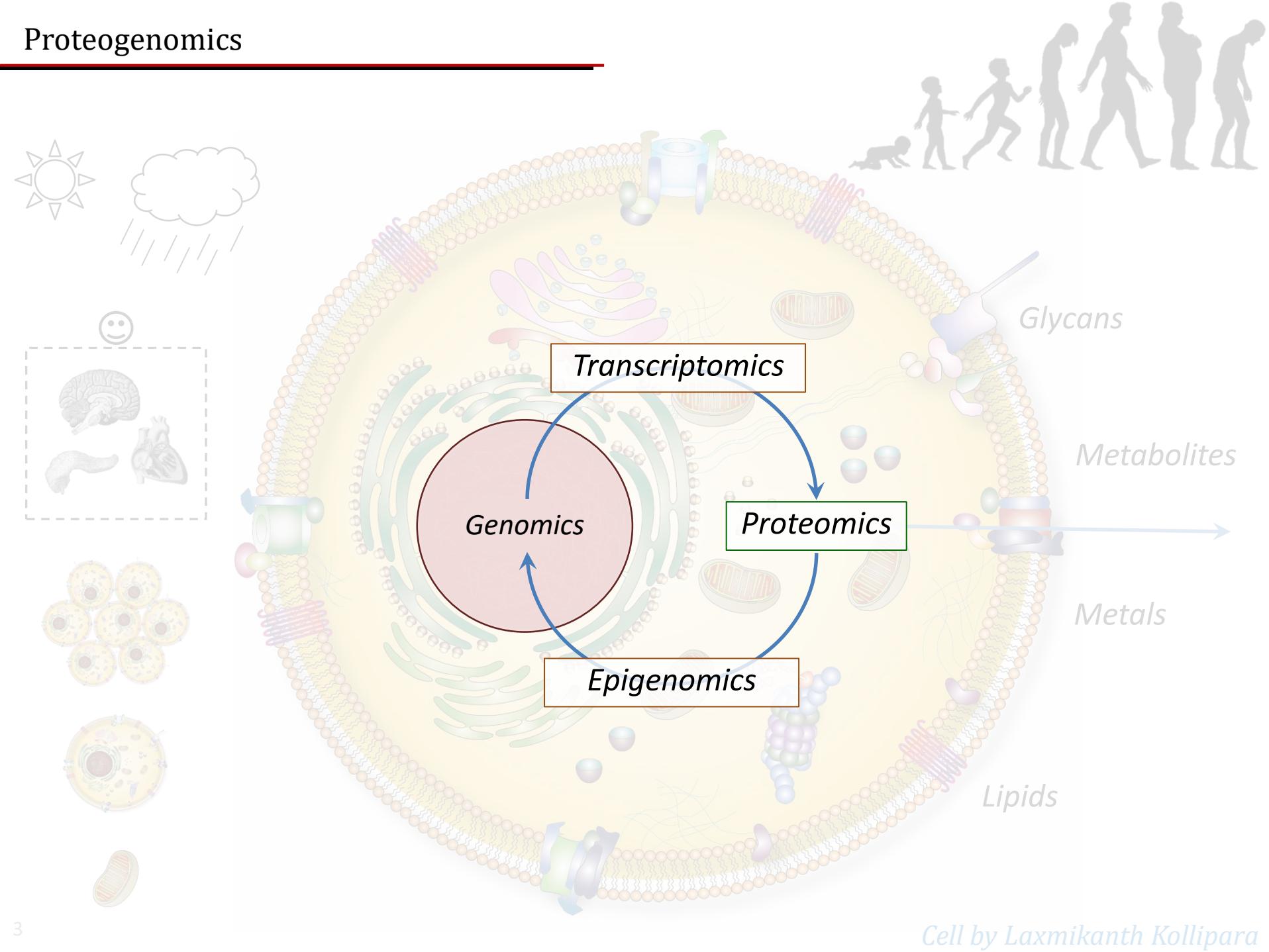
*Center for Medical Genetics and Molecular Medicine,  
Haukeland University Hospital, Bergen, Norway*

*KG Jebsen Center for Diabetes Research, Department of Clinical Science,  
University of Bergen, Norway*

# The Omics



# Proteogenomics

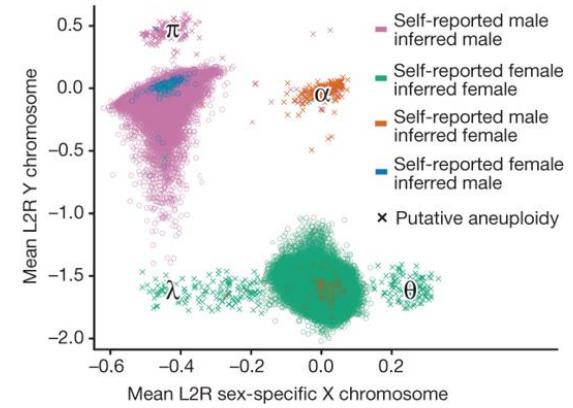
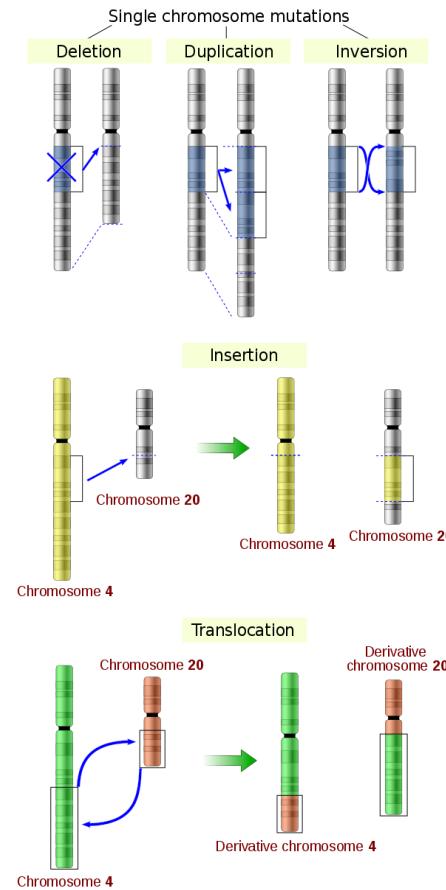
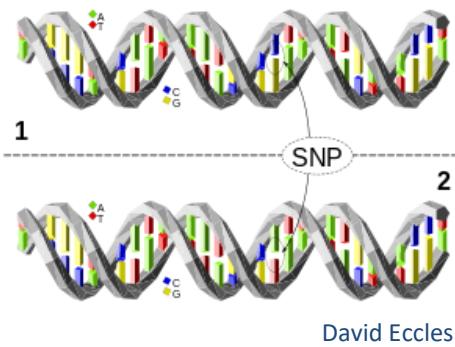


# What makes us different?

## Genetic variation

### 1- Scale of the variation?

- Small scale sequence variation, e.g. Single Nucleotide Polymorphism
- Large scale variation, e.g. copy number change
- Chromosome scale variation, e.g. chromosome duplication



$\lambda = X0$  (or mosaic XX/X0),  
 $\theta = XXX$ ,  $\alpha = XXY$ , and  $\pi = XYY$ .  
UK Biobank

# What makes us different?

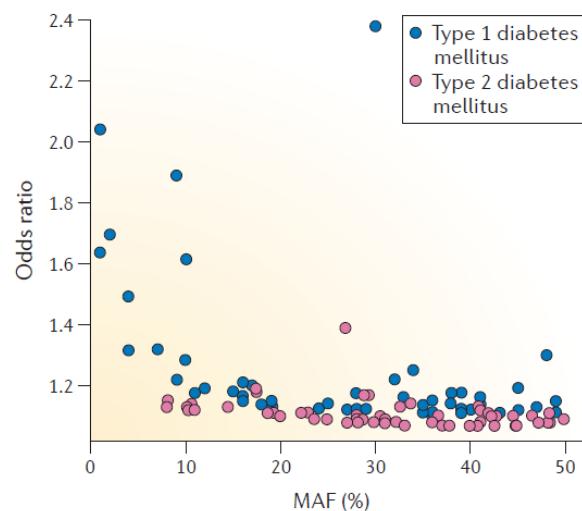
---

## Genetic variation

1- Scale of the variation?

2- Frequency in the population?

- Common variation -> complex traits
- Rare Variation -> mendelian



# What makes us different?

---

## Genetic variation

1- Scale of the variation?

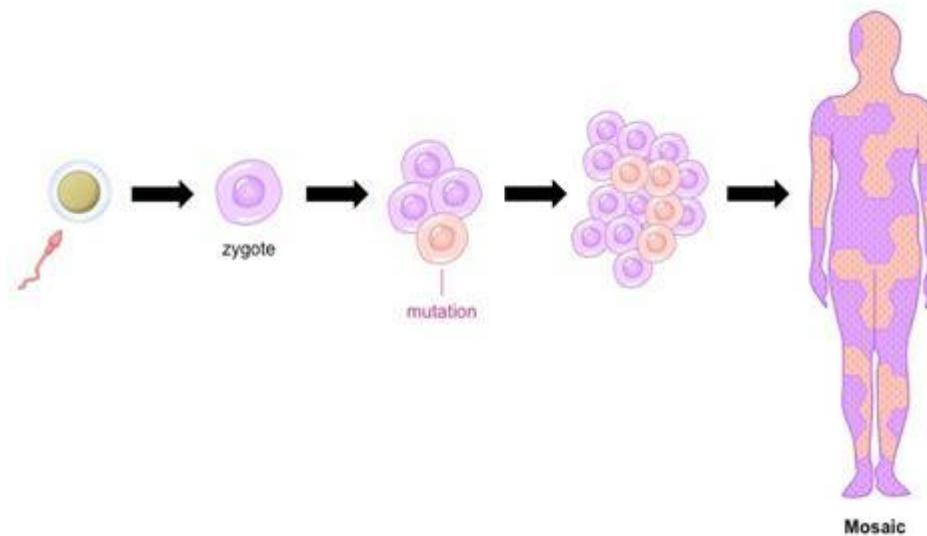
2- Frequency in the population?

3- History?

- *Ethnicity: bottleneck in ancestries*

- *Germline: germ cells*

- *Somatic: only specific cells*



# What makes us different?

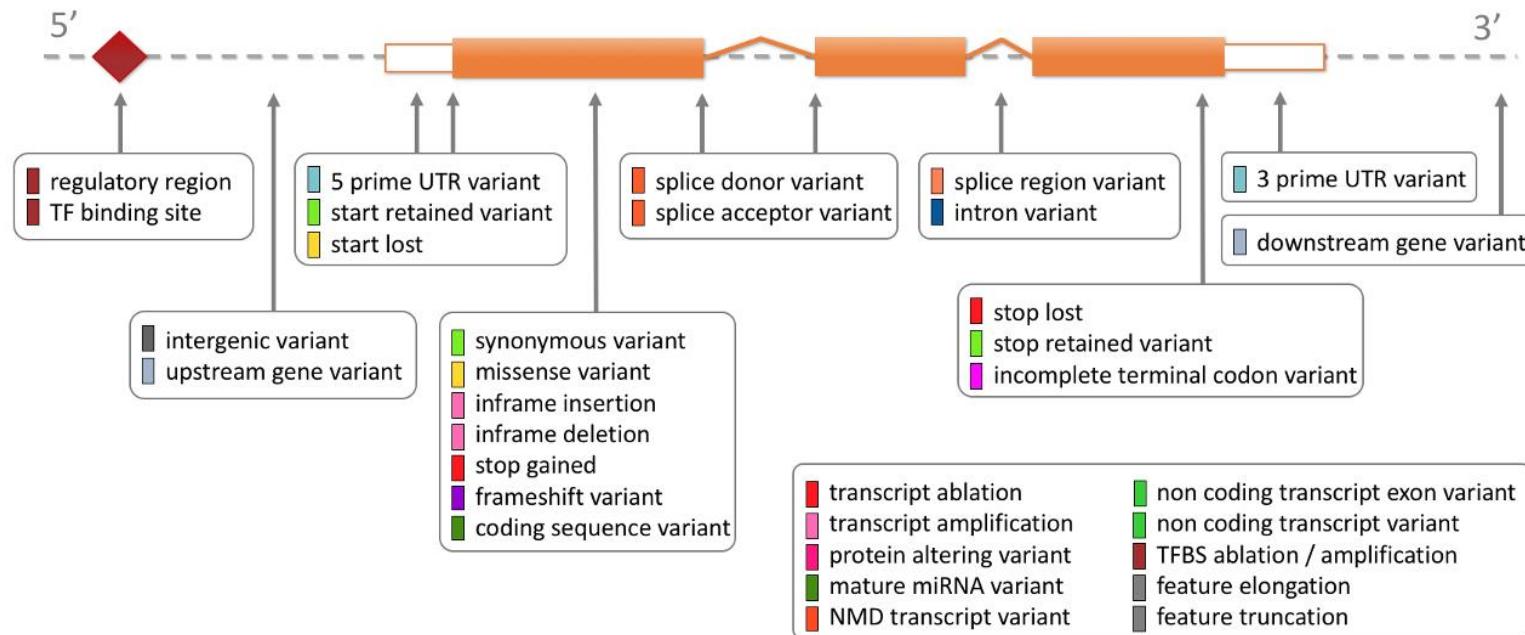
## Genetic variation

1- Scale of the variation?

2- Frequency in the population?

3- History?

4- Genomic location?



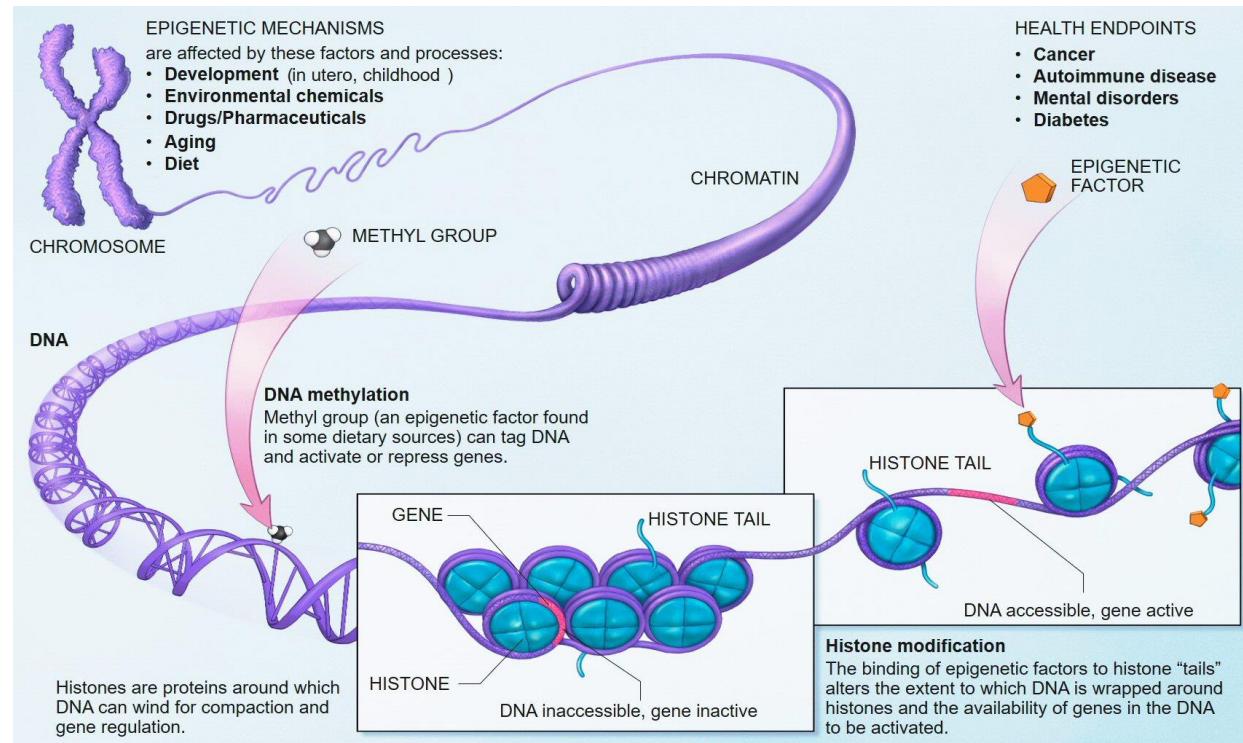
# What makes us different?

## Genetic variation

- 1- Scale of the variation?
- 2- Frequency in the population?
- 3- History?
- 4- Genomic location?

## Epigenetic variation

- 1- Multiple mechanisms
- 2- Affects gene expression



# What makes us different?

---

## *Genetic variation*

- 1- *Scale of the variation?*
- 2- *Frequency in the population?*
- 3- *History?*
- 4- *Genomic location?*

## *Epigenetic variation*

- 1- *Multiple mechanisms*
- 2- *Affects gene expression*

## *Transcriptomic variation*

- 1- *RNA translation modulation*
- 2- *Remains quite unexplored*

# What makes us different?

## Genetic variation

- 1- Scale of the variation?
- 2- Frequency in the population?
- 3- History?
- 4- Genomic location?

## Epigenetic variation

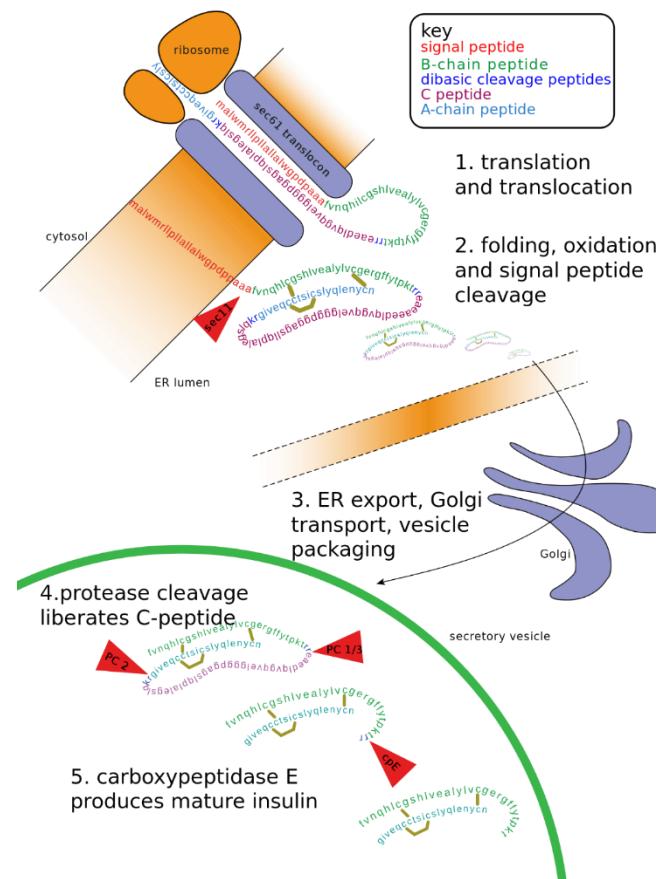
- 1- Multiple mechanisms
- 2- Affects gene expression

## Transcriptomic variation

- 1- RNA translation modulation
- 2- Remains quite unexplored

## Proteomic variation

- 1- Alternative splicing
- 2- Post-translational modifications
- 3- cleavage and folding



# What makes us different?

## Genetic variation

- 1- Scale of the variation?
- 2- Frequency in the population?
- 3- History?
- 4- Genomic location?

## Proteomic variation

- 1- Alternative splicing
- 2- Post-translational modifications
- 3- cleavage and folding

## Epigenetic variation

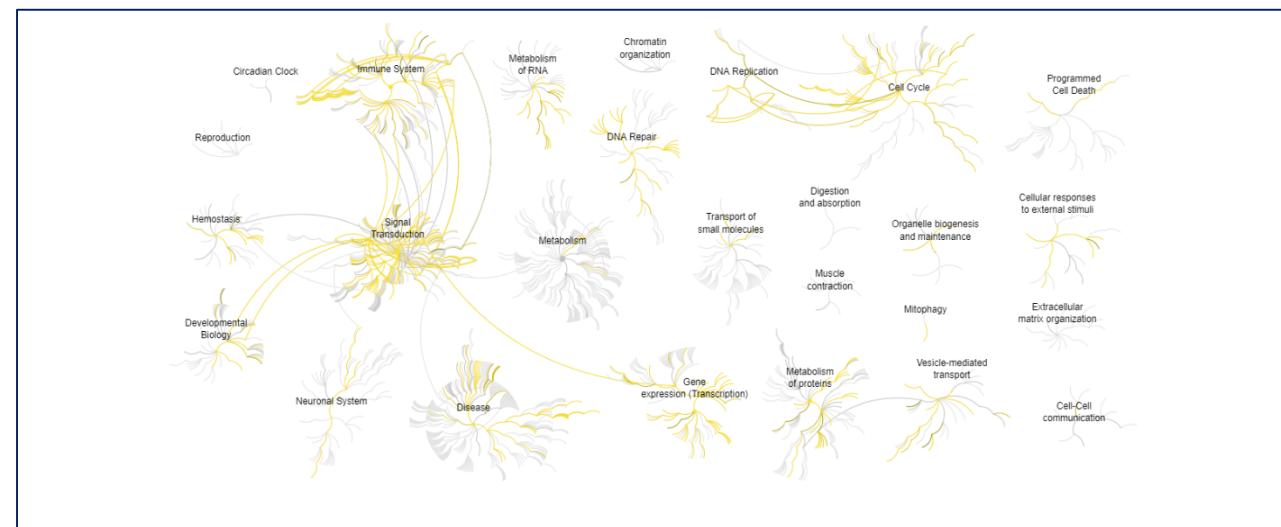
- 1- Multiple mechanisms
- 2- Affects gene expression

## Interactions and pathways

- 1- Alternative splicing
- 2- Post-translational modifications
- 3- cleavage and folding

## Transcriptomic variation

- 1- RNA translation modulation
- 2- Remains quite unexplored



# Different Resources for Different Questions

## From Genes to Proteins



Ensembl



NCBI



GenomeNet



Genome Browser

## Metabolites and Drugs



ChEBI - ChEMBL

## Proteins



## Interactions and Pathways



## Resources and Tools

The screenshot shows the Ensembl homepage with a red dashed box highlighting the main content area. At the top, there's a search bar with placeholder text "Search: All species for" and a "Go" button. Below it, a search example "e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease" is shown.

**Browse a Genome:** A section describing Ensembl as a genome browser for vertebrate genomes, mentioning comparative genomics, evolution, sequence variation and transcriptional regulation. It includes links to Human (GRCh38.p7), Mouse (GRCm38.p4), Human (GRCh37), Zebrafish (GRCz10), and a "Edit favourites" link.

**Favourite genomes:** Buttons for Human (GRCh38.p7), Human (GRCh37), Mouse (GRCm38.p4), and Zebrafish (GRCz10).

**All genomes:** A dropdown menu labeled "Select a species" and a link to "View full list of all Ensembl species".

**Did you know...?** A section about Ensembl's collaboration with major genomics projects like Blueprint, NextGen, Quantomics, HEROIC and the Neanderthal browser.

**Still using Human GRCh37?** A link to "Go to GRCh37".

**Variant Effect Predictor (VeP):** A tool for predicting the effect of variants on gene function.

**Gene expression in different tissues:** An image of a tissue sample.

**Find SNPs and other variants for my gene:** A sequence diagram showing SNPs: GGTATACATT CTTAAAGCTT CTTCATTAT GAAACATTTC.

**Retrieve gene sequence:** A sequence diagram: GCTCTAATCTCCGGGTTGC GGGCTTGTTGGCGGGCGG AGGGGGACAGATTGTTG GACCTCTGGGGCGGGTTG.

**Compare genes across species:** An image of two different species' genomes being compared.

**Use my own data in Ensembl:** A small image of a DNA sequence.

**ENCODE data in Ensembl:** An image of the ENCODE logo.

**Ensembl supports data from external projects through Track hubs:** A small image of a globe.

**What's New in Ensembl Release 86 (October 2016):** A list of updates including new mouse strains, chicken assembly, macaque assembly, mouse lemur assembly, and a Zebrafish update. Links to "Full details", "All web updates by release", and "More news on our blog".

**Tweets:** Two tweets from the @ensembl account. The first tweet discusses bottlenose dolphin immune genes, and the second tweet discusses an optimal transcriptome assembly strategy for non-model organisms.

**Footer:** Logos for EMBL-EBI and Wellcome Trust Sanger Institute, a statement about Ensembl being a joint project between them, and funding information. There are also links for "How to cite Ensembl" and "Permanent link - View in archive site".

# Training and Help

Using this website Annotation and prediction Data access API & software About us

Help & Documentation

## Try the tutorials!

**Using this website**

Our website offers lots of ways to view and interact with our genomic data - find out more!

- Adding custom tracks
- Tutorials
- Bugs
- What's New
- FAQs (Frequently Asked Questions)

[More...](#)

**Annotation & Prediction**

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

- Ensembl annotation
- Variation data
- Comparative genomics
- Regulatory build

[More...](#)

**Data access**

All of our data is open-access and can be downloaded free of charge ([disclaimer](#)). Ways to access this data include:

- Export features or sequence directly from web pages
- Extract data from our [public database](#) using Perl scripts
- Data-mining using the [BioMart](#) tool
- [FTP](#) download of complete datasets
- Access data from our [REST server](#) using any Programming language

[More...](#)

**API & Software**

Ensembl releases all its software under an Apache-style open source [licence](#). Our products include:

- [Perl API](#) for direct data access
- [REST](#) server for language agnostic access
- [Virtual machine](#) preloaded with API
- [Variant Effect Predictor](#) (VEP) and other command-line scripts
- [ehive](#) distributed processing system
- [Web frontend](#) (Apache with mod\_perl) for mirroring this website

[More...](#)

**About us**

Ensembl is a joint project between [EMBL-EBI](#) and the [Wellcome Trust Sanger Institute](#) to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.

[More...](#)

Ensembl release 86 - Oct 2016 © [WTSI](#) / [EMBL-EBI](#)

[Permanent link](#)

Get help [About us](#) [Follow us](#)

## Responsive help desk!

## Different species covered

Note: Ensembl was primarily designed for human and model organisms. The level of annotation varies between species. Other species can be accessed at [ensemblgenomes.org](http://ensemblgenomes.org).

# Integrated resources for human

**e|Ensembl** BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh38.p7) ▾

**Human**  
*Homo sapiens*

Search all categories ▾

Search Human... Go

e.g. BRCA2 or 17:63973115-64437414 or rs1333049 or osteoarthritis

Genome assembly: GRCh38.p7 (GCA\_000001405.22)

- More information and statistics
- Download DNA sequence (FASTA)
- Convert your data to GRCh38 coordinates
- Display your data in Ensembl

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart ▾ Go

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

- More about comparative analysis
- Download alignments (EMF)

Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.

- More about the Ensembl regulatory build and microarray annotation
- Experimental data sources
- Download all regulatory features (GFF)

Human (GRCh38.p7) ▾

What's New in Human release 86

- Human: updated cDNA alignments
- Human: updated RefSeq gene import
- External database references update

More news...

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

- More about this genebuild, including RNASeq gene expression models
- Download genes, cDNAs, ncRNA, proteins (FASTA)
- Update your old Ensembl IDs

Vega\* Additional manual annotation can be found in Vega

Pax6 INS FOXP2 BRCA2 DMD ssh Example gene

Example transcript

Variation

What can I find? Short sequence variants and longer structural variants, disease and other phenotypes

- More about variation in Ensembl
- Download all variants (GVF)
- Variant Effect Predictor

VeIP

ATCGAGCT ATCCAGCT ATCGAGAT Example variant

Example phenotype

Example structural variant

# Example with AQP4 Gene

Human (GRCh38.p1) ▾ Location: 18:26,852,038-26,865,818 Gene: AQP4

**Gene: AQP4** ENSG00000171885

Description aquaporin 4 [Source HGNC Symbol,Acc:HGNC:637] WCH4, MIWC

Synonyms

Location Chromosome 18: 26,852,038-26,865,818 reverse strand. GRCh38:CM000680.2

About this gene This gene has 8 transcripts ([splice variants](#)), [71 orthologues](#), [7 paralogues](#), is a member of [1 Ensembl protein family](#) and is associated with [7 phenotypes](#).

Transcripts [Show transcript table](#)

**Summary** ⓘ

Name [AQP4](#) (HGNC Symbol)

CCDS This gene is a member of the Human CCDS set: [CCDS11889.1](#), [CCDS58617.1](#)

UniProtKB This gene has proteins that correspond to the following UniProtKB identifiers: [P55087](#)

RefSeq Overlapping RefSeq Gene ID [361](#) matches and has similar biotype of protein\_coding

Ensembl version ENSG00000171885.13

Other assemblies This gene maps to [24\\_432\\_002-24\\_445\\_782](#) in GRCh37 coordinates. View this locus in the GRCh37 archive: [ENSG00000171885](#)

Gene type Known protein coding

Annotation method Annotation for this gene includes both automatic annotation from Ensembl and [Havana](#) manual curation, see [article](#).

Alternative genes This gene corresponds to the following database identifiers:  
Havana gene: [OTTHUMG00000131955](#)

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

Drag>Select: ↕

Forward strand

Genes (Comprehensive set...)

26.845Mb 26.850Mb 26.855Mb 26.860Mb 26.865Mb 26.870Mb 26.875Mb

33.78 kb

AQP4-AS1-001 > antisense  
AQP4-AS1-002 > antisense  
AQP4-AS1-003 > antisense  
AQP4-AS1-004 > antisense  
AQP4-AS1-005 > antisense  
AQP4-AS1-006 > antisense  
AQP4-AS1-007 > antisense

## Possible transcripts for this gene

## Sequence of the canonical isoform

		W S R R S V *YY YD Y R W R M * K	
Exons	A	301 AACGGGCTGGTGAACAGAAAAGCCTTACCGGTGAGATGGTCTCATCTCCCTTTGCTT 360	
		172 AACTGGGGTGGAACAGAAAAGCCTTACCGGTGAGATGGTCTCATCTCCCTTTGCTT 233	
Markup	lof	58 -N--W--G---T--E--K--P--L--P--V--D--M--V--L--I--S--L--C--F-	11

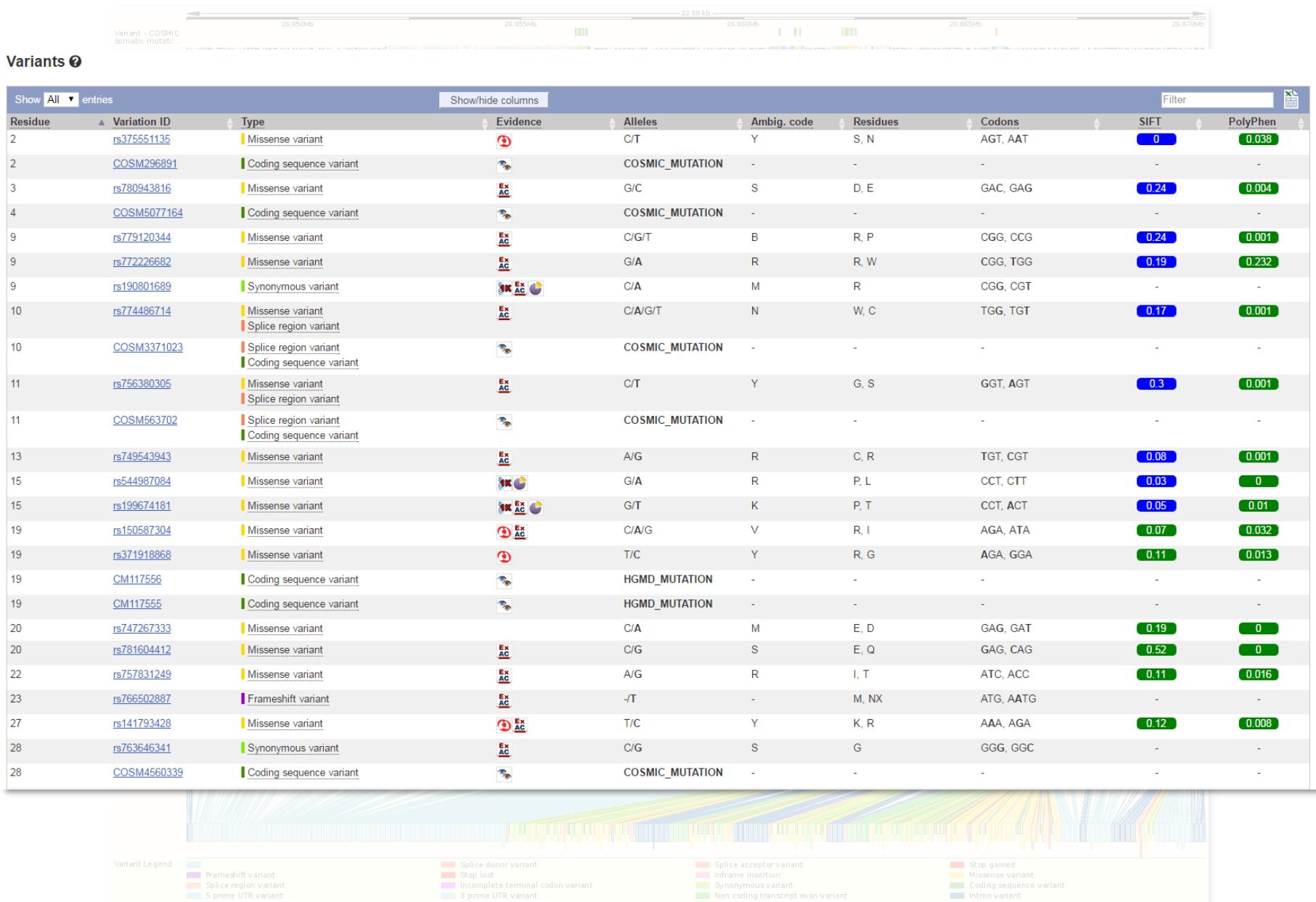
Variation: rs711087375												
	Position	18:26862293										
	Alleles	G/A										
481	TACATCGCAGCCCCAGTGCTGAGGCCATATTGGACGAGGAATCC											
352	TACATCGCAGCCCCAGTGCTGCTGGGGGCCATATTGGAGCAGGAATCC											
118	-Y---A--A---Q---C---L---G---A---I---I---G---A---G---I---											
	cDNA position	465										
	Protein	112										

	M	R*	K	WY	R	Amino acids	I/I
	Codons		Codons		Codons		
541	CCTCC	AGTGT	GCTGGGAGGCCT	GGAGTCACCAT	ATTT	CATGGAA	atC/atT
412	CCTCCC	AGTGTG	TGGGGAGGCCTGGGAGTCACCATGTT	CATGGAA			Consequences
138	-P--P-S-	V-V-G-G-L-L-G-V-T-M-M-V-H-H-G-					Synonymous variant
							<a href="#">Explore this variant</a>
							<a href="#">Gene/Transcript Locations</a>
	*	SWH	*	R		Y	
601	ATGGT	CTC	C	GGTTGAGTT	CATAATCACATT	CAATTGGTGTTA	TACTAT
472	CATGGT	CTC	T	GGTTGAGTT	GAAATCACATT	CAATTGGTGTT	A
158	-H-G-I-L-L-V-F-I-I-T-T-F-O-I-V-F-T-I-F-A-S-						

2 features	
Somatic mutation: COSM5519546	Variation: rs760876531
Position 18:26862292	Position 18:26862292
Alleles COSMIC_MUT...	Alleles C/T
cDNA position 466	cDNA position 466
Protein position 113	Protein position 113
Consequences Coding sequence variant	Amino acids A/T Codons Gcc/Acc
<a href="#">Explore this variant</a>	Consequences Mis sense variant
<a href="#">Gene/Transcript Locations</a>	<a href="#">Explore this variant</a>
<a href="#">Phenotype Data</a>	

Variation: rs771087375	
Position	18.26862293
Alleles	G/A
cDNA position	465
Protein position	112
Amino acids	I/I
Codons	atC/atT
Consequences	Synonymous variant
<a href="#">Explore this variant</a>	
<a href="#">Gene/Transcript Locations</a>	

# Variants



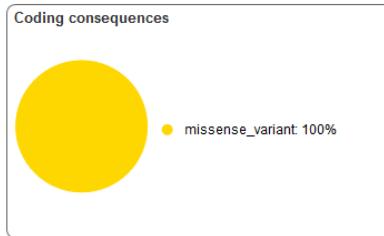
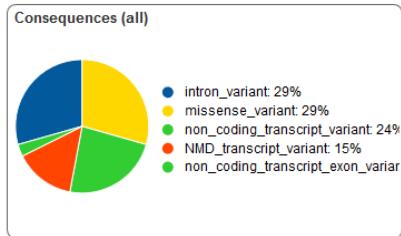
# Variant Effect Predictor

## Variant Effect Predictor results

Job details 

Summary statistics 

Category	Count
Variants processed	1
Variants filtered out	0
Novel / existing variants	0 (0.0) / 1 (100.0)
Overlapped genes	2
Overlapped transcripts	21
Overlapped regulatory features	0



## Results preview

Navigation (per variant)  
Page:   1 of 1   | Show:  All variants

Filters  
Uploaded variant  is  Add

Download  
All:     
BioMart:  

Show/hide columns (25 hidden)																		
Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene	Feature type	Feature	Biotype	Exon	Intron	cDNA position	CDS position	Protein position	Amino acids	Codons	Effect	rs ID
rs193922593	<a href="#">12:120979049-120979049</a>	T	missense_variant	MODERATE	HNF1A	<a href="#">ENSG00000135100</a>	Transcript	<a href="#">ENST00000257555</a>	protein_coding	1/10	-	507	281	94	P/L	CCT/CTT	 	
rs193922593	<a href="#">12:120979049-120979049</a>	T	missense_variant	MODERATE	HNF1A	<a href="#">ENSG00000135100</a>	Transcript	<a href="#">ENST00000400024</a>	protein_coding	1/7	-	482	281	94	P/L	CCT/CTT	 	
rs193922593	<a href="#">12:120979049-120979049</a>	T	non_coding_transcript_exon_variant	MODIFIER	HNF1A	<a href="#">ENSG00000135100</a>	Transcript	<a href="#">ENST00000402929</a>	retained_intron	1/6	-	416	-	-	-	-	-	 
rs193922593	<a href="#">12:120979049-120979049</a>	T	intron_variant, non_coding_transcript_variant	MODIFIER	HNF1A-AS1	<a href="#">ENSG00000241388</a>	Transcript	<a href="#">ENST00000433033</a>	antisense	-	1/3	-	-	-	-	-	-	 
rs193922593	<a href="#">12:120979049-120979049</a>	T	intron_variant, non_coding_transcript_variant	MODIFIER	HNF1A-AS1	<a href="#">ENSG00000241388</a>	Transcript	<a href="#">ENST00000535301</a>	antisense	-	1/1	-	-	-	-	-	-	 
rs193922593	<a href="#">12:120979049-120979049</a>	T	intron_variant, non_coding_transcript_variant	MODIFIER	HNF1A	<a href="#">ENSG00000135100</a>	Transcript	<a href="#">ENST00000536955</a>	processed_transcript	-	1/1	-	-	-	-	-	-	 
rs193922593	<a href="#">12:120979049-120979049</a>	T	intron_variant, non_coding_transcript_variant	MODIFIER	HNF1A	<a href="#">ENSG00000241388</a>	Transcript	<a href="#">ENST00000537361</a>	antisense	-	1/1	-	-	-	-	-	-	 
rs193922593	<a href="#">12:120979049-120979049</a>	T	intron_variant, non_coding_transcript_variant	MODIFIER	HNF1A-AS1	<a href="#">ENSG00000135100</a>	Transcript	<a href="#">ENST00000538626</a>	processed_transcript	-	1/1	-	-	-	-	-	-	 
rs193922593	<a href="#">12:120979049-120979049</a>	T	missense_variant, NMD_transcript_variant	MODERATE	HNF1A	<a href="#">ENSG00000135100</a>	Transcript	<a href="#">ENST00000538646</a>	nonsense-mediated_decay	1/6	-	399	281	94	P/L	CCT/CTT	 	

NCBI Resources How To Sign in to NCBI

ClinVar ClinVar Search ClinVar for gene symbols, HGVS expressions, conditions, and more Advanced

Home About Access Help Submit Statistics FTP

**NEW** Click here to see the new Variation Report design!

**NM\_000545.6(HNF1A):c.281C>T (p.Pro94Leu)**

Variation ID: 36816  
Review status: criteria provided, single submitter

**Interpretation** Go to: ↻ ⓘ

Clinical significance: Likely pathogenic  
Last evaluated: Aug 18, 2011  
Number of submission(s): 1  
Condition(s): Maturity-onset diabetes of the young, type 3 [MedGen - OMIM]  
[See supporting ClinVar records](#) ⓘ

**Allele(s)** Go to: ↻ ⓘ

**NM\_000545.6(HNF1A):c.281C>T (p.Pro94Leu)**

Allele ID: 45477  
Variant type: single nucleotide variant  
Cytogenetic location: 12q24.31  
Genomic location: Chr12: 120979049 (on Assembly GRCh38)  
Chr12: 121416852 (on Assembly GRCh37)  
Protein change: P94L  
HGVS: NG\_011731.2:g.5304C>T  
NM\_000545.6:c.281C>T  
NP\_000536.5:p.Pro94Leu  
[...more](#)

Links: ClinGen: CA214295  
dbSNP: rs193922593  
NCBI 1000 Genomes Browser: rs193922593  
Molecular consequence: NM\_000545.6:c.281C>T: missense variant [Sequence Ontology SO:0001583]

**1 Affected gene**  
HNF1 homeobox A (HNF1A) [Gene - OMIM - Variation Viewer]  
Search ClinVar for variants within HNF1A  
Search ClinVar for variants including HNF1A

**Variant frequency in dbGaP** ⓘ  
No dbGaP data has been submitted for this variant.

**Browser views**  
RefSeqGene  
Variation Viewer [GRCh38 - GRCh37]  
UCSC [GRCh38/hg38 - GRCh37/hg19]

**Related information**  
dbSNP  
Functional Class  
Gene  
MedGen  
OMIM  
Related genes (specific)

# Uniprot – Protein knowledgebase

The screenshot shows the UniProt homepage with a blue header bar containing the UniProt logo, a search bar labeled "UniProtKB", and links for "Advanced", "Search", "Help", and "Contact". Below the header, there's a navigation bar with tabs: "BLAST", "Align", "Retrieve/ID mapping", and "Peptide search". A main banner states: "The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information." To the left, a red dashed box highlights the "UniProtKB" section, which includes "UniProt Knowledgebase", "Swiss-Prot (552,884)" (manually annotated and reviewed), and "TrEMBL (70,656,157)" (automatically annotated and not reviewed). To the right, there are four boxes: "UniRef" (UniProt Reference Clusters), "UniParc" (a comprehensive database of protein sequences), "Proteomes" (proteomes for species with completely sequenced genomes), and "Supporting data" sections for "Literature citations", "Taxonomy", "Diseases", "Subcellular locations", and "Keywords". On the far right, there's a "News" sidebar with links to "Forthcoming changes", "UniProt release 2016\_10", and "Protein spotlight" on "Seeing Red".

TrEMBL: Automatically inferred from the Genome  
Swiss-Prot: Reviewed with evidence from the literature

# Uniprot – Protein knowledgebase

UniProtKB results

Filter by:

- Reviewed (20,114) Swiss-Prot
- Unreviewed (50,788) TrEMBL
- Popular organisms Human (70,902)
- Proteomes UP000005640 (70,902)

View by:

- Results table
- Taxonomy
- Keywords
- Gene Ontology
- Enzyme class
- Pathway
- UniRef

Your results in sequence clusters with identity of:  
100%, 90% or 50%

- Demo
- Help video

UniProtKB organism:"Homo sapiens (Human) [9606]" AND proteome:up000005640

Advanced Search

Help Contact About UniProtKB Basket

1 to 25 of 70,902 Show 25

Entry	Entry name	Protein names	Gene names	Organism	Length
P30450	1A26_HUMAN	HLA class I histocompatibility anti...	HLA-A HLA	Homo sapiens (Human)	365
P10314	1A32_HUMAN	HLA class I histocompatibility anti...	HLA-A HLA	Homo sapiens (Human)	365
P30453	1A34_HUMAN	HLA class I histocompatibility anti...	HLA-A HLA	Homo sapiens (Human)	365
P30456	1A43_HUMAN	HLA class I histocompatibility anti...	HLA-A HLA	Homo sapiens (Human)	365
P30464	1B15_HUMAN	HLA class I histocompatibility anti...	HLA-B HLAB	Homo sapiens (Human)	362
P03989	1B27_HUMAN	HLA class I histocompatibility anti...	HLA-B HLAB	Homo sapiens (Human)	362
P30479	1B41_HUMAN	HLA class I histocompatibility anti...	HLA-B HLAB	Homo sapiens (Human)	362
P30486	1B48_HUMAN	HLA class I histocompatibility anti...	HLA-B HLAB	Homo sapiens (Human)	362
P30487	1B49_HUMAN	HLA class I histocompatibility anti...	HLA-B HLAB	Homo sapiens (Human)	362
P30492	1B54_HUMAN	HLA class I histocompatibility anti...	HLA-B HLAB	Homo sapiens (Human)	362
Q29836	1B67_HUMAN	HLA class I histocompatibility anti...	HLA-B HLAB	Homo sapiens (Human)	362
P30498	1B78_HUMAN	HLA class I histocompatibility anti...	HLA-B HLAB	Homo sapiens (Human)	362
Q29963	1C06_HUMAN	HLA class I histocompatibility anti...	HLA-C HLAC	Homo sapiens (Human)	366
P10321	1C07_HUMAN	HLA class I histocompatibility anti...	HLA-C HLAC	Homo sapiens (Human)	366
Q14738	2A5D_HUMAN	Serine/threonine-protein phosphatase...	PPP2R5D	Homo sapiens (Human)	602
P63151	2ABA_HUMAN	Serine/threonine-protein phosphatase...	PPP2R2A	Homo sapiens (Human)	447
P26439	3BHS2_HUMAN	3 beta-hydroxysteroid dehydrogenase...	HSD3B2 HSDB3B	Homo sapiens (Human)	372

# Non-model organisms

UniProtKB results

Filter by:

- Reviewed (1,419)  
Swiss-Prot
- Unreviewed (24,682)  
TrEMBL

Popular organisms  
**Pig (26,101)**

Proteomes  
**UPI000008227 (26,101)**

View by

Results table

Taxonomy  
Keywords  
Gene Ontology  
Enzyme class  
Pathway

UniRef

Your results in sequence clusters with identity of:  
100%, 90% or 50%

Demo

Help video

UniProtKB results

UniProtKB organism:"Sus scrofa (Pig) [9823]" AND proteome:up000008227

Advanced Search

Entry	Entry name	Protein names	Gene names	Organism	Length
P50129	SHT2A_PIG	5-hydroxytryptamine receptor 2A	HTR2A	Sus scrofa (Pig)	470
Q29006	SHT4R_PIG	5-hydroxytryptamine receptor 4	HTR4	Sus scrofa (Pig)	137
P14332	6PGD_PIG	6-phosphogluconate dehydrogenase, d...	PGD	Sus scrofa (Pig)	250
Q09136	AAPK1_PIG	5'-AMP-activated protein kinase cat...	PRKAA1 AMPK1	Sus scrofa (Pig)	385
Q767L0	ABCF1_PIG	ATP-binding cassette sub-family F m...	ABCF1	Sus scrofa (Pig)	807
P06199	ACM2_PIG	Muscarinic acetylcholine receptor M...	CHRM2	Sus scrofa (Pig)	466
P83662	ACMSD_PIG	2-amino-3-carboxymuconate-6-semiald...	ACMSD	Sus scrofa (Pig)	138
Q29016	ACRBP_PIG	Acrosin-binding protein	ACRBP	Sus scrofa (Pig)	537
P24540	ACYP1_PIG	Acylphosphatase-1	ACYP1 ACYPE	Sus scrofa (Pig)	101
F15PM8	AAK1_PIG	AP2-associated protein kinase 1	AAK1	Sus scrofa (Pig)	968
P79274	ACADL_PIG	Long-chain specific acyl-CoA dehydr...	ACADL	Sus scrofa (Pig)	430
P39092	A1BG_PIG	Alpha-1B-glycoprotein	A1BG	Sus scrofa (Pig)	40
P79307	A4_PIG	Amyloid beta A4 protein	APP	Sus scrofa (Pig)	770
Q28948	AAPK2_PIG	5'-AMP-activated protein kinase cat...	PRKAA2 AMPK2	Sus scrofa (Pig)	552
Q5EE05	ABHD5_PIG	1-acylglycerol-3-phosphate O-acyltr...	ABHD5	Sus scrofa (Pig)	349
P79273	ACADS_PIG	Short-chain specific acyl-CoAdehyd...	ACADS	Sus scrofa (Pig)	413
Q29090	2ABA_PIG	Serine/threonine-protein phosphat...	PPP2R2A	Sus scrofa (Pig)	426

UniProtKB results

Filter by:

- Reviewed (2)  
Swiss-Prot

Popular organisms  
**TYREX (2)**

UniProtKB results

UniProtKB taxonomy:"Tyrannosaurus [436494]"

Advanced Search

Entry	Entry name	Protein names	Gene names	Organism	Length
P0C2W2	CO1A1_TYREX	Collagen alpha-1(I) chain	COL1A1	Tyrannosaurus rex (Tyrant lizard king)	570
P0C2W4	CO1A2_TYREX	Collagen alpha-2(I) chain	COL1A2	Tyrannosaurus rex (Tyrant lizard king)	18

# Example with AQP4 Gene

UniProtKB results

**Filter by:**

- Reviewed (1) Swiss-Prot
- Unreviewed (7) TrEMBL
- Popular organisms** Human (8)
- View by**
- Results table**
- Taxonomy
- Keywords
- Gene Ontology
- Enzyme class
- Pathway

Entry	Entry name	Protein names	Gene names	Organism	Length
P55087	AQP4_HUMAN	Aquaporin-4	AQP4	Homo sapiens (Human)	323
L0R6C7	L0R6C7_HUMAN	Alternative protein AQP4	AQP4	Homo sapiens (Human)	55
F1DSG4	F1DSG4_HUMAN	Aquaporin 4 transcript variant a	AQP4 hCG_37174	Homo sapiens (Human)	323
V9PBN7	V9PBN7_HUMAN	Aquaporin 4 isoform delta4	AQP4	Homo sapiens (Human)	296
Q6L7A0	Q6L7A0_HUMAN	Aquaporin type4 transcript variant ...	AQP4	Homo sapiens (Human)	218
H0Y3H5	H0Y3H5_HUMAN	Aquaporin-4	AQP4	Homo sapiens (Human)	98
Q9H3V7	Q9H3V7_HUMAN	Aquaporin-4	AQP4	Homo sapiens (Human)	11
J3KRM4	J3KRM4_HUMAN	Aquaporin-4	AQP4	Homo sapiens (Human)	185

Entry	Entry name	Protein names	Gene names	Length	Annotation	Protein existence
P55087	AQP4_HUMAN	Aquaporin-4	AQP4	323	Annotation score: <span style="color:blue;">●●●●○</span>	Evidence at protein level
L0R6C7	L0R6C7_HUMAN	Alternative protein AQP4	AQP4	55	Annotation score: <span style="color:blue;">●○○○○</span>	Predicted
F1DSG4	F1DSG4_HUMAN	Aquaporin 4 transcript variant a	AQP4 hCG_37174	323	Annotation score: <span style="color:blue;">●○○○○</span>	Evidence at transcript level
V9PBN7	V9PBN7_HUMAN	Aquaporin 4 isoform delta4	AQP4	296	Annotation score: <span style="color:blue;">●○○○○</span>	Evidence at protein level
Q6L7A0	Q6L7A0_HUMAN	Aquaporin type4 transcript variant ...	AQP4	218	Annotation score: <span style="color:blue;">●○○○○</span>	Evidence at transcript level
H0Y3H5	H0Y3H5_HUMAN	Aquaporin-4	AQP4	98	Annotation score: <span style="color:blue;">●○○○○</span>	Evidence at protein level
Q9H3V7	Q9H3V7_HUMAN	Aquaporin-4	AQP4	11	Annotation score: <span style="color:blue;">●○○○○</span>	Predicted
J3KRM4	J3KRM4_HUMAN	Aquaporin-4	AQP4	185	Annotation score: <span style="color:blue;">●○○○○</span>	Evidence at protein level

# Example with Protein Aquaporin 4

UniProtKB - P55087 (AQP4\_HUMAN)

Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequences (2)

Cross-references

Entry information

Miscellaneous

Similar proteins

Top

UniProtKB Advanced Search

BLAST Align Retrieve/ID mapping Peptide search

Help Contact Basket

Search

Protein | Aquaporin-4

Gene | AQP4

Organism | Homo sapiens (Human)

Status | Reviewed - Annotation score: 10000 - Experimental evidence at protein level<sup>1</sup>

Function<sup>i</sup>

Forms a water-specific channel. Osmoreceptor which regulates body water balance and mediates water flow within the central nervous system.

GO - Molecular function<sup>i</sup>

- glycerol channel activity Source: GO\_Central
- water channel activity Source: Reactome
- water transmembrane transporter activity Source: ProtInc

GO - Biological process<sup>i</sup>

- cellular response to interferon-gamma Source: UniProtKB
- cellular water homeostasis Source: GO\_Central
- ion transmembrane transport Source: GO\_Central
- multicellular organismal water homeostasis Source: UniProtKB
- renal water homeostasis Source: Reactome
- transport Source: ProtInc
- water transport Source: Reactome

Complete GO annotation...

Keywords - Biological process<sup>i</sup>

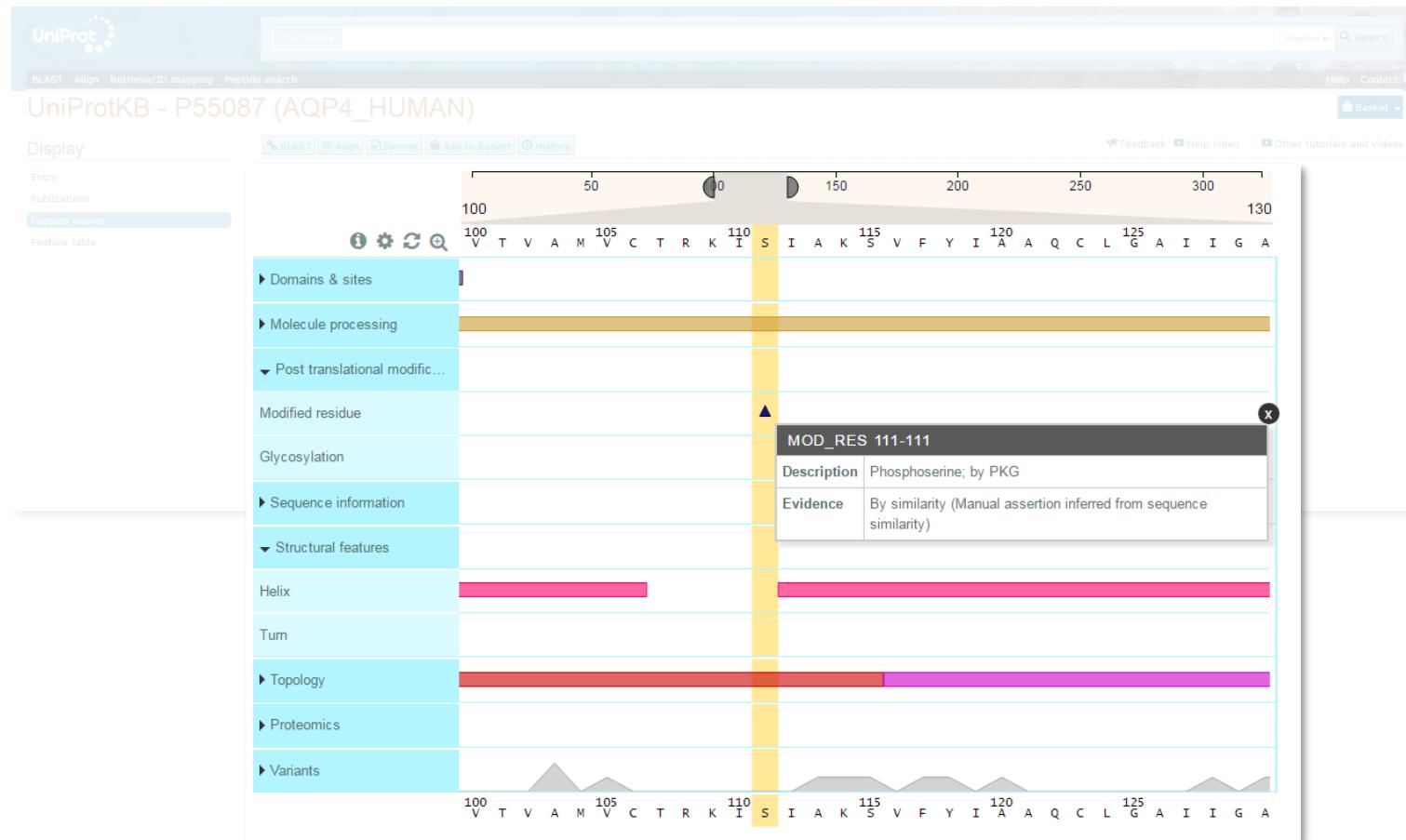
Transport

# Example with Protein Aquaporin 4

The screenshot shows the UniProtKB Publications page for the protein AQP4\_HUMAN (P55087). The top navigation bar includes links for BLAST, Align, Retrieve/ID mapping, Peptide search, UniProtKB dropdown, Advanced search, Search, Help, Contact, and a Basket icon. The main title is "Publications related to P55087 - AQP4\_HUMAN". On the left, there is a sidebar with a "Display" section containing links for Entry, Publications (which is selected), Feature viewer, and Feature table. Below this is a "Filter by:" section with dropdown menus for Source (UniProtKB/Swiss-Prot, reviewed), Computationally mapped (183), Category (PTM / Processing (4), Interaction (2), Structure (1), Sequences (5), Unclassified (179)), and Scale (Large (6), Small (184)). The main content area lists five publications:

- "cDNA cloning, gene organization, and chromosomal localization of a human mercurial insensitive water channel. Evidence for distinct transcriptional units."**  
Yang B., Ma T., Verkman A.S.  
*J. Biol. Chem.* 270:22907-22913(1995) [PubMed] [Europe PMC] [Abstract]  
Cited for: NUCLEOTIDE SEQUENCE [mRNA].  
Category: Sequences.  
Tissue: Fetal brain.  
Source: UniProtKB/Swiss-Prot (reviewed).
- "A water channel closely related to rat brain aquaporin 4 is expressed in acid- and pepsinogen-secretory cells of human stomach."**  
Misaka T., Abe K., Iwabuchi K., Kusakabe Y., Ichinose M., Miki K., Emori Y., Arai S.  
*FEBS Lett.* 381:208-212(1996) [PubMed] [Europe PMC] [Abstract]  
Cited for: NUCLEOTIDE SEQUENCE [mRNA].  
Category: Sequences.  
Source: UniProtKB/Swiss-Prot (reviewed).
- "The human AQP4 gene: definition of the locus encoding two water channel polypeptides in brain."**  
Lu M., Lee M.D., Smith B.L., Jung J.S., Agre P., Verdijk M.A.J., Merkx G., Rijss J.P.L., Deen P.M.T.  
*Proc. Natl. Acad. Sci. U.S.A.* 93:10908-10912(1996) [PubMed] [Europe PMC] [Abstract]  
Cited for: NUCLEOTIDE SEQUENCE [mRNA].  
Category: Sequences.  
Source: UniProtKB/Swiss-Prot (reviewed).  
This publication is cited by 1 other entry.
- "DNA sequence and analysis of human chromosome 18."**  
Nusbaum C., Zody M.C., Borowsky M.L., Kamal M., Kodira C.D., Taylor T.D., Whittaker C.A., Chang J.L., Cuomo C.A., Dewar K., Fitzgerald M.G., Yang X., Abouelleil A., Allen N.R., Anderson S., Bloom T., Bugalter B., Butler J., Lander E.S.  
*Nature* 437:551-555(2005) [PubMed] [Europe PMC] [Abstract]  
Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
Category: Sequences.  
Source: UniProtKB/Swiss-Prot (reviewed).  
This publication is cited by 1123 other entries.
- "The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC)."**  
The MGC Project Team  
*Genome Res.* 14:2121-2127(2004) [PubMed] [Europe PMC] [Abstract]  
Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
Category: Sequences.  
Tissue: Lung.  
Source: UniProtKB/Swiss-Prot (reviewed).  
This publication is cited by 50506 other entries.

# Example with Protein Aquaporin 4



# Example with Protein Aquaporin 4

UniProtKB - P55087 (AQP4\_HUMAN)

Display

Entry Publications Feature viewer Feature table

All None

Function Names & Taxonomy Subcellular location Pathology & Biotech **PTM / Processing** Expression Interaction Structure

Family & Domains Sequences (2) Cross-references Entry information Miscellaneous Similar proteins Top

UniProtKB Advanced Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact Basket

Feedback Help video Other tutorials and videos

Protein | Aquaporin-4  
Gene | AQP4  
Organism | Homo sapiens (Human)  
Status | Reviewed - Annotation score: 88888 - Experimental evidence at protein level<sup>1</sup>

**PTM / Processing<sup>1</sup>**

Molecule processing

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Chain <sup>1</sup>	1 – 323		323 Aquaporin-4		PRO_0000063948	Add BLAST

Amino acid modifications

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Modified residue <sup>1</sup>	111 – 111		1 Phosphoserine; by PKG			
Glycosylation <sup>1</sup>	153 – 153		1 N-linked (GlcNAc...)			
Modified residue <sup>1</sup>	180 – 180		1 Phosphoserine; by PKC			
Glycosylation <sup>1</sup>	206 – 206		1 N-linked (GlcNAc...)			
Modified residue <sup>1</sup>	276 – 276		1 Phosphoserine			
Modified residue <sup>1</sup>	285 – 285		1 Phosphoserine			
Modified residue <sup>1</sup>	289 – 289		1 Phosphothreonine			
Modified residue <sup>1</sup>	321 – 321		1 Phosphoserine			

**Post-translational modification<sup>1</sup>**  
Phosphorylation by PKC at Ser-180 reduces conductance by 50%. Phosphorylation by PKG at Ser-111 in response to glutamats increases conductance by 40% (By similarity).

# Link to PhosphoSitePlus

**PhosphoSitePlus®**  
with grant support from    

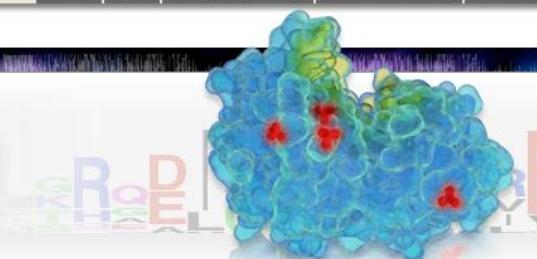
from  **Cell Signaling  
TECHNOLOGY®**

[ABOUT](#) | [USING](#) | [PRIVACY & COOKIES](#) | [CURATION PROCESS](#) | [CONTACT](#)

PhosphoSitePlus® (PSP) is an online systems biology resource providing comprehensive information and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, ubiquitination, acetylation and methylation. See [About PhosphoSite](#) above for more information.

Please cite the following reference for this resource: Hornbeck PV, et al (2015) *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations*. Nucleic Acids Res. 43:D512-20. [reprint]

## A PROTEIN MODIFICATION RESOURCE



**PROTEIN OR SUBSTRATE SEARCH**

Protein Name:  **SEARCH**

**ADVANCED SEARCH AND BROWSE OPTIONS**

-  [Protein, Sequence, or Reference Search](#)
-  [Site Search](#)
-  [Comparative Site Search](#)
-  [Browse MS2 Data By Disease](#)
-  [Browse MS2 Data by Cell Line](#)
-  [Browse MS2 Data by Tissue](#)

**WHAT'S NEW**

August 2016 **CST-LINCS BD2K** tyrosine kinase inhibition TMT data is available for download. Click [here](#) to download the quantitative TMT data and a description of the experimental and analytical process.

June 2016 **Cytoscape plugin** has been updated for Cytoscape 3.0-3.2. We are working or plugin for Cytoscape 3.3-3.x.

Dec 2014 Download **PhosphoSitePlus, 2014: mutations, PTMs and recalibrations**. Nucleic Acids Res.(2015) 43:D512-20.

Aug 2014 **Download PTMVar dataset:** Overlap of disease missense mutations & genetic variants, with their corresponding PTMs and flanking sequences.

**SITE STATISTICS**

	TOTAL	NON-REDUNDANT
Proteins:	53,503	20,252
PTMs, all types:	496,049	396,635
PTMs, low-throughput (LTP) methods:	22,656	17,622
PTMs, high-throughput (HTP) MS/MS:	485,387	388,696
MS peptides:	2,294,489	481,986
Number curated papers:	20,858	

**MODIFICATION SITE STATISTICS, NON-REDUNDANT:**

Acetylation:	36,665	Caspase cleavage:	484
Di-Methylation:	2,840	Methylation:	215
Mono-Methylation:	14,470	O-Galnac:	2,118
O-GlcNAc:	1,475	Phospho-Ser:	160,443
Phospho-Thr:	66,985	Phospho-Tyr:	42,908
Succinylation:	4,628	Sumoylation:	6,955
Tri-Methylation:	334	Ubiquitylation:	56,911

 PhosphoSite, created by Cell Signaling Technology is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Information about permissions beyond the scope of this license are available at <http://www.phosphosite.org/staticContact>.

# Example with Aquaporin 4



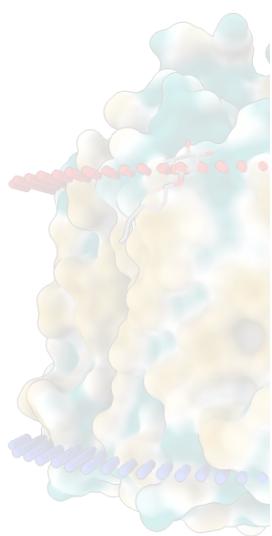
## Sites Implicated In

cell motility, altered: S180-p  
activity, inhibited: S180-p  
intracellular localization: S276-p

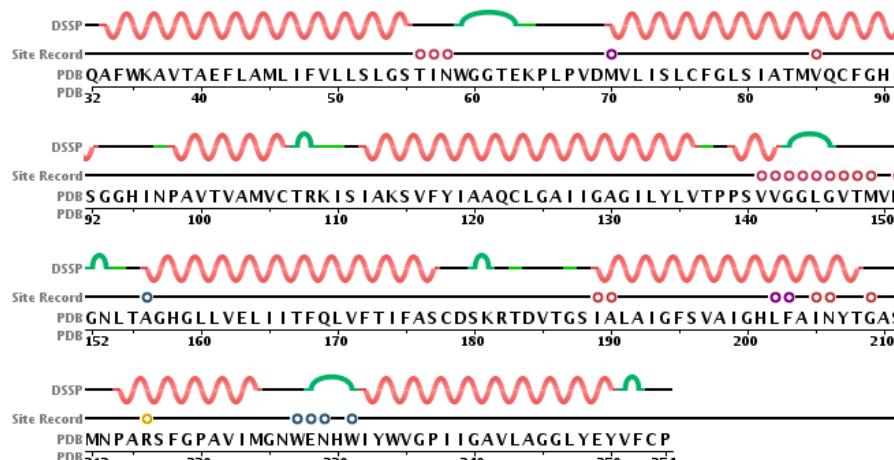
# Link to PDB

The screenshot shows the RCSB PDB homepage. The top navigation bar includes links for Deposit, Search, Visualize, Analyze, Download, Learn, More, and MyPDB Login. Below the header is the RCSB PDB logo and a search bar. The main content area features a map of the world, a search bar, and links for Advanced Search and Browse by Annotations. On the left, a sidebar menu is highlighted with a red dashed border, containing links for Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. To the right of the sidebar, there is a section titled "A Structural View of Biology" which includes a brief description and a link to "Discovering Biology Through Crystallography". Another section on the right is titled "November Molecule of the Month" featuring a 3D molecular model of "Aminopeptidase 1 and Autophagy".

# Example with Aquaporin 4



Sequence Chain View

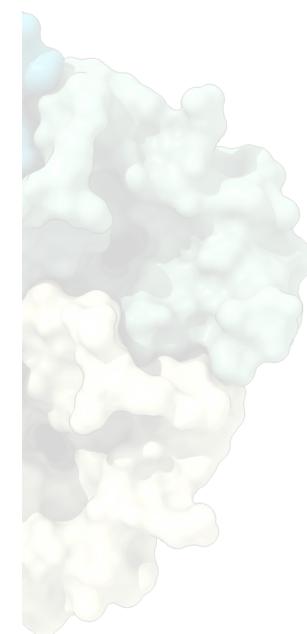


**Site Record Legend**

- BINDING SITE FOR RESIDUE GOL A 6 (SOFTWARE)
- BINDING SITE FOR RESIDUE GOL A 4 (SOFTWARE)
- BINDING SITE FOR RESIDUE GOL A 3 (SOFTWARE)
- BINDING SITE FOR RESIDUE BOG A 473 (SOFTWARE)
- BINDING SITE FOR RESIDUE GOL A 2 (SOFTWARE)
- BINDING SITE FOR RESIDUE GOL A 1 (SOFTWARE)

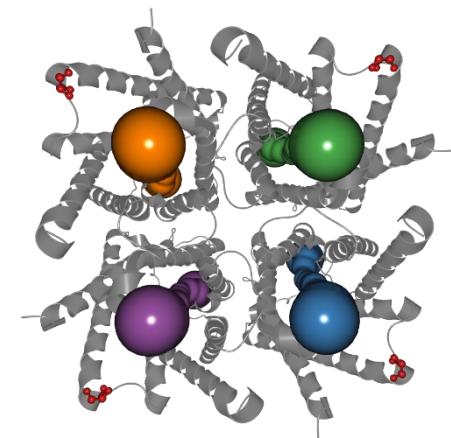
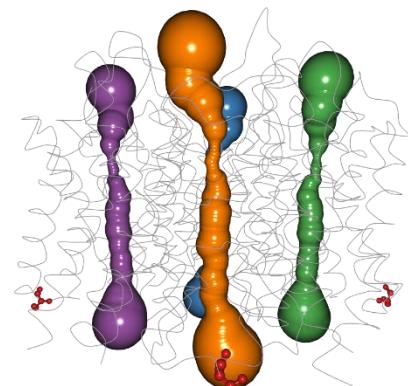
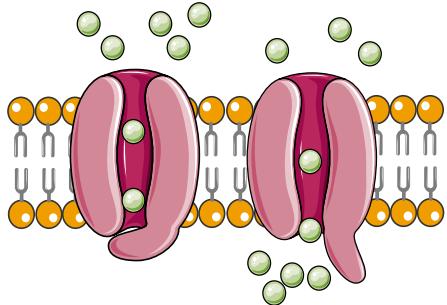
**DSSP Legend**

-  T: turn
-  empty: no secondary structure assigned
-  S: bend
-  H: alpha helix



## Example with Aquaporin 4

---



# Link to STRING

Version: 10.0

LOGIN | REGISTER

 STRING

Search Download Help My Data

Protein by name >

Protein by sequence >

Multiple proteins >

Multiple sequences >

Organisms >

Protein families ("COGs") >

Examples >

Random entry >

## SEARCH

Single Protein by Name / Identifier

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

Organism: auto-detect ▼

SEARCH

© STRING CONSORTIUM 2016

 SIB - Swiss Institute of Bioinformatics

 CPR - NNF Center for Protein Research

 EMBL - European Molecular Biology Laboratory

ABOUT

Content

References

Contributors

Statistics

INFO

Scores

Use scenarios

FAQs

Usage

ACCESS

Versions

APIs

Licensing

Partners

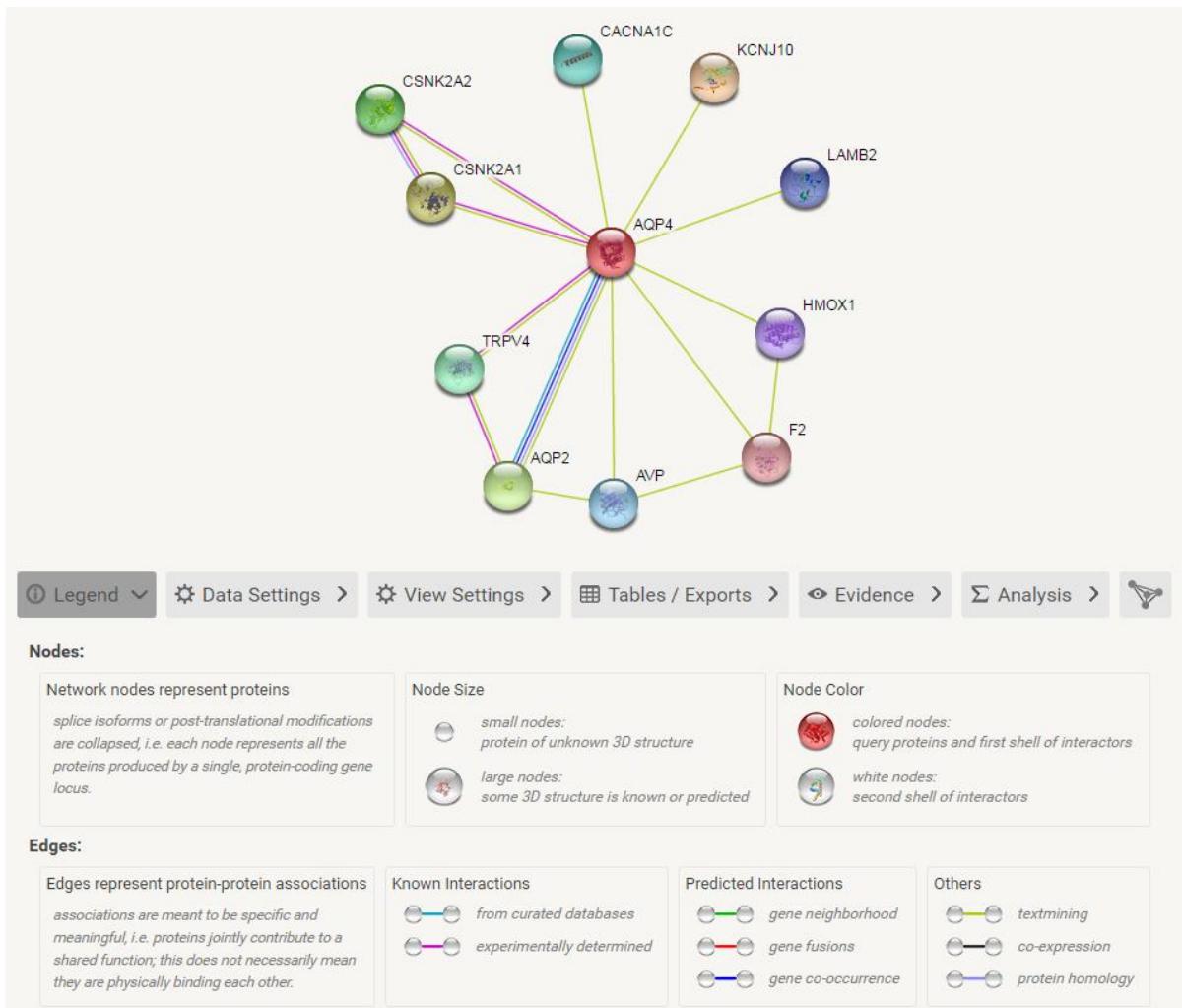
CREDITS

Funding

Datasources

Software

# Example with Aquaporin 4

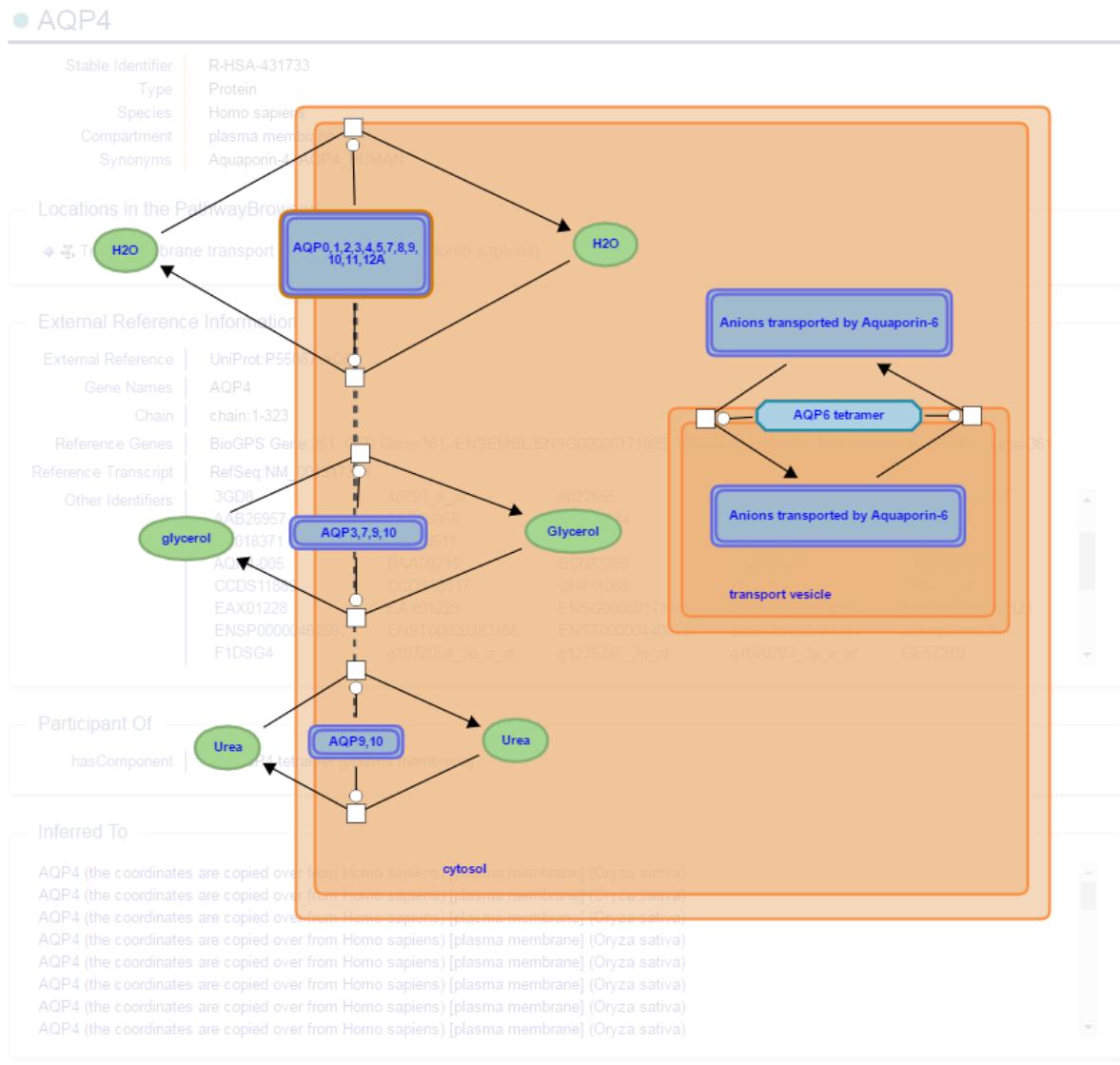


# Link to Reactome



The image shows the Reactome homepage. At the top, there is a banner with the Reactome logo and a blue background featuring a molecular structure. Below the banner is a navigation bar with links: About, Content, Documentation, Tools, Community, Download, Contact, a search bar containing "e.g. O95631, NTN1, signaling", and a "Search" button. The main content area includes several buttons: "Browse Pathways" (with a tree icon), "Analyze Data" (with a bar chart and magnifying glass icon), "Reactome FlViz app" (with a network icon), "User Guide" (with a person icon), "Data Download" (with a download arrow icon), and "Contact Us" (with an envelope icon). To the right, there is a "Tweets" section with the text "Current Version: Reactome V58" and "Tweets by @reactome". At the bottom, there is a "About Reactome" section with a detailed description of the database's purpose and version information, followed by logos for OICR, NYU Langone Medical Center, CSHL, EMBL-EBI, and a green hexagonal logo. A note at the bottom states: "The development of Reactome is supported by grants from the US National Institutes of Health (P41 HG003751 and 1U54GM114833-01), Ontario Research Fund, and the European Molecular Biology Laboratory."

## Example with Aquaporin 4



# Batch queries using Biomarts and APIs

**e|Ensembl** BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors      Login/Register

New  Count  Results       URL  XML  Perl  Help

**Dataset**  
Homo sapiens genes (GRCh38.p7)  
**Filters**  
[None selected]  
**Attributes**  
Ensembl Gene ID  
Ensembl Transcript ID

**Dataset**  
[None Selected]

**GENE:**  
**Ensembl**

Ensembl Gene ID  
 Ensembl Transcript ID  
 Ensembl Protein ID  
 Ensembl Exon ID  
 Description  
 Chromosome Name  
 Gene Start (bp)  
 Gene End (bp)  
 Strand  
 Band  
 Transcript Start (bp)  
 Transcript End (bp)  
 Transcription Start Site (TSS)  
 Transcript length (including UTRs and CDS)  
 Transcript Support Level (TSL)  
 GENCODE basic annotation

APPRIS annotation  
 Associated Gene Name  
 Associated Gene Source  
 Associated Transcript Name  
 Associated Transcript Source  
 Transcript count  
 % GC content  
 Gene type  
 Transcript type  
 Source (gene)  
 Source (transcript)  
 Status (gene)  
 Status (transcript)  
 Version (gene)  
 Version (transcript)

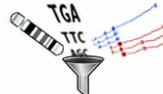
## Download a sequence or region



Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features
- ...and more!

## Customise your download



Custom datasets can be retrieved using the BioMart data-mining tool.  
You may find exploring this web-based query tool easier than extracting information direct from our databases.

## Fetch data programmatically



Write your own Perl scripts to retrieve small-to-medium datasets. All our data, as well as added functionality, is available through the Ensembl Perl API.

Use the API to retrieve gene and transcript sets, fetch alignments between sequences, compare allele frequencies and much more!

You can also use our [REST API](#) to retrieve data to process in the programming language of your choice.

## Download databases & software



All of our data and software, including pipelines and web code, is available free.

- [Download data via FTP](#)
- [Ensembl pipeline on GitHub](#)
- [Set up your own Ensembl website](#)

# Experimental protein characterization

## Human Protein Atlas

# THE HUMAN PROTEIN ATLAS



≡MENU   HELP   NEWS

SEARCH<sup>i</sup>

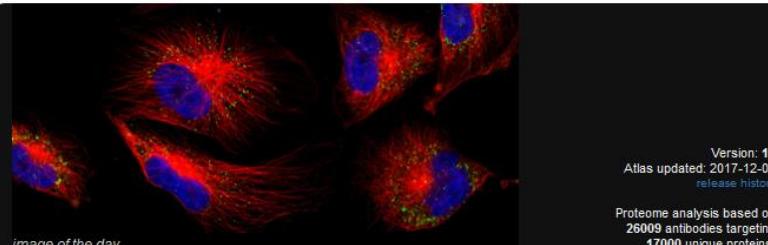
elastase  
e.g. RBM3, insulin, CD36

Search Fields »

 TISSUE ATLAS

 CELL ATLAS

 PATHOLOGY ATLAS



Version: 18  
Atlas updated: 2017-12-01  
[release history](#)

Proteome analysis based on  
26009 antibodies targeting  
17000 unique proteins.

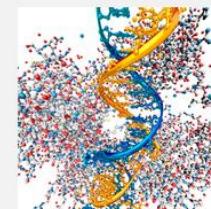
image of the day

...

## ProteomicsDB

### Welcome to ProteomicsDB!

ProteomicsDB is a effort of the Technische Universität München (TUM). It is dedicated to expedite the identification of the human proteome and its use across the scientific community.



**Browse proteins**  
Explore the human proteome protein by protein.



**Browse chromosomes**  
Explore the human proteome chromosome by chromosome.



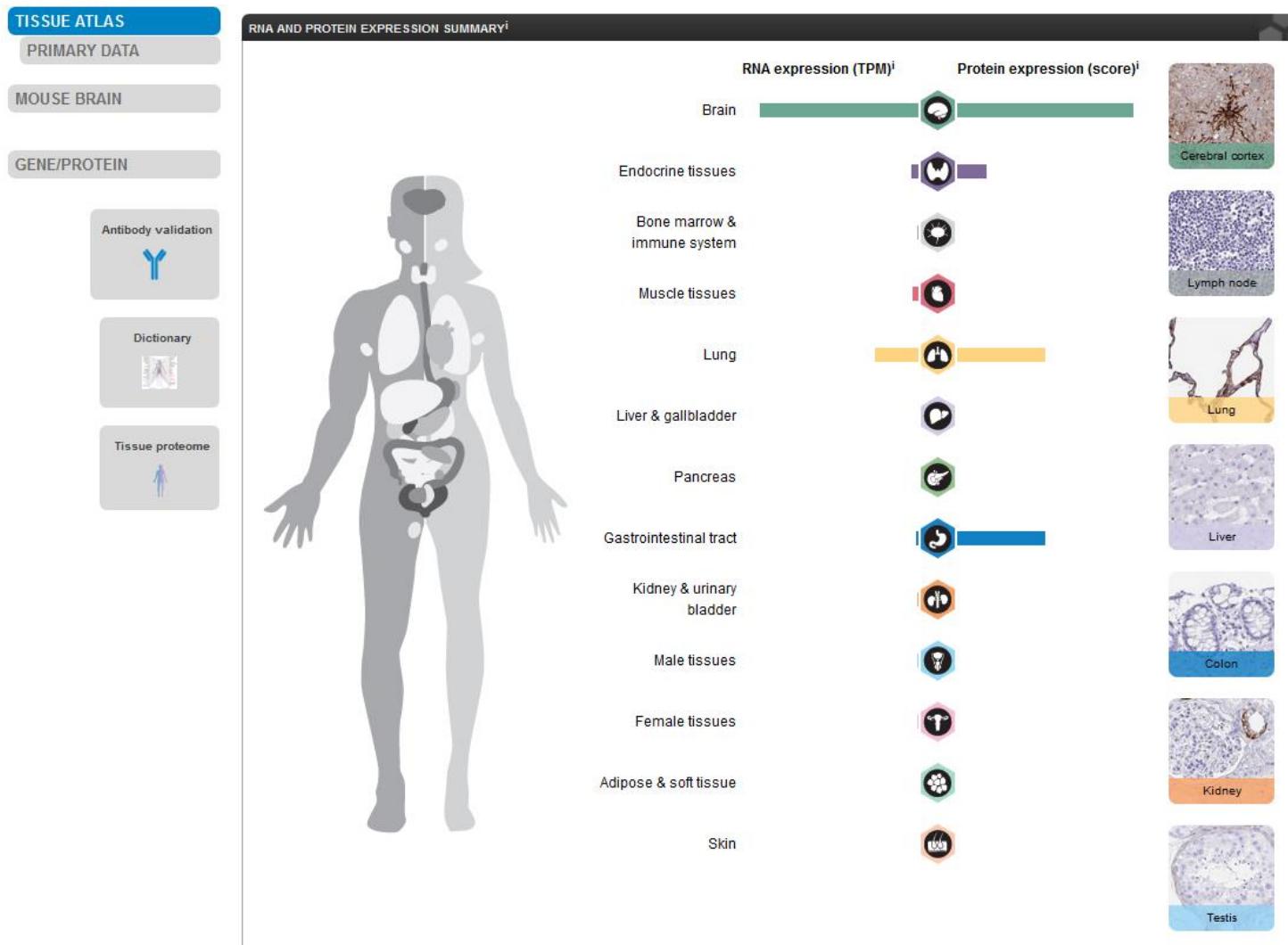
**Adopt a protein**  
Help us to fill the gaps in the human proteome.



**Data repository**  
Share your data with the community.

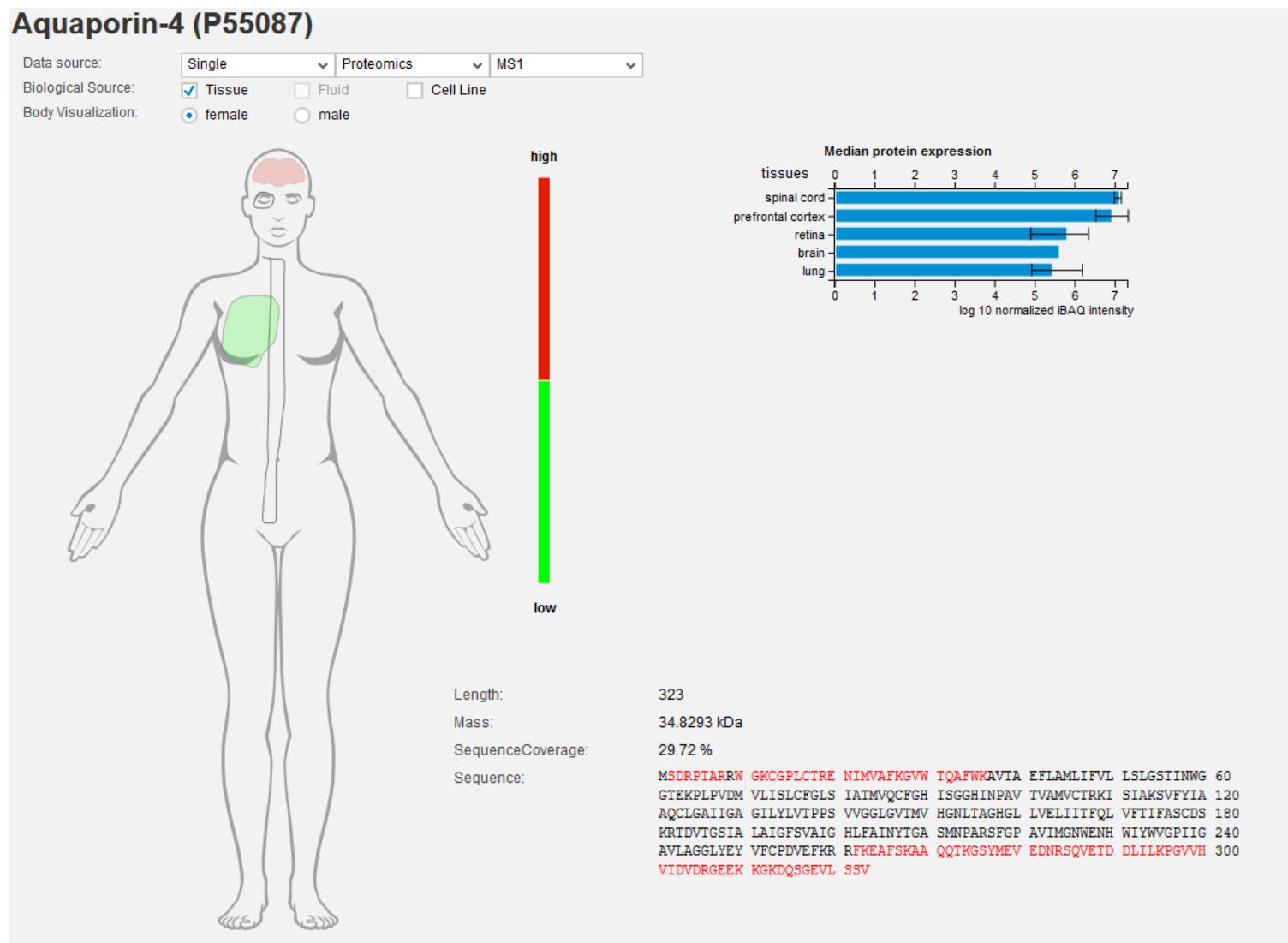
# Example with Aquaporin 4

## Human Protein Atlas



# Example with Aquaporin 4

## ProteomicsDB



# Different Resources for Different Questions

## From Genes to Proteins



Ensembl



NCBI



GenomeNet



Genome Browser

## Metabolites and Drugs



ChEBI - ChEMBL

## Proteins



## Interactions and Pathways

