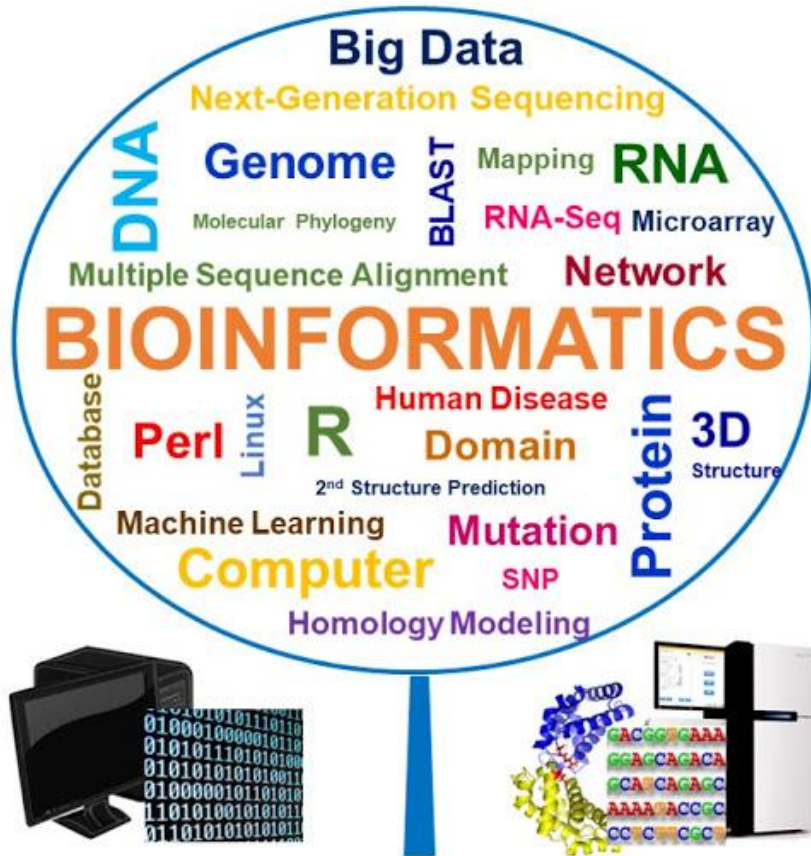
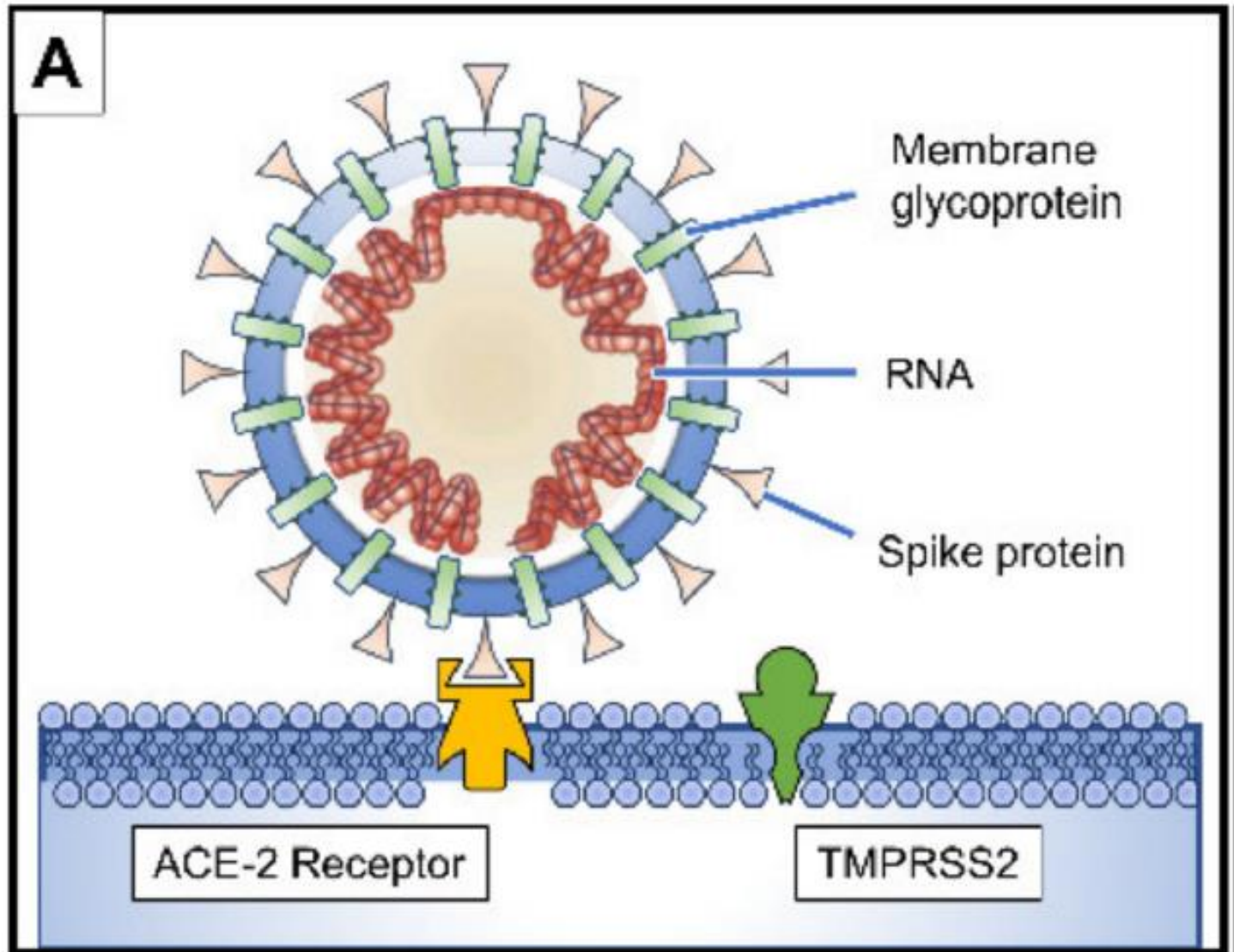


**Welcome
to everybody**

How we can use browsers to find information that we can use in the lab.



How does virus enter in the human cell?





Severe acute respiratory syndrome coronavirus 2 reference genome

Severe acute respiratory syndrome coronavirus 2 (Host: human,vertebrates)

ssRNA(+)



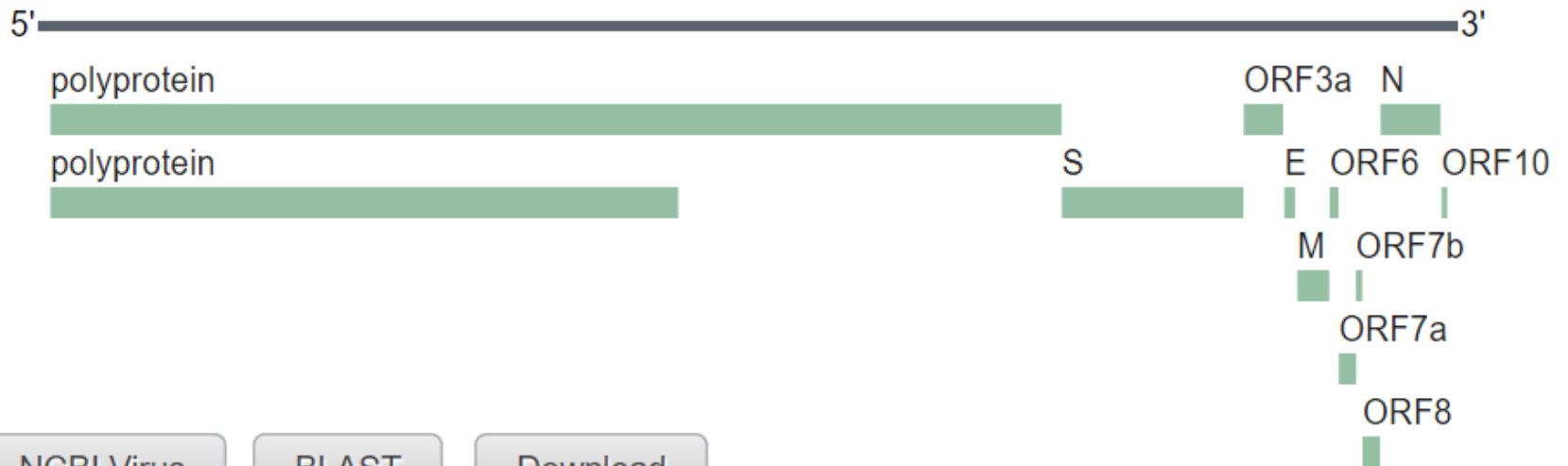
RefSeq GCF_009858895.2

[RefSeq genomic segments](#) (1)

[RefSeq Proteins](#) (28)

[PubMed](#)

[LitCovid](#)



[NCBI Virus](#) [BLAST](#) [Download](#)

Use NCBI Datasets for bulk downloading of genome sequence and annotation data.

[NCBI Datasets](#)

[Command-line tool](#)

[API documentation](#)



Severe acute respiratory syndrome coronavirus 2 reference genome

Severe acute respiratory syndrome coronavirus 2 (Host: human,vertebrates)

ssRNA(+)

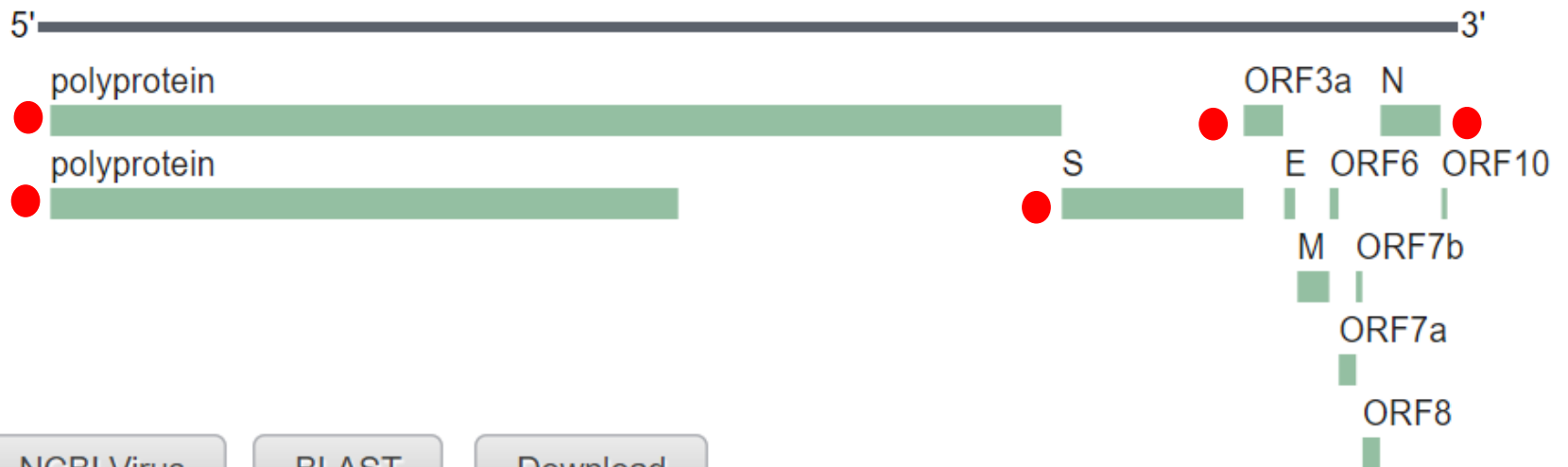
RefSeq GCF_009858895.2

[RefSeq genomic segments](#) (1)

[RefSeq Proteins](#) (28)

[PubMed](#)

[LitCovid](#)



[NCBI Virus](#) [BLAST](#) [Download](#)

Use NCBI Datasets for bulk downloading of genome sequence and annotation data.

[NCBI Datasets](#)

[Command-line tool](#)


[API documentation](#)

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[FASTA](#) [Graphics](#)

Go to:



LOCUS NC_045512 29903 bp ss-RNA linear VRL 13-MAR-2020
DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.
ACCESSION NC_045512
VERSION NC_045512.2
DBLINK BioProject: [PRJNA485481](#)
KEYWORDS RefSeq.
SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
ORGANISM [Severe acute respiratory syndrome coronavirus 2](#)
Viruses; Riboviria; Nidovirales; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.
REFERENCE 1 (bases 1 to 29903)
AUTHORS Wu,F., Zhao,S., Yu,B., Chen,Y.-M., Wang,W., Hu,Y., Song,Z.-G., Tao,Z.-W., Tian,J.-H., Pei,Y.-Y., Yuan,M.L., Zhang,Y.-L., Dai,F.-H., Liu,Y., Wang,Q.-M., Zheng,J.-J., Xu,L., Holmes,E.C. and Zhang,Y.-Z.
TITLE A novel coronavirus associated with a respiratory disease in Wuhan of Hubei province, China
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 29903)
CONSRTM NCBI Genome Project
TITLE Direct Submission
JOURNAL Submitted (17-JAN-2020) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

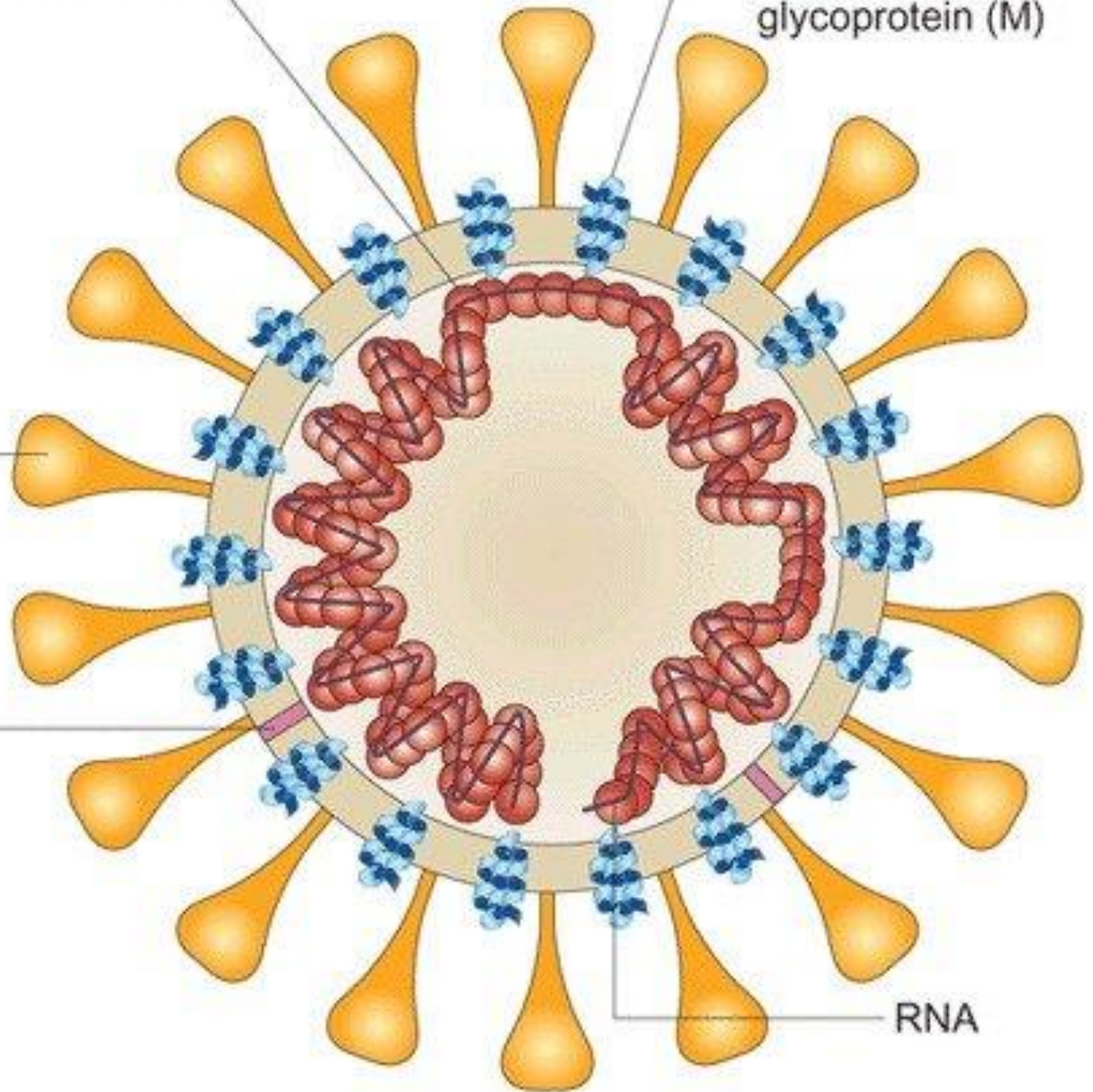
Nucleocapsid protein (N)

Membrane glycoprotein (M)

Spike protein (S)

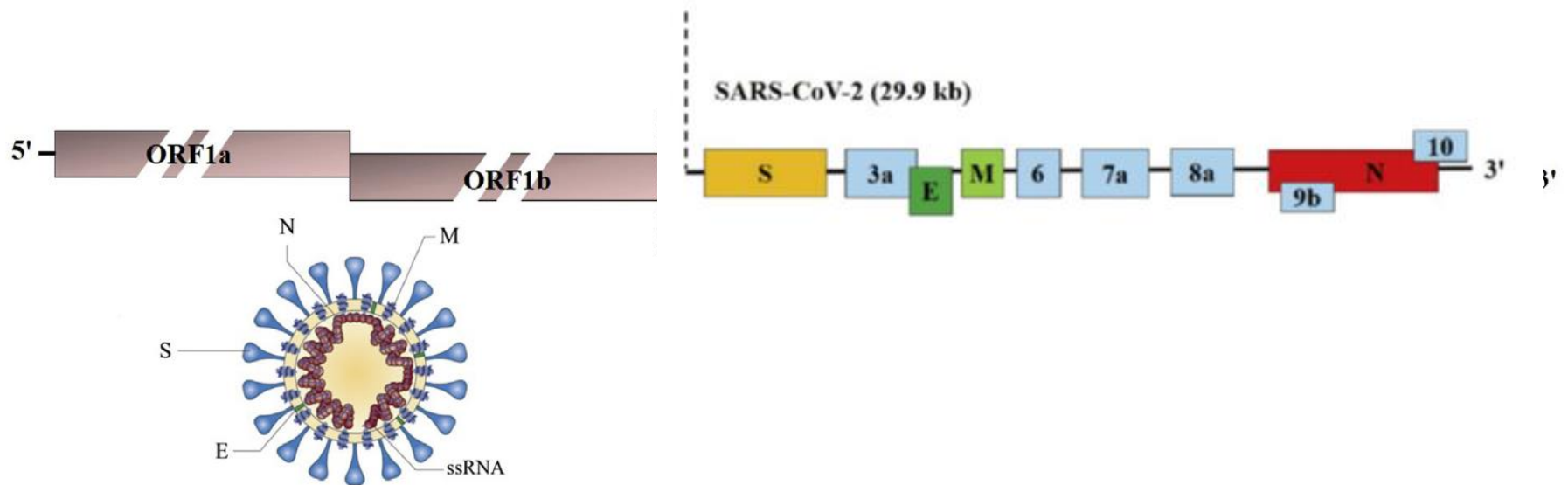
Envelope protein (E)

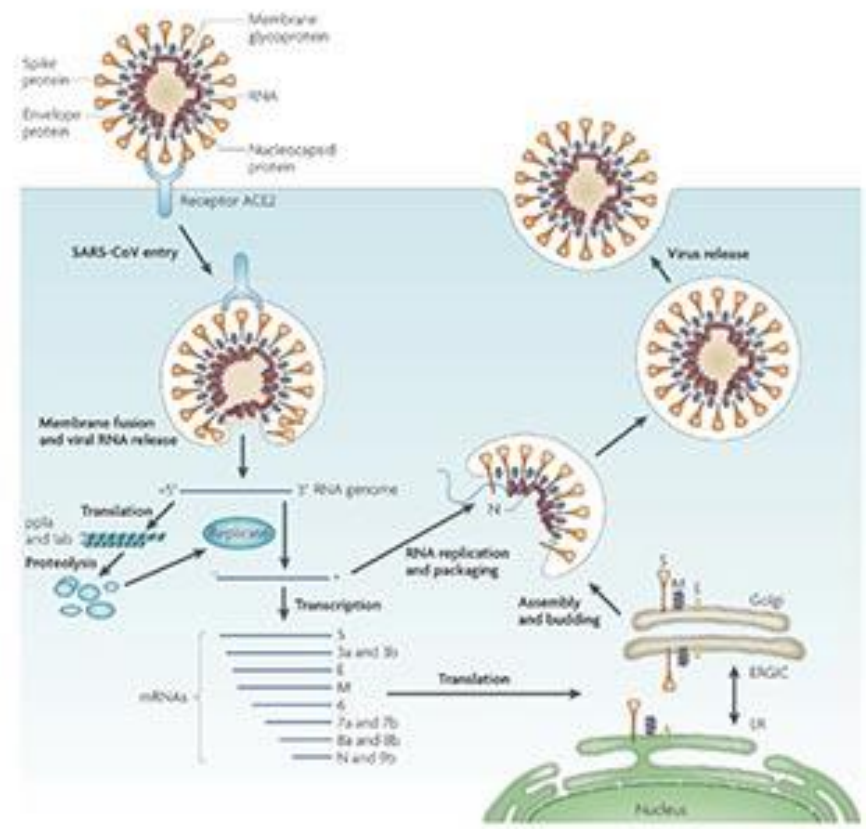
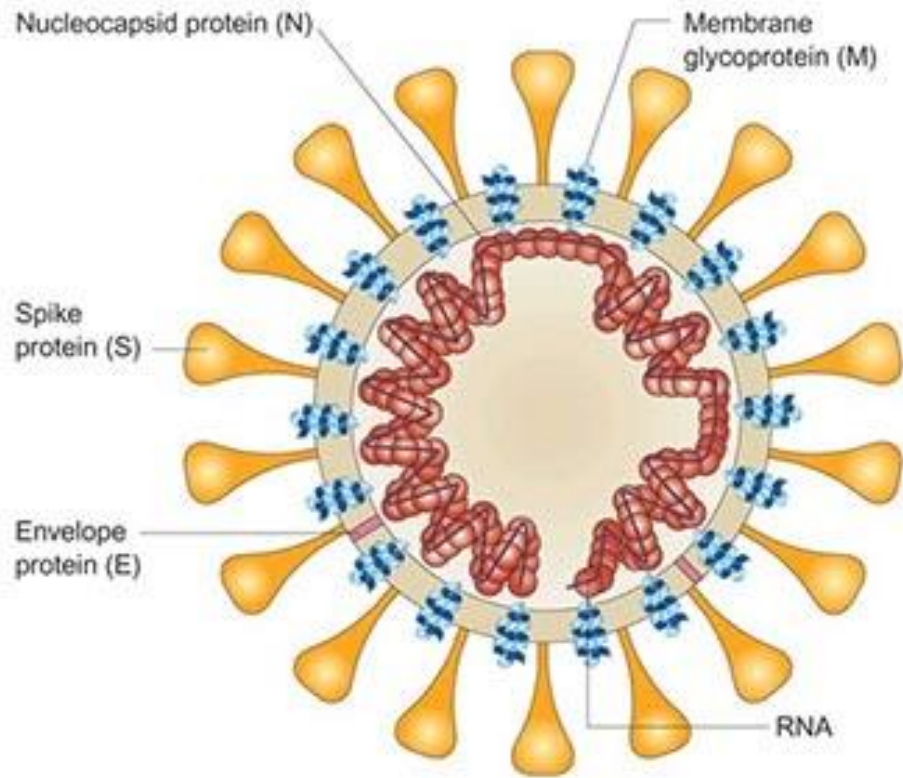
RNA

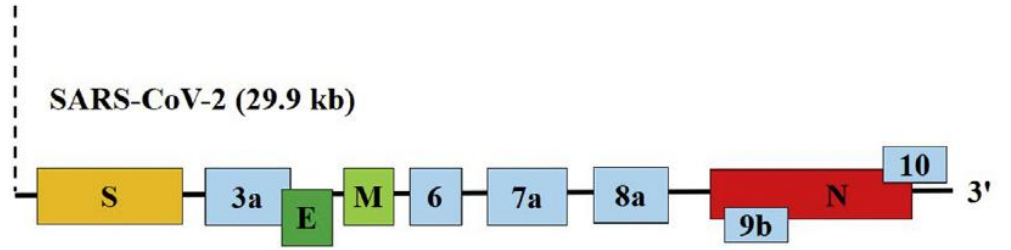


- They are enveloped viruses with round and sometimes pleiomorphic virions of approximately 80 to 120 nm in diameter.
- It contains positive-strand RNA, with the largest RNA genome (approximately 30 kb).
- The genome RNA is complexed with the basic nucleocapsid (N) protein to form a helical capsid found within the viral membrane.
- These are spike (S), the type I glycoprotein that forms the peplomers on the virion surface, giving the virus its corona- or crown-like morphology in the electron microscope.
- The membrane (M) protein, a protein that spans the membrane three times and has a short N-terminal ectodomain and a cytoplasmic tail; and small envelope membrane protein (E), a highly hydrophobic protein.

The 5'-terminal two-thirds of the genome **ORF1a/b** encodes polyproteins, which form the viral replicase transcriptase complex. The other ORFs on the one-third of the genome encode four main structural proteins: **spike (S), envelope (E), nucleocapsid (N) and membrane (M) proteins**, as well as several accessory proteins.







gene

21563..25384
/gene="S"
/locus_tag="GU280_gp02"
/gene_synonym="spike glycoprotein"
/db_xref="GeneID:[43740568](#)"

CDS

21563..25384
/gene="S"
/locus_tag="GU280_gp02"
/gene_synonym="spike glycoprotein"
/note="structural protein; spike protein"
/codon_start=1
/product="surface glycoprotein"
/protein_id="[YP_009724390.1](#)"
/db_xref="GeneID:[43740568](#)"
/translation="MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFR
SSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTKSNIIIR
GWIFGTTLDSKTQSLIIVNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPO
GFSALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGWTAGAAAYYVGYLQPRTFL
LKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN
LCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCF

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[GenBank](#) [Graphics](#)

>NC_045512.2:21563-25384 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

```
ATGTTTGT TTTTCTTG TTTTATTGCC ACTAGTCTCTAGT CAGTGTGTTAATCTTACA ACCAGA AACTCAAT
TACCCCTGC ATACACTAATTCTTT CACACGTGGT GTTTATTACCCTGACAAA GTTTTTCAGATCCTCAGT
TTTACATTCA AACTCAGGACTTGTTC TACCTTTCTTTTCCA ATGTTACTTGGTTC CATGCTATACATGTC
TCTGGGACCA ATGGTACTAAGAGGTTTG ATAACCCTGTCCTACC ATTTAATGATGGT GTTATTTTGCTT
CCTACTGAGA AGTCTAACATAATAAG AGGCTGGATTTTTGGT ACTACTTTAGATT CGAAGACCCAGTCCCT
ACTTATTGTTA ATAACGCTACTAATG TTGTTATTAAGTCTGTG AATTTCAATTTTGTAA TGATCCATTT
TTGGGTGTTT ATTACCACAAA AACAACAAAAGTTGGATGG AAAGTGAGTTCAGAGT TTTATTCTAGTGCGA
ATAATTGCAC TTTTGAATATGTCTCTCAGCCTTTTCTTATGGACCTTGAAGGAAA ACAGGGTAATTTCAA
AAATCTTAGG GAATTTGTGTTAAGAATATTGATGGTTATTTT AAAATATATTCTAAGCACACGCCTATT
AATTTAGTGC GTGATCTCCCTCAGG GTTTTTCGGCTTTAGAAC CATTGGTAGATTTGCCAATAGGTATTA
ACATCACTAG GTTTCAAAC TTTACTTGCCTTTACATAGAA GTTATTTGACTCCTGGT GATTCTTCTTCAGG
TTGGACAGCT GGTGCTGCAGCTT ATTTATGTGGGTTATCTTCAACCTAGGACTTTTCTATTA AAAATATAAT
GAAAAATGGA ACCATTACAGATGCTGTAGACTGTGCACCTTGACCCTCTCTCAGAAA CAAAGTGACGTTGA
AATCCTTCACTGTAGAAAAAGGAATCTATCAA ACTTCTAAC TTTAGAGTCCAACCAACAGAATCTATTGT
TAGATTTCTA ATATTACAACTTGTGCCCTTTTGGTGAAGTTTTT AACGCCACCAGATTTGCATCTGTT
TATGCTTGGAA CAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATCCGCATCAT
TTTCCACTTTT AAGTGTATGGAGTGTCTCCTACTAAATTAATGATCTCTGCTTTACTAATGTCTATGC
AGATTCATTTGT AATTAGAGG TGATGAAGTCAGACAAATCGCTCCAGGGCAA ACTGGAAAGATTGCTGAT
TATAATTATAA ATTACAGATGATTTTACAGGCTGCGTTATAGCTTGG AATTCTAACAATCTTGATTCTA
AGGTTGGTGGT AATTATAAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGA
```

5' - ATGTTTGT TTTTCTTG TTTTATTGCC ACTAGTCTCTAGT CAGTGTGTTAATCTTACA ACCAGA AACTCAAT-3'

5' - **ATGTTTGT TTTTCTTG TTTT** - 3' 

3' - TACAAACAAA AAGAACA AAAATAACGGTGATCAGAGATCAGTCACACAATTAGAATGTTGGTCTTGAGTTA-5'

 3' **ACTGAGGATCTGAAA ACTTTGTCA** - 5'

5' - TACCCCTGC ATACACTAATTCTTT CACACGTGGT GTTTATTACCCTGACAAA GTTTTTCAGATCCTCAGT-3'



Contents lists available at [ScienceDirect](#)

Journal of Pharmaceutical Analysis

journal homepage: www.elsevier.com/locate/jpa

Molecular immune pathogenesis and diagnosis of COVID-19

Xiaowei Li ^{a, b}, Manman Geng ^{a, b}, Yizhao Peng ^{a, b}, Liesu Meng ^{a, b, *}, Shemin Lu ^{a, b, *}

^a Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Xi'an Jiaotong University Health Science Center, Xi'an, Shaanxi, 710061, China

^b Key Laboratory of Environment and Genes Related to Diseases, Ministry of Education, Xi'an, Shaanxi, 710061, China

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[FASTA](#) [Graphics](#)

Go to:

LOCUS NC_045512 29903 bp ss-RNA linear VRL 30-MAR-2020
DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.
ACCESSION NC_045512
VERSION NC_045512.2
DBLINK BioProject: [PRJNA485481](#)
KEYWORDS RefSeq.
SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
ORGANISM [Severe acute respiratory syndrome coronavirus 2](#)
Viruses; Riboviria; Nidovirales; Coronidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.
REFERENCE 1 (bases 13476 to 13503)
AUTHORS Baranov,P.V., Henderson,C.M., Anderson,C.B., Gesteland,R.F., Atkins,J.F. and Howard,M.T.
TITLE Programmed ribosomal frameshifting in decoding the SARS-CoV genome

gene

```
21563..25384  
/gene="S"  
/locus_tag="GU280_gp02"  
/gene_synonym="spike glycoprotein"  
/db_xref="GeneID:43740568"
```

CDS

```
21563..25384  
/gene="S"  
/locus_tag="GU280_gp02"  
/gene_synonym="spike glycoprotein"  
/note="structural protein; spike protein"  
/codon_start=1  
/product="surface glycoprotein"  
/protein_id="YP_009724390.1"  
/db_xref="GeneID:43740568"  
/translation="MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFR  
SSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIR  
GWIFGTTLDSKTQSLLIVNNATNVVIKVFCEQFCNDPFLGVYYHKNNKSWMESEFRVY  
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPO  
GFSALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGWTAGAAAYVGYLQPRFTL  
LKYNENGTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITN  
LCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCF  
TNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYN  
YLYRLFRRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY
```


Severe acute respiratory syndrome-related coronavirus isolate SARS-CoV-2/KMS1/human/2020/CHN, complete genomeSequence ID: [MT226610.1](#) Length: 29899 Number of Matches: 1Range 1: 26295 to 26318 [GenBank](#) [Graphics](#)▾ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
48.1 bits(24)	0.004	24/24(100%)	0/24(0%)	Plus/Plus
Query 1	ACTTCTTTTTCTTGCTTTCGTGGT	24		
Sbjct 26295	ACTTCTTTTTCTTGCTTTCGTGGT	26318		

[CDS](#)

```
25393..26220
/gene="ORF3a"
/locus_tag="GU280_gp03"
/codon_start=1
/product="ORF3a protein"
/protein_id="YP_009724391.1"
/db_xref="GeneID:43740569"
/translation="MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFG
WLIIVGVALLAVFQSASKIITLKKRWQLALSKGVHFCNLLLLFVTVYSHLLLVAAGLE
APFLYLYALVYFLQSINFVRIIMRLWLCWKRKNPLLYDANYFLCWHTNCYDYCIPY
NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQ
LSTDTGVEHVTFIYNKIVDEPEEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL"
```

[gene](#)

```
26245..26472
/gene="E"
/locus_tag="GU280_gp04"
/db_xref="GeneID:43740570"
```

[CDS](#)

```
26245..26472
/gene="E"
/locus_tag="GU280_gp04"
/note="ORF4; structural protein; E protein"
/codon_start=1
/product="envelope protein"
/protein_id="YP_009724392.1"
/db_xref="GeneID:43740570"
/translation="MYSFVSEETGTLIVNSVLLFLAFVWFLLVTLAILTALRLCAYCC
NIVNVSLVKPSFYVYSRVKNLNSSRPDLLV"
```

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/KMS1/human/2020/CHN, complete genome

GenBank: MT226610.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS MT226610 29899 bp RNA linear VRL 27-MAR-2020
DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate
SARS-CoV-2/KMS1/human/2020/CHN, complete genome.
ACCESSION MT226610
VERSION MT226610.1
KEYWORDS .
SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
ORGANISM [Severe acute respiratory syndrome coronavirus 2](#)
Viruses; Riboviria; Nidovirales; Coronaviridae;
Orthocoronavirinae; Betacoronavirus; Sarbecovirus.
REFERENCE 1 (bases 1 to 29899)
AUTHORS Xu,X., Liao,Y., Wang,L., Zhou,X., Xie,Z., Chen,H., Fan,S., Liu,L.,
Zheng,H., Jiang,G. and Li,Q.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2020) Institute of Medical Biology, Chinese
Academy of Medical Sciences and Peking Union Medical College, No.
935, Jiao Ling Road, Kunming, Yunnan 650118, China
COMMENT ##Assembly-Data-START##
Sequencing Technology :: Sanger dideoxy sequencing
##Assembly-Data-END##

NCBI-Q8

← → ↻ covid-19.uniprot.org/uniprotkb?query=*&facets=reviewed:true



* Search

^ COVID-19 UniProtKB 37 results

This site provides the latest available pre-release UniProtKB data for the SARS-CoV-2 coronavirus and other entries relating to the COVID-19 outbreak. Therefore, data and functionality provided here may differ from the main Uniprot.org website which is updated every eight weeks. This site will be updated as new relevant information becomes available, independent of the general UniProt release schedule.

This data can also be accessed via our FTP on
ftp://ftp.uniprot.org/pub/databases/uniprot/pre_release/

Download

Status

Reviewed (Swiss-Prot) (37)

Popular organisms

Human (9)

Other organisms

You can view our latest webinar about Sars-CoV-2 and this portal from [here](#), presented on the 2nd April 2020.

Please go to the [UniProt.org](#) website for all other entries and functionalities.

You can view all COVID-19 related publications contributed by the community [here](#).

Latest update: 28-April-2020

P0DTC1 · R1A_SARS2

Replicase polyprotein 1a · Severe acute respiratory syndrome coronavirus 2 · 4405 amino-acids · 5/5

[#Endonuclease](#) [#Hydrolase](#) [#Nuclease](#) [#Protease](#) [#RNA-binding](#) [#Thiol protease](#) · [#Activation of host autophagy by virus](#) [#Decay of host mRNAs by virus](#) [#Eukaryotic host gene expression shutoff by virus](#) [#Eukaryotic host translation shutoff by virus](#) [#Host gene expression shutoff by virus](#) [#Host mRNA suppression by virus](#) [#Host-virus interaction](#) [#Inhibition of host innate immune response by virus](#) [#Inhibition of host interferon signaling pathway by virus](#) [#Inhibition of host IRF3 by virus](#) [#Inhibition of host ISG15 by virus](#) [#Inhibition of host RLR pathway by virus](#) [#Modulation of host ubiquitin pathway by viral deubiquitinase](#) [#Modulation of host ubiquitin pathway by virus](#) [#Hbl conjugation pathway](#) [#Viral immunoevasion](#)

Report bug



BLAST [Download](#) [Add a Publication](#)

Entry

Function

Names & Taxonomy

Subcellular Location

Pathology & BioTech

Protein Processing

Expression

Interaction

Structure

Family & Domains

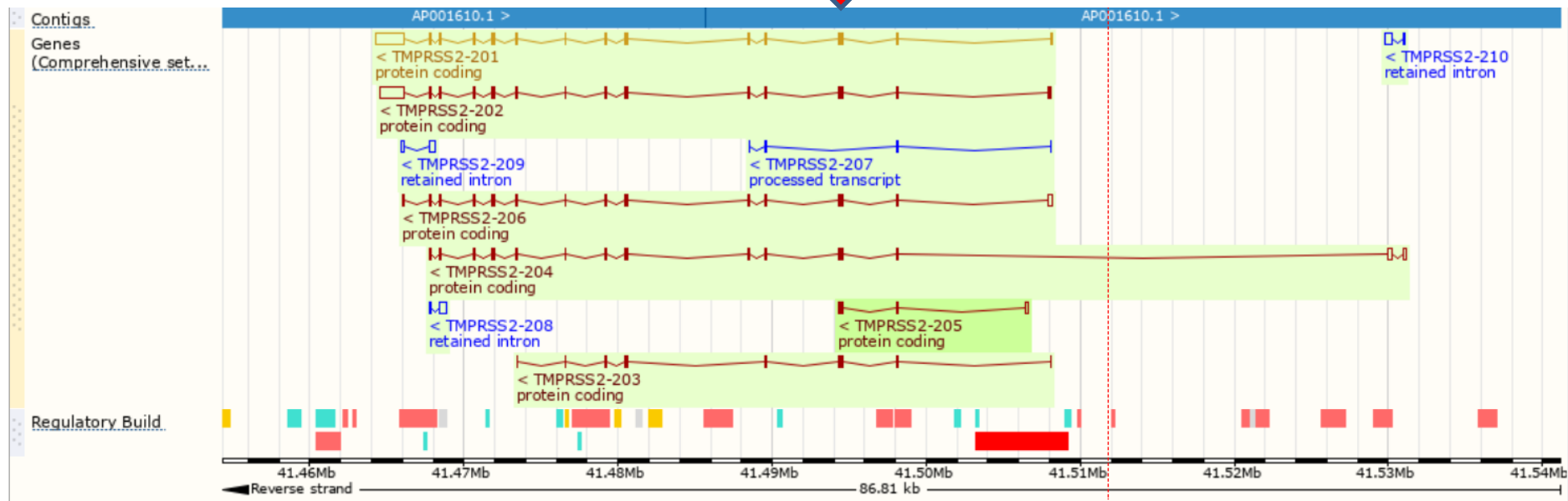


Type	Positions	Description
region of interest	319-541	Receptor-binding domain (RBD) 1 automatic annotation 1 publication
region of interest	437-508	Receptor-binding motif; binding to human ACE2 By similarity
Other site of interest	685-686	Cleavage 1 automatic annotation
region of interest	788-806	Fusion peptide 1 automatic annotation

Ensembl- Second Part-Q2

Show **All** entries Show/hide columns Filter

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence					ctgggacaccgcctcctgagattaaagcgagagccaggggcgggcgggc
1	ENSE00001296629	41,508,158	41,508,081	-	-	78	GAGTAGGCGCGAGCTAAGCAGGAGGCCGAGGGCGGAGGGCGAGGGCGAGGGCGGGGAGCG CCGCCTGGAGCGCGGCAG
	Intron 1-2	41,508,080	41,498,190			9,891	gtgagcgccgcccgtaccagggtcc.....acagtccctgtaattctttcttctag
2	ENSE00003502036	41,498,189	41,498,119	-	0	71	GTCATATTGAACATTCCAGATACCTATCATTACTCGATGCTGTTGATAACAGCAAGATGG CTTTGAACTCA
	Intron 2-3	41,498,118	41,494,579			3,540	gtaagtgggttaattattaccttct.....attcatcttcttttcaactgtttag
3	ENSE00003788834	41,494,578	41,494,356	0	1	223	GGGTCACCACCAGCTATTTGGACCTTACTATGAAAACCATGGATACCAACCGGAAAACCCC TATCCCGCACAGCCCACTGTGGTCCCCACTGTCTACGAGGTGCATCCGGCTCAGTACTAC CCGTCCCGCGTGCCCCAGTACGCCCGAGGGTCTGACGCAGGCTTCCAACCCGTCGTC TGACGCAGCCCAAATCCCCATCCGGGACAGTGTGCACCTCAA
	Intron 3-4	41,494,355	41,489,594			4,762	gtaggattctctttgtgtttatttc.....cgttgcactctcctcttcttccag
4	ENSE00003500399	41,489,593	41,489,507	1	1	87	AGACTAAGAAAGCACTGTGCATCACCTTGACCCTGGGGACCTTCTCGTGGGAGCTGCGC TGGCCGCTGGCTACTCTGGAAGTTCA
	Intron 4-5	41,489,506	41,488,514			993	gtaagtgcaggagcctcgatccca.....acctctcttctgttttctctgcaag
5	ENSE00001308618	41,488,513	41,488,394	1	1	120	TGGGCAGCAAGTGTCCAACTCTGGGATAGAGTGCAGCTCCTCAGGTACCTGCATCAACC CCTCTAAGTGGTGTGATGGCGTGTACACTGCCCGCGGGGAGGACGAGAATCGGTGTG



Ensembl- Second Part-Q9

Human (GRCh38.p13) ▾

Location: 21:41,464,305-41,531,116

Gene: **TMPRSS2**

- Gene-based displays**
- [-] Summary
 - [-] Splice variants
 - [-] Transcript comparison
 - [-] Gene alleles
 - [-] Sequence
 - [-] Secondary Structure
 - [-] Comparative Genomics
 - [-] Genomic alignments
 - [-] Gene tree
 - [-] Gene gain/loss tree
 - [-] Orthologues
 - [-] Paralogues

Gene: **TMPRSS2** ENSG00000184012

Description transmembrane serine protease 2 [Source:HGNC Symbol;Acc:HGNC:11876]

Gene Synonyms PRSS10

Location [Chromosome 21: 41,464,305-41,531,116](#) reverse strand.
GRCh38:CM000683.2

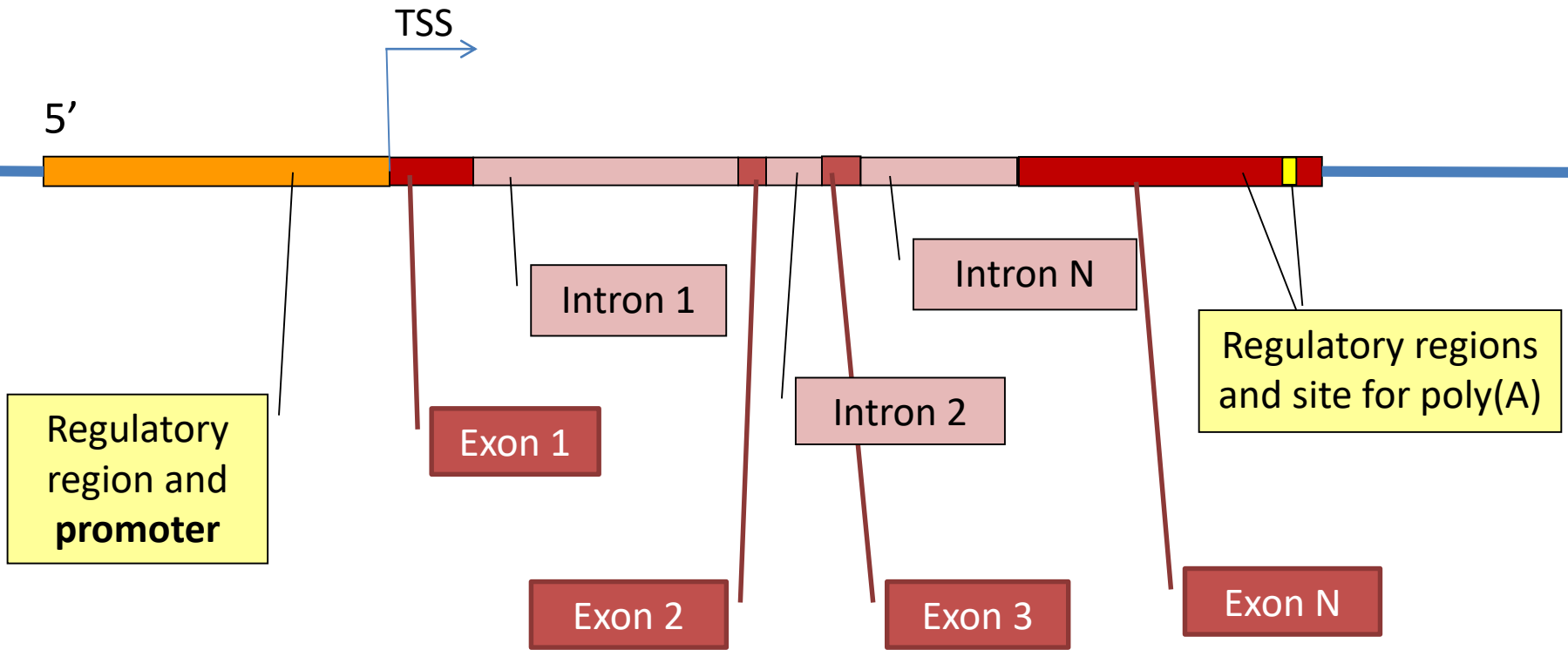
About this gene This gene has 10 transcripts ([splice variants](#)), [301 orthologues](#), [17 paralogues](#), is a member of [1 Ensembl protein family](#) and is associated with [75 phenotypes](#).

Transcripts [Hide transcript table](#)

Filter

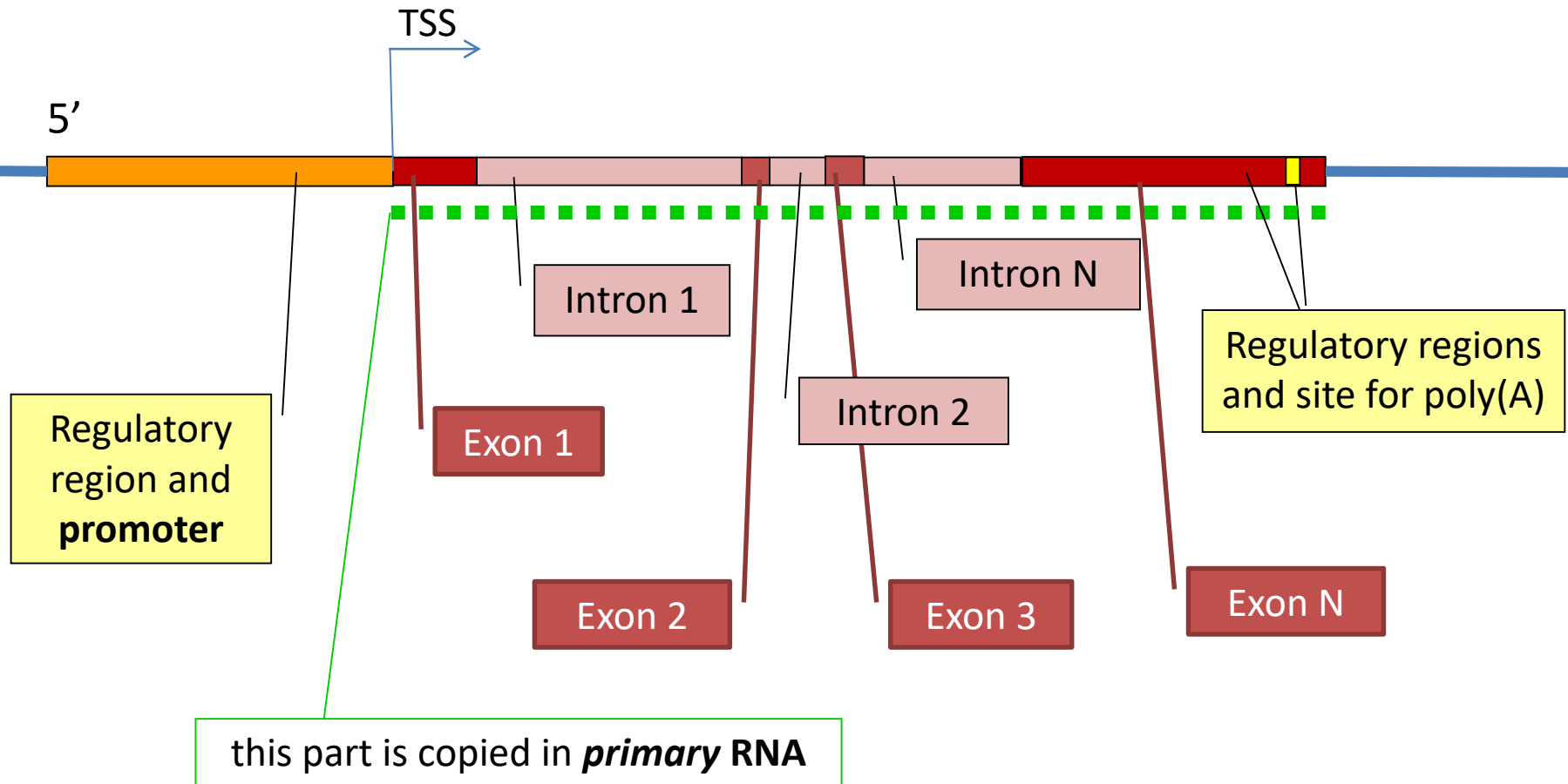
Variant ID	Chr: bp	Alleles	Global MAF	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA coord	SIFT	Poly-Phe n	CAD	REV EL	Meta LR	Mutation Assessment	Transcript
rs1201213753	21:41498140	G/A/C	-	SNP	dbSNP		-	5 prime UTR variant	-	-	-	-	-	-	-	-	ENST00000332149.10
rs1201213753	21:41498140	G/A/C	-	SNP	dbSNP		-	5 prime UTR variant	-	-	-	-	-	-	-	-	ENST00000332149.10
COSV59822939	21:41498140	COSMIC_MUTATION	-	somatic SNV	COSMIC		-	5 prime UTR variant	-	-	-	-	-	-	-	-	ENST00000332149.10
rs1331618004	21:41498141	T/C	-	SNP	dbSNP		-	5 prime UTR variant	-	-	-	-	-	-	-	-	ENST00000332149.10

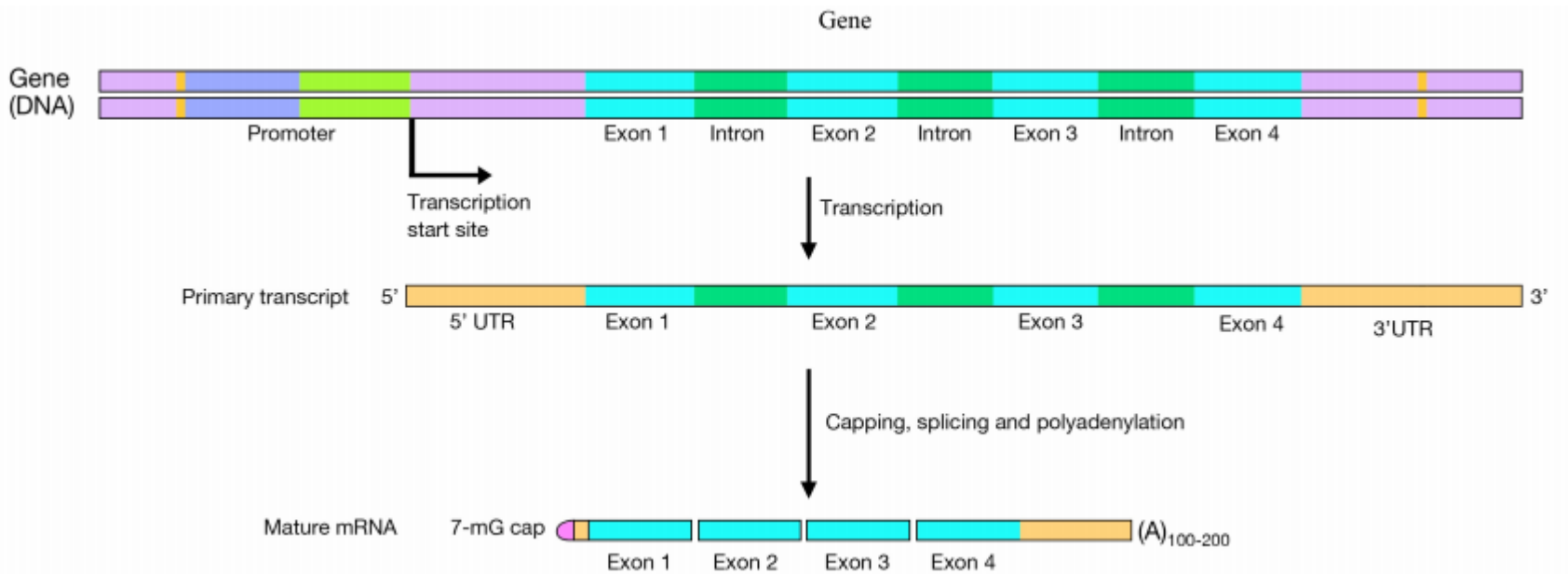
Eukaryotic protein coding genes



TSS= Transcription start site

Eukaryotic protein coding genes

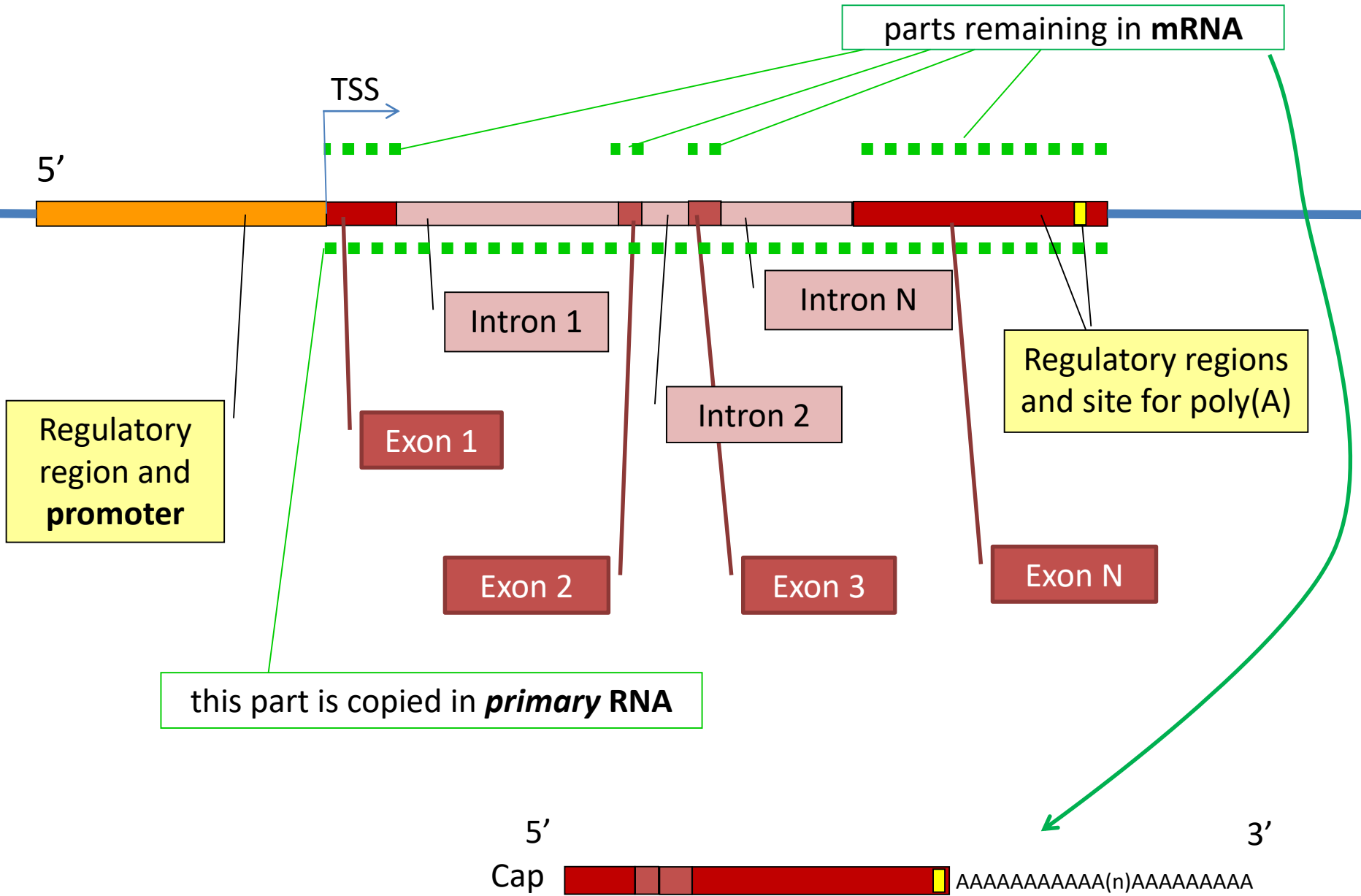




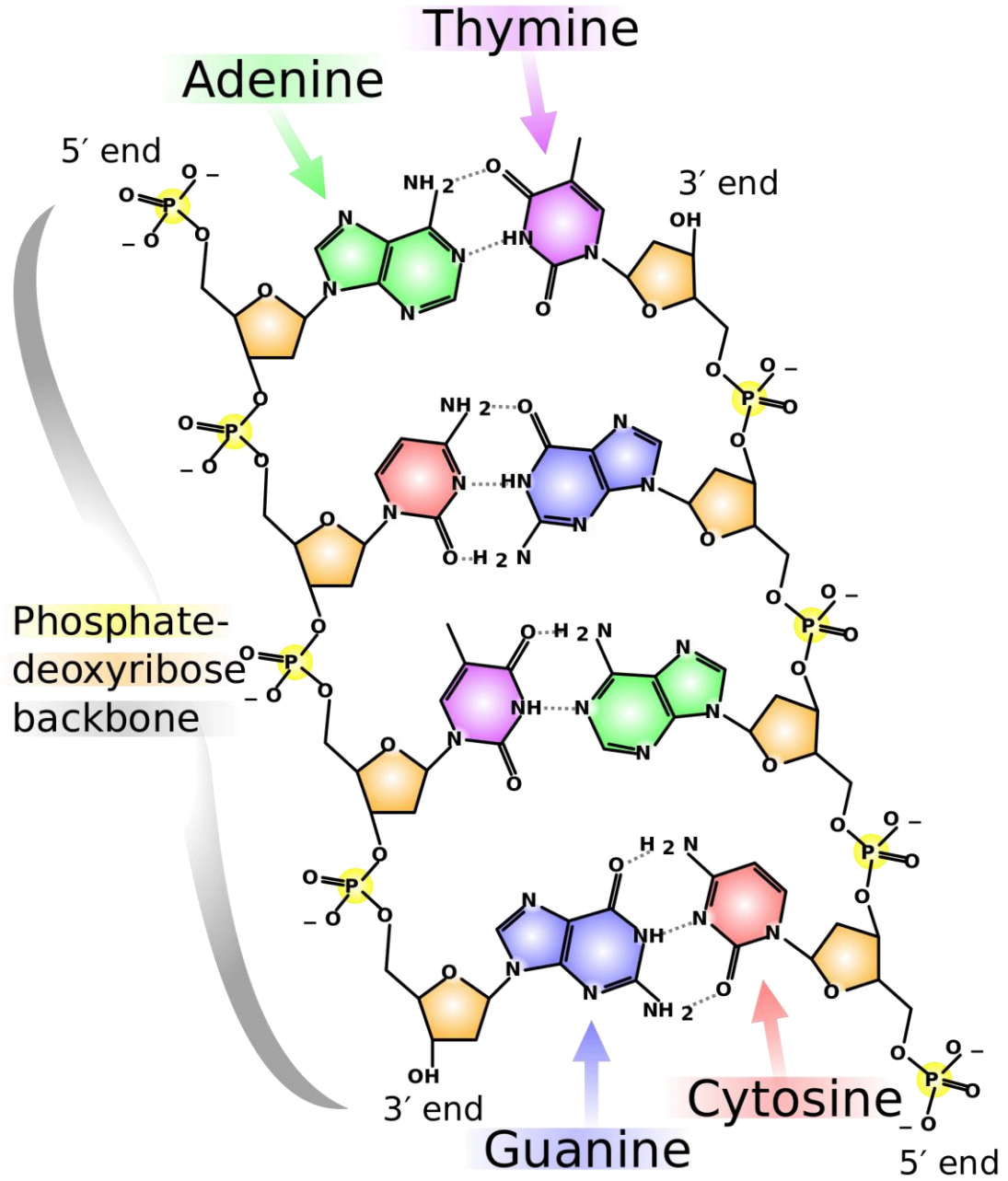
5' UTR: untraslated region

3' UTR: untraslated region

Eukaryotic protein coding genes



COMPLEMENTARY In molecular biology

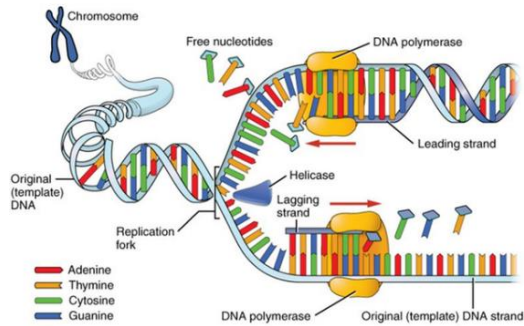


Complementary in molecular biology

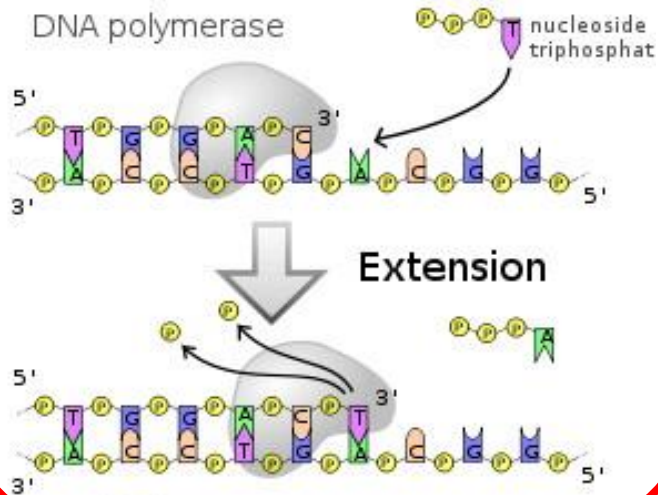
is the base principle of **DNA replication** and **transcription**.

The base complement A=T shares two hydrogen bonds
the base pair G≡C has three hydrogen bonds

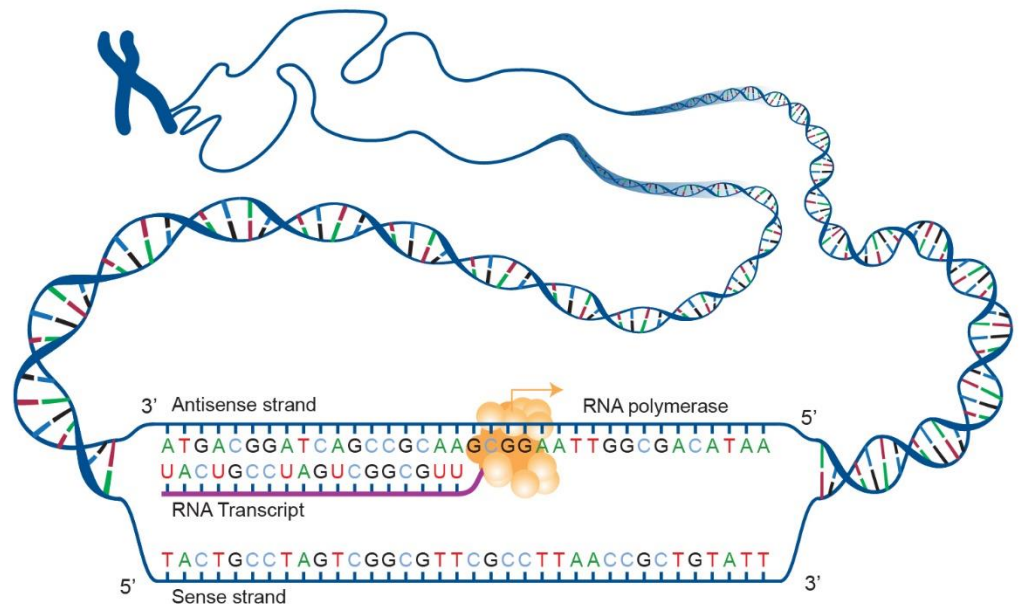
DNA REPLICATION



DNA polymerase



TRANSCRIPTION



Complementary

is the base principle of the molecular biology techniques to study gene expression.

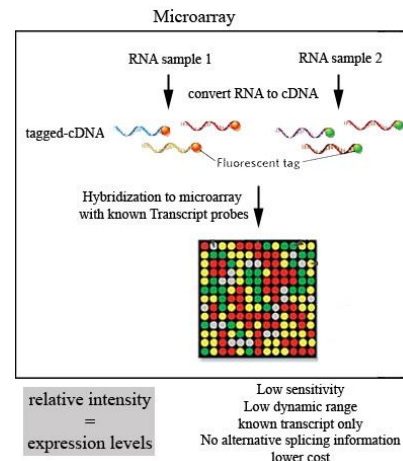
ANALYSIS of SINGLE GENE

PCR

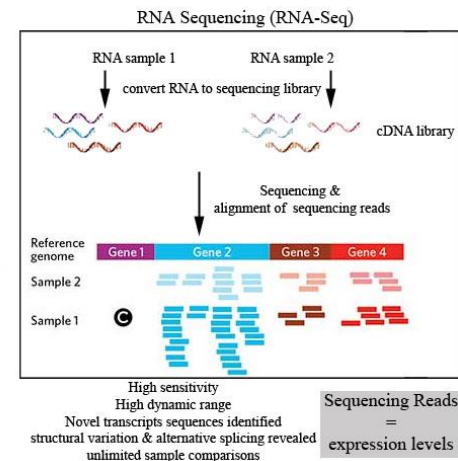


ANALYSIS of MULTIPLE GENES

microarray



RNA-Seq



ANALYSIS SINGLE GENE

PCR

Template DNA



Primers bind to template

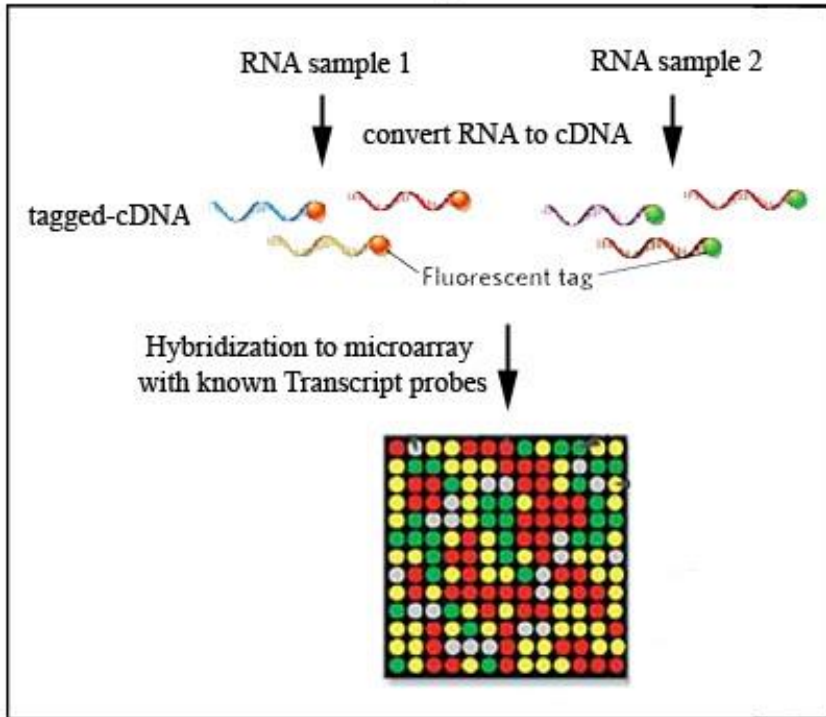


Tag polymerase extends primers



MULTIPLE GENES ANALYSIS

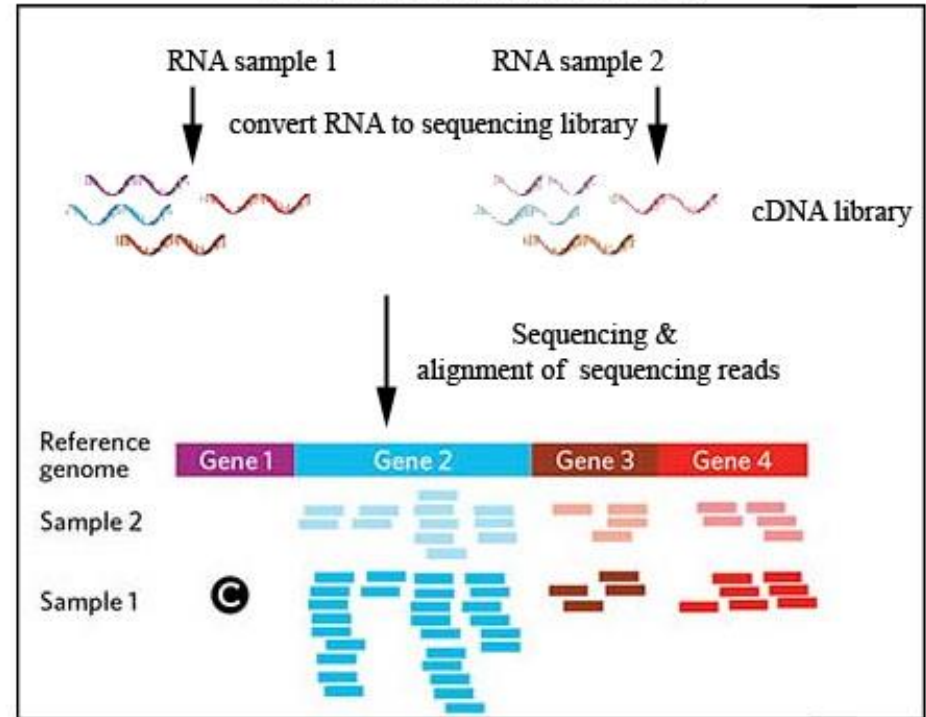
Microarray



relative intensity
=
expression levels

Low sensitivity
Low dynamic range
known transcript only
No alternative splicing information
lower cost

RNA Sequencing (RNA-Seq)



High sensitivity
High dynamic range
Novel transcripts sequences identified
structural variation & alternative splicing revealed
unlimited sample comparisons

Sequencing Reads
=
expression levels

Complementary

is the base principle of therapeutic approach based on Antisense oligonucleotide (ASO)

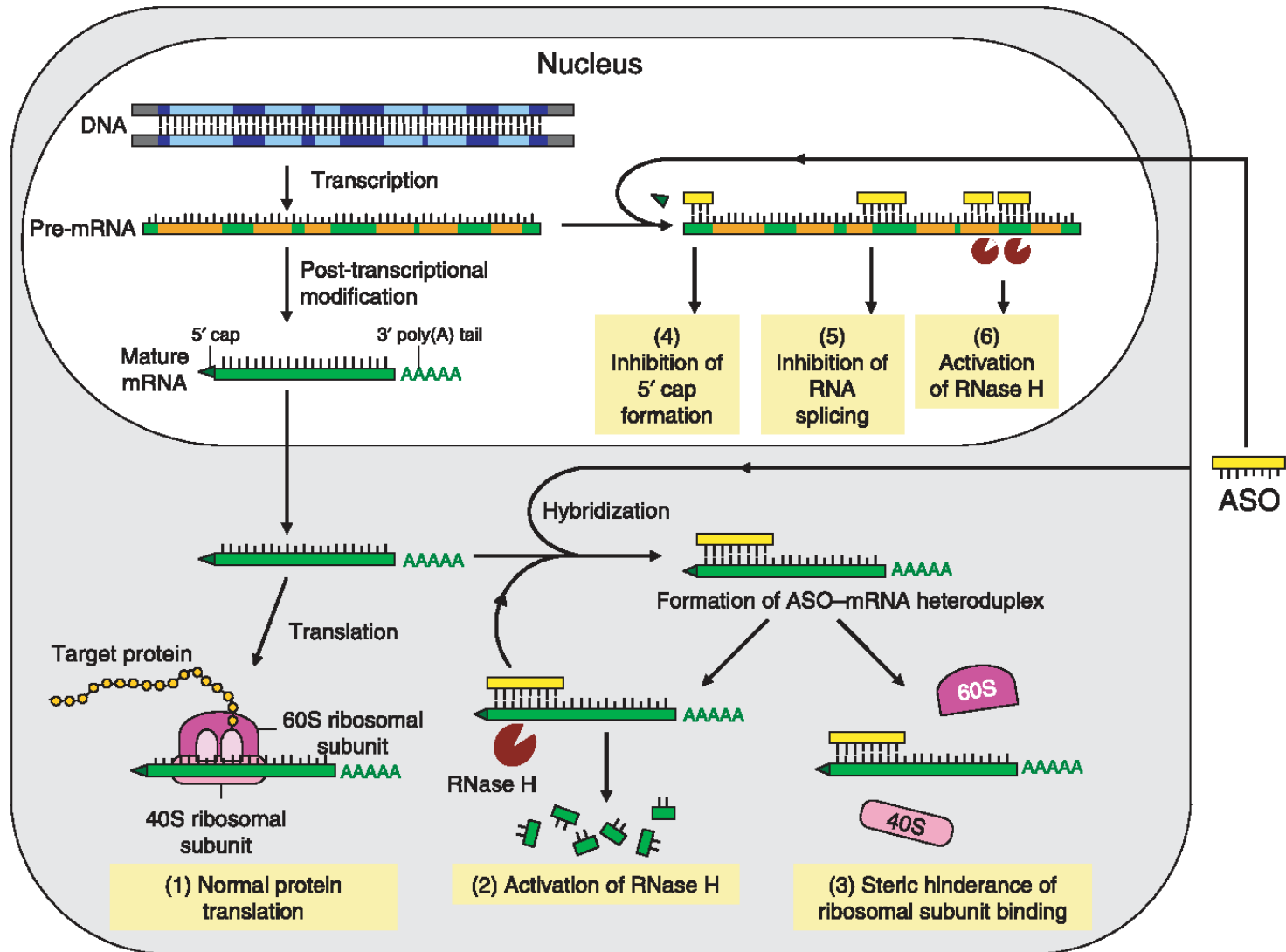


Fig. 1. Mechanism of Antisense Oligonucleotide (ASO) action. 1. ASO hybridizes with pre-mRNA in the nucleus. 2. The ASO inhibits

Human protein atlas

Q5

proteinatlas.org/ENSG00000184012-TMPRSS2/tissue

THE HUMAN PROTEIN ATLAS

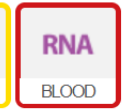
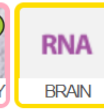
MENU HELP NEWS

TMPRSS2

Search

Fields

TMPRSS2



TISSUE ATLAS

PRIMARY DATA

GENE/PROTEIN

ANTIBODIES
AND
VALIDATION



Dictionary



GENERAL INFORMATION¹

Gene name ¹	TMPRSS2
Gene description	Transmembrane serine protease 2
Protein class ¹	Cancer-related genes Enzymes
Predicted location ¹	Intracellular,Membrane
Number of transcripts ¹	6

HUMAN PROTEIN ATLAS INFORMATION¹

RNA tissue specificity ¹	Tissue enhanced (intestine, pancreas, prostate)
RNA tissue distribution ¹	Detected in many
Protein evidence ¹	Evidence at protein level
Protein expression ¹	Membranous expression in prostate, gastrointestinal tract, kidney and pancreas.

IMMUNOHISTOCHEMISTRY DATA RELIABILITY

Data reliability description ¹	Antibody staining mainly consistent with RNA expression data.
Reliability score ¹	Enhanced
Antibodies ¹	HPA035787

SHOW MORE

Human protein atlas

Q2

proteinatlas.org/ENSG00000184012-TMPRSS2/tissue

THE HUMAN PROTEIN ATLAS

TMPRSS2

Search

Fields »

MENU HELP NEWS

TISSUE ATLAS

PRIMARY DATA

GENE/PROTEIN

ANTIBODIES AND VALIDATION



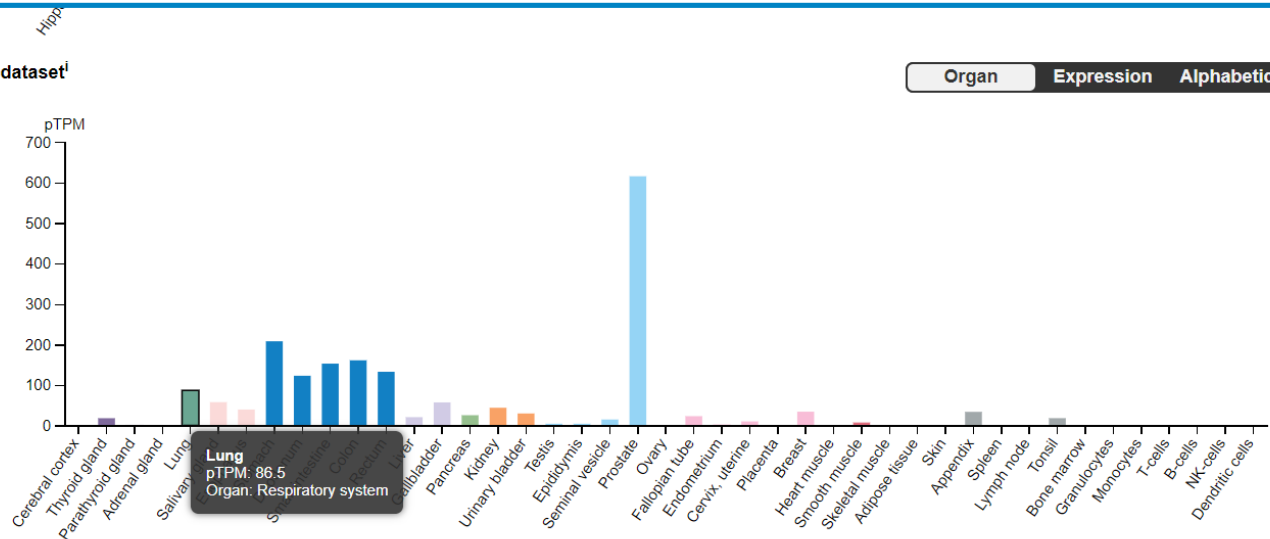
Dictionary



Tissue proteome



HPA dataset¹



Organ

Expression

Alphabetical

GTEx dataset¹

Organ

Expression

Alphabetical

TISSUE ATLAS

PRIMARY DATA

GENE/PROTEIN

ANTIBODIES AND VALIDATION



Dictionary



Tissue proteome



RNA EXPRESSION OVERVIEW¹

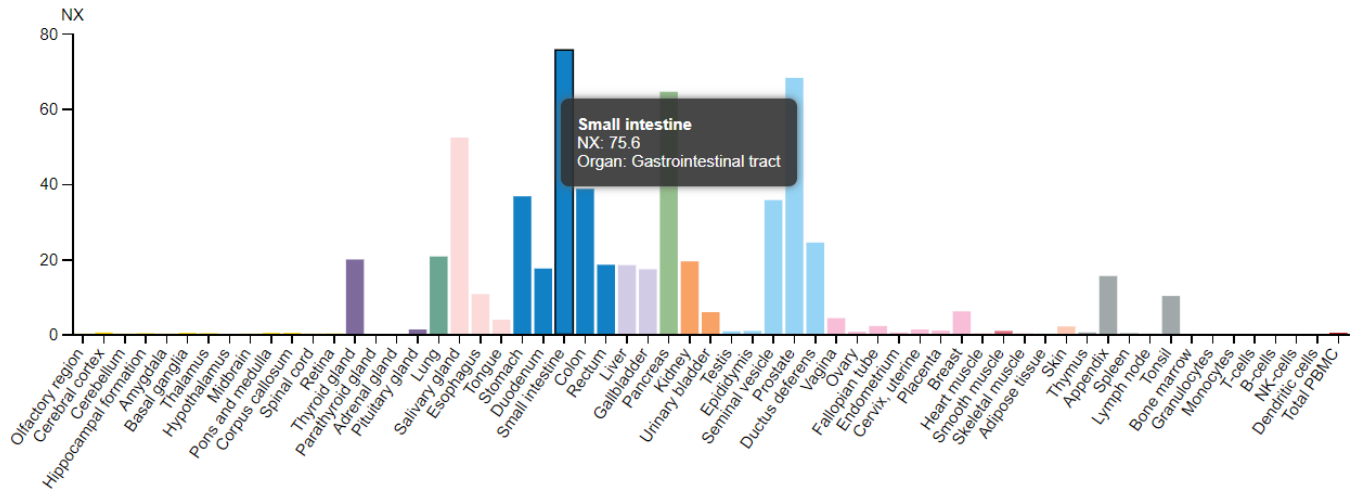
Consensus dataset¹

RNA tissue specificity: Tissue enhanced (intestine, pancreas, prostate)

Organ

Expression

Alphabetical



Q9



Single Cell Expression Atlas

Single cell gene expression across species

- Home
- Gene search
- Browse experiments
- Release notes
- Help
- Support

Lung Cell Atlas: Single cell RNA sequencing analysis of fresh resected human lung tissue - Drop-seq dataset

Single-cell RNA-Seq mRNA baseline

Number of cells: 4,599

Organism: *Homo sapiens*

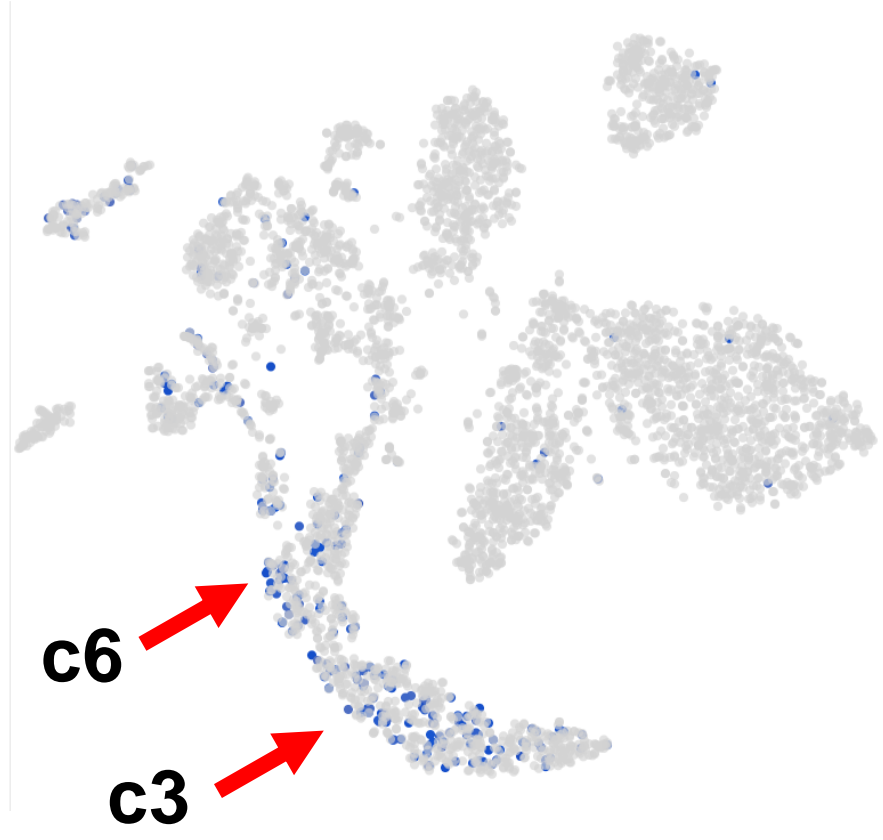
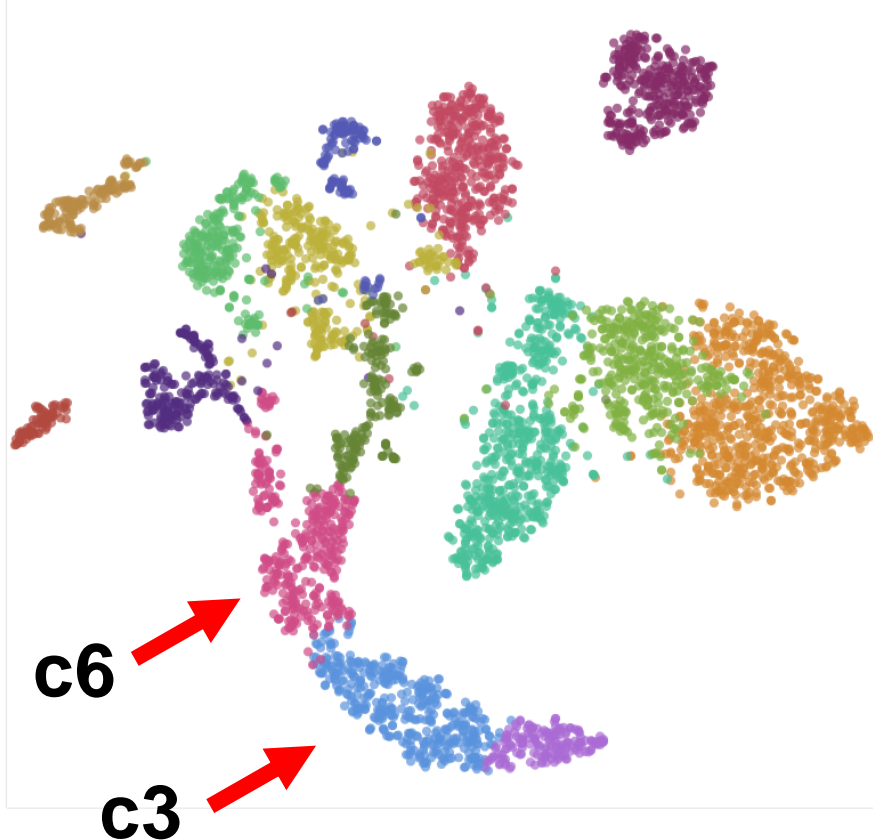
Publication:

- Vieira Braga FA, Kar G, Berg M, Carpaj OA, Polanski K et al. (2019) *A cellular census of human lungs identifies novel cell states in health and in asthma.*

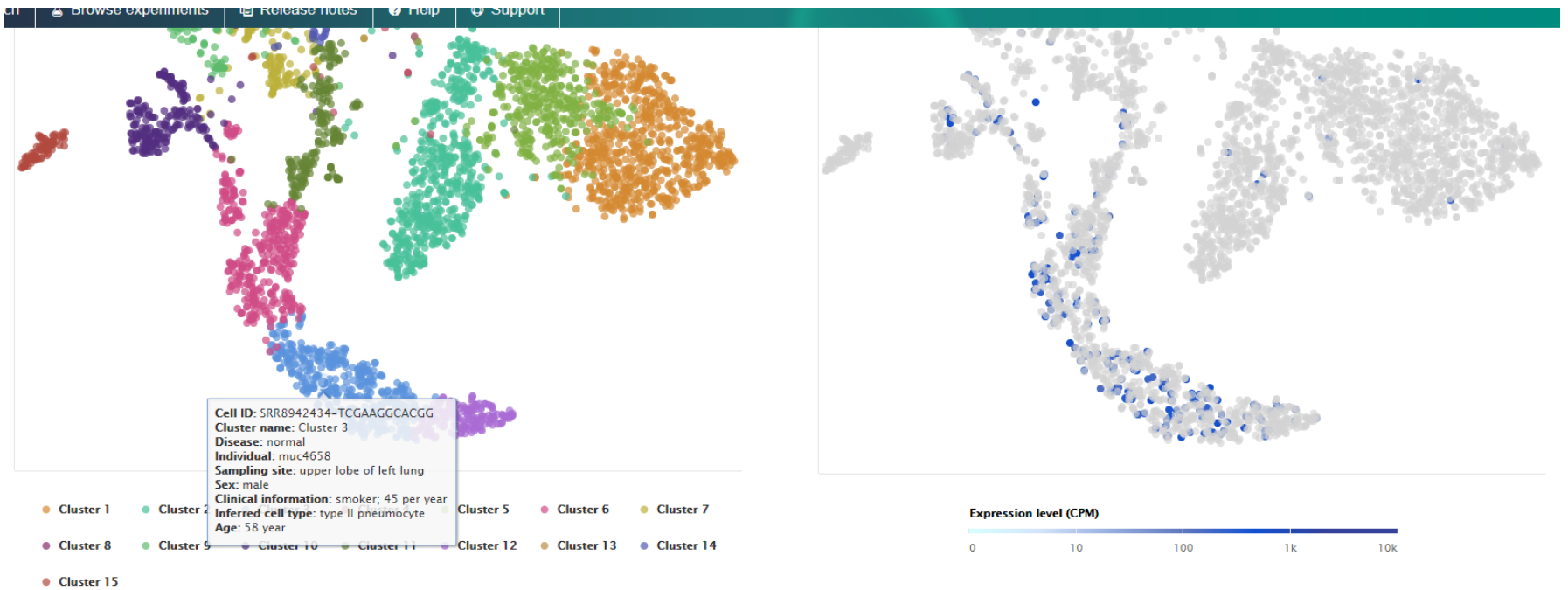
ebi.ac.uk/gxa/sc/experiments/E-GEOD-130148/results/tsne?geneId=ENSG00000184012

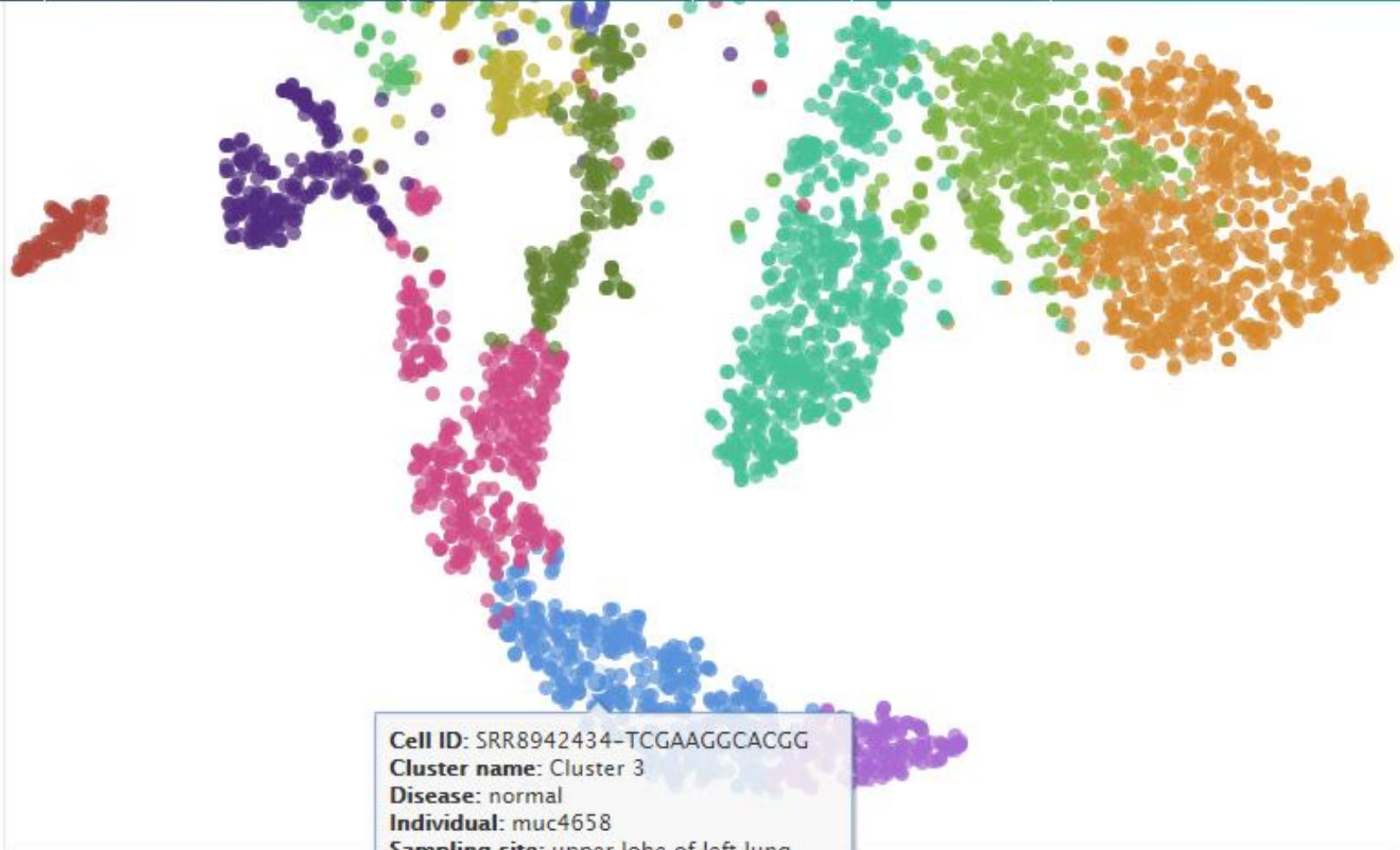


- Browse experiments
- Release notes
- Help
- Support



Q10












Cell ID: SRR8942434-TCGAAGGCACGG
Cluster name: Cluster 3
Disease: normal
Individual: muc4658
Sampling site: upper lobe of left lung
Sex: male
Clinical information: smoker; 45 per year
Inferred cell type: type II pneumocyte
Age: 58 year



- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5
- Cluster 6
- Cluster 7
- Cluster 8
- Cluster 9
- Cluster 10
- Cluster 11
- Cluster 12
- Cluster 13
- Cluster 14
- Cluster 15

Q8

Adjusted p -value	Log ₂ -fold change	Gene	Condition	Category	Pathway
2.0205 × 10 ⁻⁴⁰	5.6	TMPRSS2	'replicatively senescent; grown until growth arrest' vs 'spontaneously immortal cells; control'	phenotype, treatment	Pathway Profiling of Replicative and Induced Senescence
		TMPRSS2	'interleukin 13; 100 nanogram per milliliter; air-liquid interface (day 14)' vs 'untreated; submerged (day 8)'	compound, growth condition	Genome-wide expression profiling of an in vitro model for studying esophageal epithelial differentiation
		TMPRSS2	'tumor' vs 'normal'	clinical information	Identification of a Novel Angiogenesis and Tumor Suppressor Gene Rab25 in Esophageal Squamous Cell Carcinoma

		TMPRSS2	'replicatively senescent; grown until growth arrest' vs 'spontaneously immortal cells; control'	phenotype, treatment	Pathway Profiling of Replicative and Induced Senescence
		TMPRSS2	'interleukin 13; 100 nanogram per milliliter; air-liquid interface (day 14)' vs 'untreated; submerged (day 8)'	compound, growth condition	Genome-wide expression profiling of an in vitro model for studying esophageal epithelial differentiation
		TMPRSS2		information	Identification of a Novel Angiogenesis and Tumor Suppressor Gene Rab25 in Esophageal Squamous Cell Carcinoma
		TMPRSS2		type	Gene expression profiling by RNA-seq of human iPSC and iPSC-derived cardiomyocytes from an Yoruban individual (NA19101)
		TMPRSS2		, time	Transcriptome analysis of human brain microvascular endothelial cells response to interactive DIII domain of protein E of WNV and interactive DIII domain of protein E of TBEV using RNA-seq
		TMPRSS2		treatment	Rapid neurogenesis through transcriptional activation in human stem cell (RNA-Seq)
		TMPRSS2	'lung carcinoma' vs 'normal'	biopsy site, disease	Gene expression profiling of the adjacent airway field cancerization in early stage NSCLC

Genome-wide expression profiling of an in vitro model for studying esophageal epithelial differentiation
'interleukin 13; 100 nanogram per milliliter; air-liquid interface (day 14)' vs 'untreated; submerged (day 8)'

Property	Test value (N=3)	Reference value (N=3)
compound	interleukin 13 100 nanogram per milliliter	untreated
growth condition	air-liquid interface (day 14)	submerged (day 8)
cell line	EPC2-hTERT	EPC2-hTERT
cell type	hTERT-immortalized human esophageal epithelial cell	hTERT-immortalized human esophageal epithelial cell
organism	Homo sapiens	Homo sapiens

TCGA

Q5

cbioportal.org/datasets



Join our new webinar series to learn how to use cBioPortal effectively. Second webinar **Thursday May 7th 11am-12pm EDT**. [Click for More Info!](#) or [View Recorded Webinars](#)



[Data Sets](#) [Web API](#) [R/MATLAB](#) [Tutorials/Webinars](#) [FAQ](#) [News](#) [Visualize Your Data](#) [About](#)

Login

Datasets

The table below lists the number of available samples per cancer study and data type.

pancrea

Columns

Name ▲	Reference	All	Sequenced	CNA	RNA-Seq
Acinar Cell Carcinoma of the Pancreas (JHU, J Pathol 2014)	Jial et al. J Pathol 2014	23	23	0	0
Cystic Tumor of the Pancreas (Johns Hopkins, PNAS 2011)	Wu et al. PNAS 2011	32	32	0	0
Pancreatic Adenocarcinoma (ICGC, Nature 2012)	Blankin et al. Nature 2012	99	99	0	0
Pancreatic Adenocarcinoma (QCMG, Nature 2016)	Bailey et al. Nature 2016	456	383	0	96
Pancreatic Adenocarcinoma (TCGA, Firehose Legacy)		186	150	184	179
Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas)	TCGA, Cell 2018	184	179	183	177
Pancreatic Cancer (UTSW, Nat Commun 2015)	Witkiewicz et al. Nat Commun 2015	109	109	109	0
Pancreatic Neuroendocrine Tumors (Johns Hopkins University, Science 2011)	Jiao et al. Science 2011	10	10	0	0
Pancreatic Neuroendocrine Tumors (Multi-Institute, Nature 2017)	Scarpa et al. Nature 2017	98	98	0	0

Data sets of published studies were curated from literature. Data sets of legacy TCGA studies were downloaded from the [Broad Institute Firehose](#) and updated quarterly.

TCGA Q5

Modify Query



Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas)

All samples (184 patients/samples) - TMPRSS2

OncoPrint

Cancer Types Summary

Plots

Mutations

Co-expression

Comparison

Survival

CN Segments

Groups: (drag to reorder)

Altered group (0)

Unaltered group (184)

TMPRSS2 (0)

Select all | Deselect all

Overlap

Survival

Clinical

Mutations

Copy-number

mRNA

Survival types: Overall

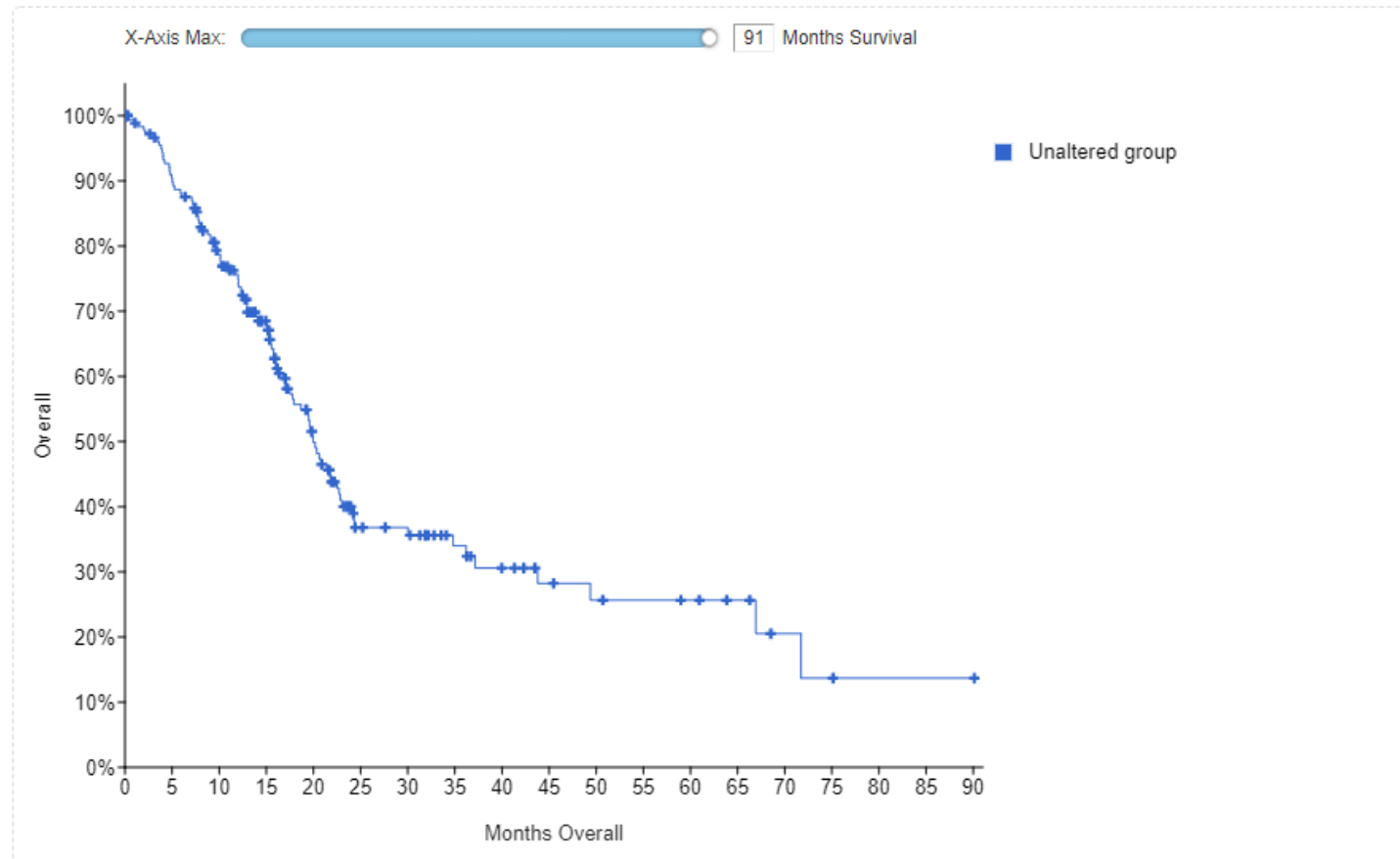
Disease Free

Progression Free

Disease-specific

Overall

Overall patient survival status.



TCGA

Q6

Join our new webinar series to learn how to use cBioPortal effectively. Second webinar **Thursday May 7th 11am-12pm EDT** [Click for More Info!](#) or [View Recorded Webinars](#)



[Data Sets](#) [Web API](#) [R/MATLAB](#) [Tutorials/Webinars](#) [FAQ](#) [News](#) [Visualize Your Data](#) [About](#)

Logi

Modify Query



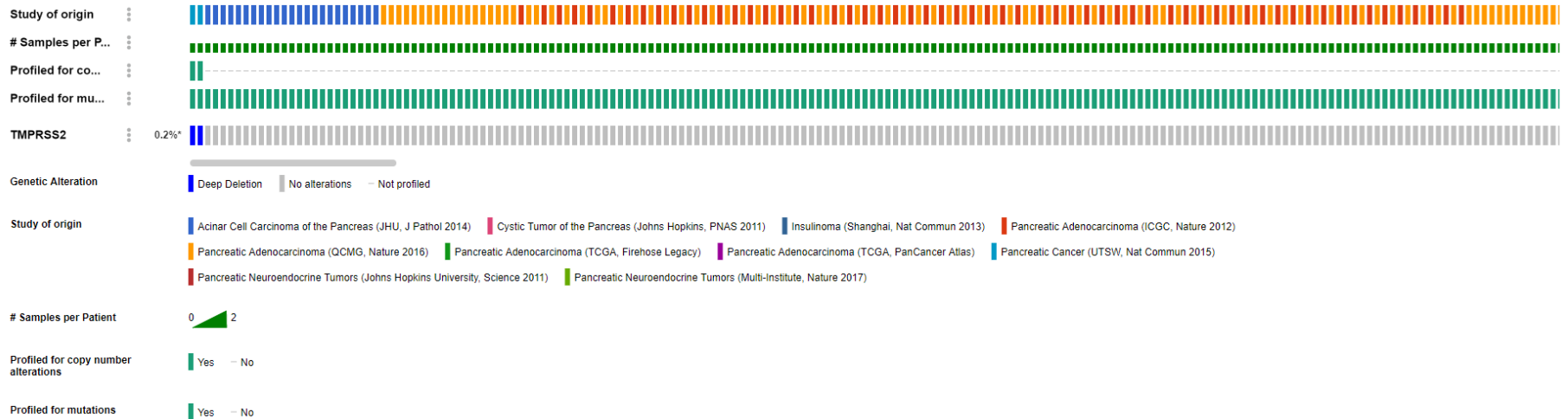
Combined Study (1207 samples)

Querying 1206 patients / 1207 samples in 10 studies - TMPRSS2

Queried gene is altered in
 • 2 (<0.1%) of queried patients
 • 2 (<0.1%) of queried samples

[OncoPrint](#) [Cancer Types Summary](#) [Mutations](#) [Comparison](#) [Survival](#) [CN Segments](#) [Pathways](#) [Expression](#) [Download](#)

Add Clinical Tracks 217 ▾ Sort ▾ Mutations ▾ View ▾ Download ▾ 🔍 100 % 🔍 📐



What is enrichment analysis?

Enrichment analysis is a computational method for inferring knowledge about an input gene set by comparing it to annotated gene sets representing prior biological knowledge.

Enrichment analysis checks whether an input set of genes significantly overlaps with annotated gene sets.

For more information, you can read this excellent review about enrichment analysis [1].

What is an enrichment term?

Each gene set within the Enrichr database is associated with a functional term or an *enrichment term* such as a pathway, cell line, or disease. The output of Enrichr are ranked lists of terms, one list for each gene set library. The most highly ranked enrichment terms for the user's input gene list provide knowledge about the input list.

What is a gene set library?

A *gene set library* is a set of related gene sets or enrichment terms.

Each enrichment term in Enrichr's results pages is organized by its gene set library. These libraries have been constructed from many sources such as published studies and major biological and biomedical online databases. Others have been created for and only available through Enrichr. For example, the ChEA 2015 library is a set of functional terms representing transcription factors profiled by ChIP-seq in mammalian cells. Each term is associated with a collection of putative targets inferred from the peaks identified in each ChIP-seq study.

enrichr

Q6

<http://amp.pharm.mssm.edu/Enrichr/>

← → ↻ Non sicuro | amp.pharm.mssm.edu/Enrichr/#stats

ENCODE_Histone_Modifications_2015	412	29065	2123	↓
ENCODE_TF_ChIP-seq_2014	498	21493	3713	↓
ENCODE_TF_ChIP-seq_2015	816	26382	1811	↓
Enrichr_Libraries_Most_Popular_Genes	121	5902	109	↓
Enrichr_Submissions_TF-Gene_Cocurrence	1722	12486	299	↓
Epigenomics_Roadmap_HM_ChIP-seq	383	22288	4368	↓
ESCAPE	315	25651	807	↓
Gene_Perturbations_from_GEO_down	2460	30832	302	↓
Gene_Perturbations_from_GEO_up	2460	31132	298	↓
GeneSigDB	2139	23726	127	↓
Genome_Browser_PWMs	615	13362	275	↓
GO_Biological_Process_2013	941	7682	78	↓
GO_Biological_Process_2015	5192	14264	58	↓
GO_Biological_Process_2017	3166	13822	21	↓
GO_Biological_Process_2017b	10125	13247	49	↓
<u>GO_Biological_Process_2018</u>	5103	14433	36	↓
GO_Cellular_Component_2013	205	7324	172	↓
GO_Cellular_Component_2015	641	13236	82	↓
GO_Cellular_Component_2017	636	10427	38	↓
GO_Cellular_Component_2017b	816	8002	143	↓
GO_Cellular_Component_2018	446	8655	61	↓
GO_Molecular_Function_2013	402	8469	122	↓
GO_Molecular_Function_2015	1136	12753	57	↓
GO_Molecular_Function_2017	972	10601	25	↓
GO_Molecular_Function_2017b	3271	10089	45	↓