

Data sources and species

Data sources

A Web Service is available to query currently available data sources:

<http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/#!/Meta/getVersion>

Please, find below a summary of all data sources information:

Category	Data source	Version/Date		
		CellBase v3 (March 2015)	CellBase v4 (June 2016)	CellBase v5 (Autumn 2020)
Core	Ensembl Core	v79	v82	v99
Protein	UniProt	March 2015	Release 2015_10	Release 2020_02
	InterPro	v50	v54	--
	Polyphen2/Sift	Ensembl 79	Ensembl v82	Ensembl v99
Variation	Ensembl Variation	v79	v82	--
	1000 genomes project		Phase 3 2016-05	Phase 3 2016-05
	ExAC		0.3.1	--
	GoNL		Release 5	--
	UK10K		2016-05	2016-05
	ESP		2016-05	--
	gnomAD Genomes			v3
	TOPMed			2020-04
	DiscovEHR			2020-04
	GenomeAsia 100K			2020-04
Regulatory	Ensembl Regulatory	v79	v82	v99
	Ensembl Motif features			v99
	mirBase			22.1
	mirTarBase			7.0
Conservation	PhastCons		June 2016	--
	PhyloP		June 2016	--
	GERP++		June 2016	--
Clinical	ClinVar	March 2015	2016-12	2020-02
	COSMIC	v71	v79	v91
	HPO		2015-11	2020-04
	DisGeNET		Version 3.0	7.0
	Disease ontology			2020-05
Biological Networks	Reactome	v51	June 2016	--
	IntAct	March 2015	June 2016	--
Ontologies	Gene Annotation			2020-05
	Gene ontology (basic)			2020-05
Others	DGIdb		2.0	v3.0.2

Table of Contents:

- [Data sources](#)
- [Species](#)
- [\[DEPRECATED\] Release v3](#)
 - [Data sources and versions](#)
 - [Core features](#)
 - [Protein](#)
 - [Variation](#)
 - [Regulatory](#)
 - [Clinical association](#)
 - [Conservation scores](#)
 - [Systems biology](#)
 - [Others](#)
 - [Available species](#)

	Gene Expression Atlas		June 2016	2.0.14
	CADD		v1.3	--
	gnomAD constraints			2.1.1
	Repeats			Dec 2013

Species

A Web Service is available to query all available species and assemblies:

<http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/rest/v4/meta/species>

Please, find below a summary of available species data:

Id	Common Name	Scientific Name	Assembly	
			CellBase v3 (March 2015)	CellBase v4 (June 2016)
hsapiens	Human	Homo sapiens	GRCh37.p13	GRCh37.p13, GRCh37.p8
mmusculus	Mouse	Mus musculus	GRCm38.p2	GRCm38.p5
rnorvegicus	Rat	Rattus norvegicus	Rnor_5.0	Rnor_6.0
ptroglodytes	Chimp	Pan troglodytes	CHIMP2.1.4	CHIMP2.1.4
agambiae	Anopheles	Gambiae	AgamP4	AgamP4
athaliana	Arabidopsis	Thaliana	THAIR10	THAIR10
btaurus	Bos	Taurus	UMD3.1	UMD3.1
celegans	Caenorhabditis	Elegans	WBcel235	WBcel235
cfamiliaris	Canis	Familiaris	CanFam3.1	CanFam3.1
csabaeus	Chlorocebus	Sabaeus	ChISab1.1	ChISab1.1
cintestinales	Ciona	Intestinalis	KH	KH
drerio	Danio	Rerio	zv9	GRCz10
dmelanogaster	Drosophila	Melanogaster	BDGP6	BDGP6
ggallus	Gallus	Gallus	4.0	5.0
gmax	Glycine	Max	V1.0	V1.0
ggorilla	Gorilla	Gorilla	gorGor3.1	gorGor3.1
lmajor	Leishmania	Major	ASM272v2	ASM272v2
osativa	Oryza	Sativa	IRGSP-1.0	IRGSP-1.0
olatipes	Oryzias	Latipes	HdrR	HdrR
oaries	Ovis	Aries	Oar_v3.1	Oar_v3.1
pfalciparum	Plasmodium	Falciparum	ASM276v1	ASM276v1
scerevisiae	Saccharomyces	Cerevisiae	R64-1-1	R64-1-1
slycopersicum	Solanum	Lycopersicum	SL2.40	
sscrofa	Sus	Scrofa	Sscrofa10.2	Sscrofa10.2
vvinifera	Vitis	Vinifera	IGGP_12x	IGGP_12x
zmays	Zea	Mays	AGPv3	AGPv3
afumigatus	Aspergillus	Fumigatus		A1163
anidulans	Aspergillus	Nidulans		ASM1142v1
ecoli	Esherichia	Coli		HUSEC2011CHR1
fcatus	Felis	Catus		
mpneumoniae	Mycoplasma	Pneumoniae		M129 (ASM2734v1)

sbicolor	Sorghum	Bicolor		Sorbi1
----------	---------	---------	--	--------

[DEPRECATED] Release v3

Data sources and versions

Core features

- **Ensembl Release 79** (March 2015): Core data for all species are built from Ensembl v79, so ****_Homo sapiens_**** uses now assembly GRCh38.p2 and GENCODE 22, you can query the rest of assemblies at **[**Ensembl table of assemblies**]**(<http://www.ensembl.org/info/website/archives/assembly.html>). These includes genome sequence, gene sets, variation and regulation. **Ensembl Release 75** (Feb 2014) is used only for keeping old ****_Homo sapiens_ GRCh37**** assembly.

Protein

- **UniProt** (Release March 2015)
- **InterPro v50** v50 (Release Feb 2015)
- Polyphen2/Sift from Ensembl v79

Variation

- Ensembl v79 Variation (dbSNP 142)
- Population frequencies: 1000 genomes project, ESP (ExAC in preparation).

Regulatory

- Ensembl v79 Regulatory
- miRNAs:
- miRBase (Release 21)
- miRTarBase (Release 4.5)
- TargetScan (Release 6.0)

Clinical association

- ClinVar (Release March 2015)
- GWAS Catalog
- COSMIC v71 (Release March 2015)

Conservation scores

- PhastCons
- PhyloP
- (GERP++ in preparation)

Systems biology

- IntAct (Release March 2015)
- (Reactome 51 in preparation)

Others

- Gene Expression Atlas
- Gene disease association: DisGeNET
- DGldb: <http://dgidb.genome.wustl.edu/>

Available species

`species`	Name	Scientific name	Assembly
hsapiens	human	Homo sapiens	GRCh37.p13
mmusculus	mouse	Mus musculus	GRCm38.p2
rnorvegicus	rat	Rattus norvegicus	Rnor_5.0

ptroglodytes	chimp	Pan troglodytes	CHIMP2.1.4
ggorilla	gorilla	Gorilla gorilla	gorGor3.1
pabelii	orangutan	Pongo abelii	PPYG2
mmulatta	macaque	Macaca mulatta	MMUL 1.0
sscrofa	pig	Sus scrofa	Sscrofa10.2
cfamiliaris	dog	Canis familiaris	CanFam 3.1
ecaballus	horse	Equus caballus	Equ Cab 2
ocuniculus	rabbit	Oryctolagus cuniculus	OryCun2.0
ggallus	chicken	Gallus gallus	Galgal4
btaurus	cow	Bos taurus	UMD3.1
fcatus	cat	Felis catus	Felis_catus_6.2
drerio	zebrafish	Danio rerio	Zv9
cintestinalis		Ciona intestinalis	KH
dmelanogaster	fruitfly	Drosophila melanogaster	BDGP 5
dsimulans		Drosophila simulans	dsim_caf1
dyakuba		Drosophila yakuba	dyak_caf1
agambiae	mosquito	Anopheles gambiae	AgamP4
celegans	worm	Caenorhabditis elegans	WS235
scerevisiae	yeast	Saccharomyces cerevisiae	R64-1-1
spombe		Schizosaccharomyces pombe	ASM294v2
afumigatus		Aspergillus fumigatus	TIGR
aniger		Aspergillus niger	DSM
anidulans		Aspergillus nidulans	ASM1142v1
aoryzae		Aspergillus oryzae	NITE
pfalciparum	malaria parasite	Plasmodium falciparum	3D7
lmajor		Plasmodium falciparum	ASM276v1
athaliana		Arabidopsis thaliana	TAIR10
alyrata		Arabidopsis lyrata	v.1.0
bdistachyon		Brachypodium distachyon	v1.0
osativa		Oryza sativa Indica	ASM465v1
gmax		Glycine max	V1.0
vvinifera		Vitis vinifera	IGGP_12x
zmays		Zea mays	AGPv3