

Download Sources

Optional - Skip this step



We have already processed all these data and json documents are available through our FTP server for those users who wish to skip this section.

Download the JSON from here:

http://bioinfo.hpc.cam.ac.uk/downloads/cellbase/v4/homo_sapiens_grch37/mongodb/

http://bioinfo.hpc.cam.ac.uk/downloads/cellbase/v4/homo_sapiens_grch38/mongodb/

And follow the instructions here: [Load Data](#)

The first step to creating a CellBase instance is to download the data files. Download can be done through the CellBase CLI.

```
$ cellbase/build/bin$ ./cellbase.sh download --data genome,gene
```

The --data argument is required and is a comma separated list of data types to download. See below for the full list.

Type	Data sources
genome	• Ensembl
gene	• Ensembl • DGIdb • UniProt gene mappings • Gene Expression Atlas • HPO gene annotation • GNomad
variation **	• 1000 genomes • ExAC • GoNL • UK10K • ESP
variation_functional_score	• CADD
regulation	• Ensembl
protein	• UniProt • InterPro • Polyphen/Sift
conservation **	• PhaseCons • PhyloP • GERP++

clinical_variants **	• ClinVar • COSMIC • HPO • DisGeNET
repeats	• UCSC
svs	• DGV
all **	Downloads all of the above

See [Download Sources](#) for details on versions and available organisms.

** Please note that many files are very large and can take several hours to download.

For example, to download all human (GRCh37) data from all sources and save it into the ``tmp/data/cellbase/v4/`` directory, run:

```
cellbase/build/bin$ ./cellbase-admin.sh download -a GRCh37 --common
/tmp/data/cellbase/v4/common/ -d all -o /tmp/data/cellbase/v4/ -s
hsapiens
```

If download was successful, you can proceed to building the json objects that should be loaded into the corresponding database: [Building the CellBase database](#)