

A step-by-step guide for generating the corrected gene proximity (CGP) map from a genome-wide Hi-C data file (on a Windows machine)

1. Go to <https://julialang.org/downloads/> to download and install the latest version of Julia (version 1.1, June 2019).
2. Download and unzip the file **CGPmap.zip** to a folder, e.g.,
C:\Users\XXXXXX\Downloads\CGPmap.
3. As an example, this instruction uses a Hi-C file named
ENCODE3-A549C-HindIII-R1__hg19__genome__C-40000-raw.hdf5,
which is downloaded from
<https://www.encodeproject.org/experiments/ENCSR444WCZ/>.
This pipeline works for any existing Hi-C data files in the same format.
4. Open Julia and type
julia> cd("C:\\Users\\XXXXXX\\Downloads\\CGPmap")
and press Enter to navigate to the CGPmap folder.
5. In Julia, press
]
to enter the package manager.
6. Type
**(v1.1) pkg> add HDF5 JLD MAT Graphs DataFrames LinearAlgebra
Statistics Interpolations**
and press Enter to install packages required to run the pipeline.
7. Press Backspace to quit package manager.
8. Type
julia> include("CGPMap_example.jl")
and press Enter to generate the corresponding CGP map for the Hi-C data.
This could take some time to complete, depending on the performance of the machine.
9. Concretely, this function takes a genome-wide Hi-C map as input, and produces a square, symmetric matrix, which is the CGP map, as output. The matrix has 20,227 rows and columns, corresponding to 20,227 human protein-coding genes. The matrix elements indicate the “corrected” spatial proximity between genes, where a positive value implies that two genes are closer than expectation in 3D and a negative values implies the opposite. The names and

id's for the 20,227 genes are stored in the file

all_exp_TPM_gencode19_pc_nr_no_mt.mat.

10. Type

```
julia> matfile = matopen("CorrectedGeneProximityMap.mat","w")
```

```
julia> write(matfile,"varname",CGPmap)
```

```
julia> close(matfile)
```

and press Enter to export the resultant CGP map to a MAT-file named

CorrectedGeneProximityMap.mat.

11. Perform downstream analysis using the exported MAT-file.