OBTAINING AND EXLPORING GAM DATA

Obtain the data reported in the following article:

Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, Barbieri

M, de Santiago I, Lavitas LM, Branco MR, Fraser J, Dostie J, Game L, Dillon N, Edwards PA, Nicodemi M, Pombo A.

Complex multi-enhancer contacts captured by genome architecture mapping.

*Nature*. 2017 Mar 23;**543(7646)**:519-524.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5366070/#!po=5.27638>

The data may be downloaded here:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64881>

scroll to the bottom of the page and download the following file:

GSE64881\_segmentation\_at\_30000bp.passqc.multibam.txt.gz

Explore the data:

Each column of the data is a nuclear profile (NP), which was obtained by taking a random slice from the nucleus of an embryonic stem cell from a mouse. In the file, names of NPs are contained in the first row and begin with the letter ‘F.’ An example NP name is: F10A2.

Each row of the data is a genomic window of 30,000 contiguous nucleotides[[1]](#footnote-1) (a.k.a. base pairs).

The position in the genome (i.e., the coordinates) of a genomic window consist of three components:

1. Chromosome
2. Start position on the chromosome
3. Stop position on the chromosome

The first three columns of the spreadsheet denote the coordinates of a genomic window.

The following example denotes a genomic window that begins at position ‘0’ on chromosome 1 and stops at position 30000 on chromosome 1:

1. Chromosome = chr1
2. Start position on the chromosome = 0
3. Stop position on the chromosome = 30000

Each cell (*i*, *j*) of the matrix contains either a ‘1’ or a ‘0’:

* ‘1’ – genomic window *i* was present in NP *j*
* ‘0’ – genomic window *i* was not present in NP *j*

Write a program to read the file and compute the following:

1. Number of genomic windows
2. Number of NPs
3. Min, Max and Average number of windows present in an NP (consider all NPs)
4. Min, Max and Average number of NPs that detected a window (consider all windows)

1. A *nucleotide* is the basic unit of DNA. [↑](#footnote-ref-1)