**What does this tool do?**

3DPredictor web tool can predict 3D-genome contacts map for chromosome region of interest using information about CTCF binding and gene expression (RNA-seq data).

**Input file formats**

You can upload files from your computer or using FTP from external public resources, for example from ENCODE (<https://www.encodeproject.org/>)

**CTCF data**

Input data format is narrowPeak BED file (BED6+4 format).

Example:

| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| chr1 | 840081 | 840400 | treat1\_peak\_1 | 69 | . | 4.89872 | 10.50944 | 6.91052 | 158 |
| chr1 | 919419 | 919785 | treat1\_peak\_2 | 87 | . | 5.85158 | 12.44148 | 8.70936 | 130 |
| chr1 | 937220 | 937483 | treat1\_peak\_3 | 66 | . | 4.87632 | 10.06728 | 6.61759 | 154 |

Columns which are significant for prediction contain the following data:

* **1st**: chromosome name
* **2nd**: start position of peak
* **3rd**: end position of peak
* **7th**: fold-change

**RNA-seq data**

RNA-seq data should contain fields “gene\_id” and “FPKM”. “gene\_id” is Ensemble ID or gene name.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **gene\_id** | **Gene Name\*** | **Reference** | **Strand** | **Start** | **End** | **Coverage** | **FPKM** | **TPM** |
| ENSG00000185960.8 | SHOX | chrX | + | 624344 | 646823 | 0.000000 | 0.000000 | 0.000000 |
| ENSG00000002834.13 | SHOX | chrX | + | 624344 | 659411 | 0.000000 | 0.000000 | 0.000000 |

*\* Don’t be confused: you may have “Gene name” field in your RNA-seq file, but it will be ignored. If you would like to use gene name identifiers, place them into “gene\_id” column.*

**Genome**

You should choose relevant genome assembly, which corresponds to your CTCF ChIP-seq data. *For RNA-seq data, genome version does not matter as we obtain genomic coordinates for each gene/transcript in RNA-seq data from Ensemble, and do not use genomic coordinates provided in the uploaded file.*

**Coordinates**

Please enter the chromosomal region for prediction in such format:

chr12:1,000,000-1,200,000

Note that this tools predicts all 3D-genome contacts at distance <= 1.5 MB at 5 Kb resolution. *Processing 1 MB takes ~1 h, so avoid predicting long intervals when possible.*

**Model**

Here we provide list of trained models for human or mouse data at 5 Kb resolution. We recommend to use model which was not trained on the contacts from the chromosome containing region of interest, to avoid overfitting. For example, for prediction of contacts on chr1, use model trained on even chromosomes. You can choose appropriate model from the model panel on main web page.

**Output file formats**

This tool produces a *hic file* with 3D-genome contacts that will be sent to your e-mail and *brief report* of job execution. You can visualize this file using Juicebox (https://www.aidenlab.org/juicebox/).

If your job has failed, please contact [regnveig@yandex.ru](mailto:regnveig@yandex.ru).

Enjoy the prediction! We hope it might be useful for your research☺