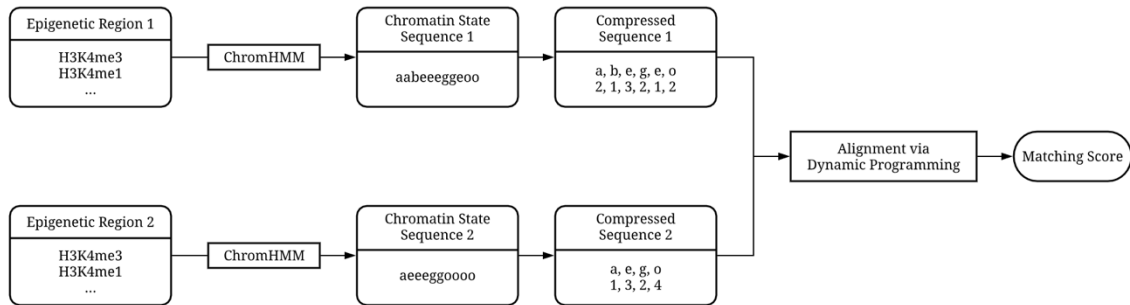


EpiAlign Web Manual

Updated on Aug 14th 2018

Haowen Zhang, Xinzhou Ge, Lingjue Xie, Wei Vivian Li, Soo Bin Kwon, Jingyi Jessica Li

Epigenome Local Alignment Search Tool (EpiAlign), a computation method developed to compare chromatin state sequences generated from multi-track epigenomic signals by the ChromHMM method. The alignment scores produced by EpiAlign algorithm quantify epigenetic similarity of genomic regions, such as genes, in different epigenome samples. The flow chart of EpiAlign algorithm is shown below:



To illustrate the utility of EpiAlign, we have done a large-scale analysis on the Roadmap epigenomic dataset, which contains epigenomic information on 127 human genomes coming from 19 different tissue types. This EpiAlign Web provides a user-friendly interface allowing users to get access to the alignment results of ~20,000 Ensembl genes and visualize the results. More specifically, the EpiAlign Web provides two features to demonstrate the applications of EpiAlign: cell-type alignment scores and pairwise alignment scores.

1. Cell-type alignment scores

An important application of EpiAlign is to assess the epigenetic similarity of chromatin state sequences over long genomic regions, such as genes, in the same or different tissues/cells. In the Section of Cell-type alignment scores, users can retrieve the alignment results of epigenomes across different tissue and cell types over a user-specified Ensembl gene. The query results include an alignment score matrix and a heatmap corresponding to the score matrix.

Step 1:

Cell-type Alignment

Ensembl Gene ID

ENSG00000154027

eg. ENSG00000154027

Apply

Input an Ensembl gene id and press the “Apply” button.

Step 2:

	ESC	iPSC	ES – deriv	Blood&T – cell	HSC&B – cell	Mesench	Epithelial	Neurosph	Thymus	Brain	Muscle	Heart	Sm. Muscle	Digestive	Other	ENCODE2012
ESC	18.29	18.50	18.44	14.63	18.51	17.97	12.28	20.25	8.62	14.44	11.88	12.82	11.81	11.75	12.47	15.27
iPSC	18.50	18.70	17.51	13.63	17.40	15.95	11.35	18.30	10.10	12.88	10.32	11.80	11.30	11.30	11.86	14.57
ES – deriv	18.44	17.51	20.56	19.25	21.96	23.06	19.89	21.50	8.83	21.13	14.29	15.40	12.53	13.06	13.99	18.96
Blood&T – cell	14.63	13.63	19.25	24.74	18.06	25.93	22.30	17.71	7.43	24.90	12.54	12.71	9.30	9.60	10.81	20.28
HSC&B – cell	18.51	17.40	21.96	18.06	24.50	21.14	15.43	21.06	9.89	17.49	14.53	14.98	13.61	13.83	14.59	18.19
Mesench	17.97	15.95	23.06	25.93	21.14	39.67	24.28	21.62	7.75	33.83	13.65	14.30	10.88	10.52	11.57	28.16
Epithelial	12.28	11.35	19.89	22.30	15.43	24.28	32.82	15.12	7.44	31.01	12.82	12.75	6.50	8.12	9.92	22.27
Neurosph	20.25	18.30	21.50	17.71	21.06	21.62	15.12	21.00	8.75	16.70	12.80	14.20	12.50	13.21	13.41	17.94
Thymus	8.62	10.10	8.83	7.43	9.89	7.75	7.44	8.75	6.00	6.85	9.10	8.90	6.50	7.62	8.00	7.12
Brain	14.44	12.88	21.13	24.90	17.49	33.83	31.01	16.70	6.85	40.40	13.04	13.16	8.78	9.17	10.50	24.49
Muscle	11.88	10.32	14.29	12.54	14.53	13.65	12.82	12.80	9.10	13.04	13.00	13.48	10.75	10.90	11.04	12.93
Heart	12.82	11.80	15.40	12.71	14.98	14.30	12.75	14.20	8.90	13.16	13.48	12.40	11.15	12.02	12.47	13.10
Sm. Muscle	11.81	11.30	12.53	9.30	13.61	10.88	6.50	12.50	6.50	8.78	10.75	11.15	11.67	11.04	10.57	9.97
Digestive	11.75	11.30	13.06	9.60	13.83	10.52	8.12	13.21	7.62	9.17	10.90	12.02	11.04	11.61	11.43	10.88
Other	12.47	11.86	13.99	10.81	14.59	11.57	9.92	13.41	8.00	10.50	11.04	12.47	10.57	11.43	11.27	11.48
ENCODE2012	15.27	14.57	18.96	20.28	18.19	28.16	22.27	17.94	7.12	24.49	12.93	13.10	9.97	10.88	11.48	19.78

[Download](#)

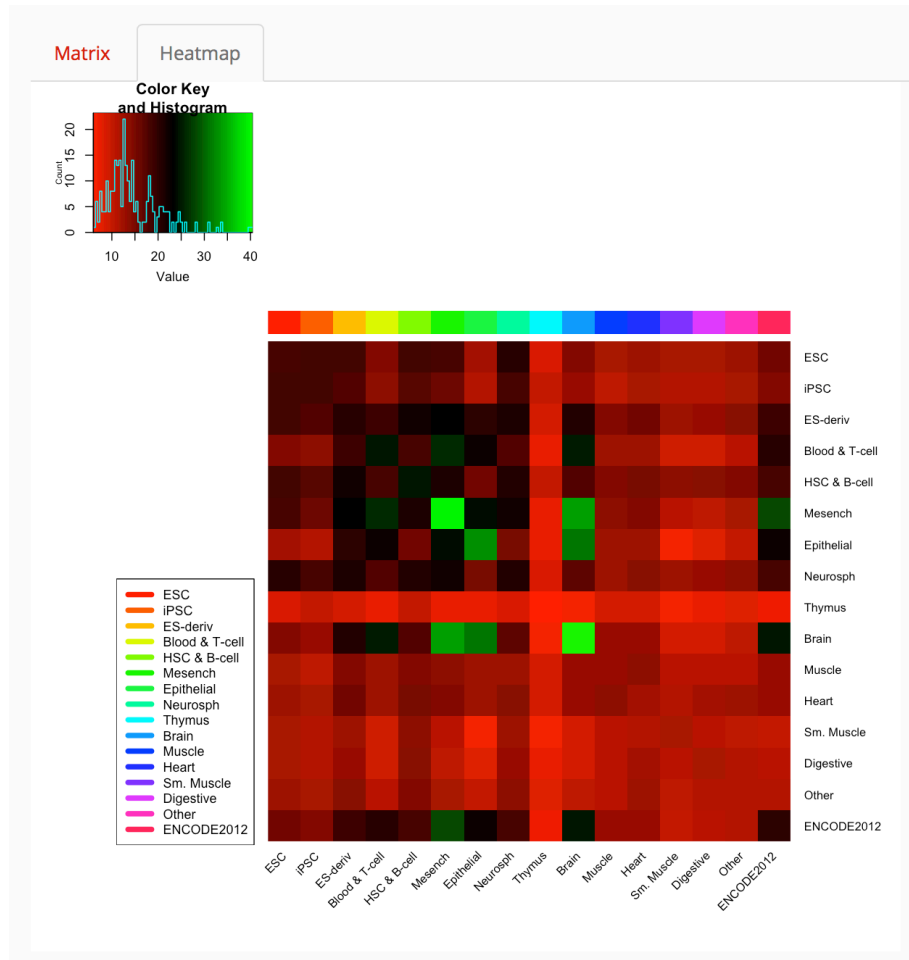
The alignment results will be shown in the form of a 16*16 matrix. We note that the three tissue types IMR90, Myosat, and Adipose each only has one sample in the Roadmap dataset, so we exclude these three tissue types from the application. The sample sizes of different tissue/cell types in Roadmap dataset are summarized in the table below.

Tissue/Cell Type	Samples Count
Adipose	1
Blood & T-cell	14
Brain	10
Digestive	12
ENCODE2012	16
Epithelial	8
ES-deriv	9
ESC	8
Heart	5
HSC & B-cell	9
IMR90	1
iPSC	5
Mesench	4
Muscle	5
Myosat	1
Neurosph	2
Other	11
Sm. Muscle	4
Thymus	2

Each entry in the alignment result matrix records the average alignment score of the chromatin state sequences of the given gene in two tissue types, corresponding to the column and row names of the matrix. The EpiAlign algorithm predicts that the tissue types which have higher alignment score are more similar in terms of the epigenomic signals on that gene region. For example, in the Roadmap dataset, there are 10 Brain genome samples. The average of 45 (10 choose 2) alignment scores between gene ENSG00000164828's chromatin state sequences within the Brain epigenome samples is 40.40, which is the highest in the matrix. Thus, EpiAlign predicts Brain cells have more similar chromatin state pattern on gene ENSG00000164828 region in comparison to other tissues. Since gene ENSG00000164828 is a Brain-associated gene. The alignment results indicate that the EpiAlign algorithm can extract chromatin patterns that carry cell-type-associated.

Press the “**Download**” button to store the alignment results in a csv file on the local machine.

Step 3:

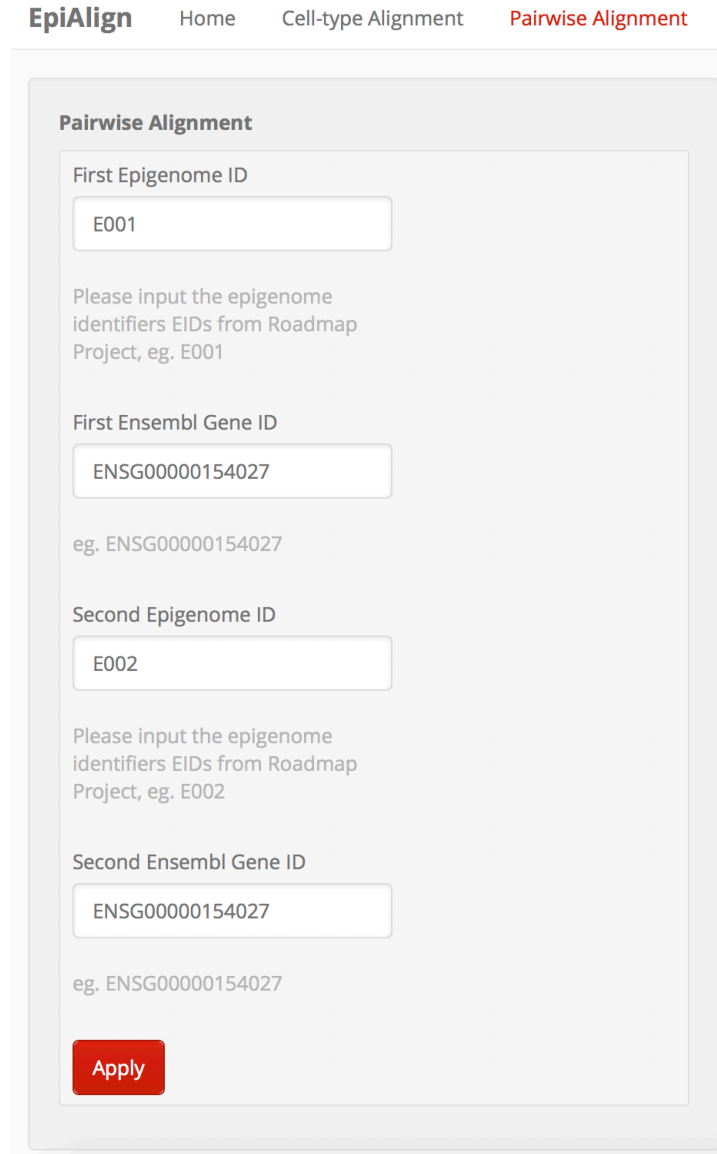


The alignment results are also provided in the form of a heatmap. The higher the average alignment score between two tissue types, the brighter the green color will be. In this example using gene ENSG00000164828, the comparison of Brain vs. Brain has the brightest green color.

2. Pairwise alignment scores

The second feature provided by this EpiAlign Web portal is to compare the chromatin state similarity between two (same or different) genomic regions on two (same or different) epigenomes.

Step 1:



The screenshot shows the EpiAlign web portal's "Pairwise Alignment" interface. The navigation bar at the top includes "EpiAlign", "Home", "Cell-type Alignment", and "Pairwise Alignment" (highlighted in red). The main content area is titled "Pairwise Alignment" and contains the following fields and instructions:

- First Epigenome ID:** A text input field containing "E001". Below it, the instruction reads: "Please input the epigenome identifiers EIDs from Roadmap Project, eg. E001".
- First Ensembl Gene ID:** A text input field containing "ENSG00000154027". Below it, the instruction reads: "eg. ENSG00000154027".
- Second Epigenome ID:** A text input field containing "E002". Below it, the instruction reads: "Please input the epigenome identifiers EIDs from Roadmap Project, eg. E002".
- Second Ensembl Gene ID:** A text input field containing "ENSG00000154027". Below it, the instruction reads: "eg. ENSG00000154027".

At the bottom of the form is a red "Apply" button.

Input the Ensembl gene ids and epigenome identifiers ids of the two regions to compare, and press the "Apply" button.

Step 2:

1	<i>seq1</i> : 12 – 529,	<i>seq2</i> : 8 – 1390
2	<i>Score</i> = 22.000000	
3	<i>m, m</i>	1, 4
4	<i>n, n</i>	12, 11
5	<i>o, o</i>	145, 178
6	<i>g, g</i>	2, 4
7	<i>o, o</i>	29, 836
8	<i>g, g</i>	6, 2
9	<i>o, o</i>	237, 217
10	<i>g, g</i>	4, 2
11	<i>o, o</i>	37, 111
12	<i>g, g</i>	2, 1
13	<i>e, e</i>	43, 17

The alignment results are displayed in a table format. In this example, the alignment regions giving the highest alignment score (22) is between the 12th to the 529th chromatin states in the first genomic region with the 8th to the 1390th chromatin state in the second genomic region. The detailed alignment results are shown from the row 3 to the row 13. The second column shows the matched chromatin states. The third column shows the number of chromatin states being matched together. For example, the row 3 means that 1 “m” state in first sequence aligns with 4 “m” states in second sequence.

3. Contact Info

If you have any questions, feel free to email us via the link provided in the contact info section.

EpiAlign

[Home](#)

[Cell-type Alignment](#)

[Pairwise Alignment](#)

[More ▾](#)

[Contact Info](#)

In case you have any questions, [feel free to contact us](#) .