Website construction  
This Website（http://www.rnanut.net/lncompare/） is built on a Linux server using the Apache+MySQL+PHP framework; all data graphics visualizations on the site are done using the open source G2 package(https://antv.alipay.com/), and the display and download of the table data is done using a JavaScript plugin called vis（http://visjs.org/）.

Similarity analysis  
In order to measure the similarity between two LncRNAs, after comparing the results of multiple similarity calculation methods, the optimal cosine similarity calculation method is adopted.

In the formula, A and B are two LncRNA feature vectors,which has deleted the NaN values. The value of the cosine similarity is the quotient of the point multiplication of the two vectors of A and B and the product of the Second paradigm of two vectors.

In order to further describe the similarity between two lncrnas, based on the cosine similarity, a mutual rank similarity calculation is performed.

In the formula, denotes the serial number of B in the descending order of Cosine similarity between all lncrnas and A, and is the same.

Enrichment analysis：

Use the hypergeometric distribution algorithm to characterize whether the lncrnas inputed have obvious enrichment on a certain feature (quantity , classification). For example: Whether the lncRNAs inputed are related to disease.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Input | Background | Total |
| Related | a | b | a + b |
| Irrelevant | c | d | c + d |
| Total | a + c | b + d | (a+b+c+d)=n |

The number of the inputed lncRNAs that associated with the diseases is a, and that unrelated to the diseases is c; The number of the lncRNAs that included in background and associated with the diseases is b, and that included in background and unrelated to the diseases is d; The total number of lncRNAs is n. The p value calculated by the formula determines whether the input lncrnas are significantly enriched in a certain feature (p < 0.005).