Supplementary Document SD5: Demonstration of comparing strings using edit distances

In order to illustrate the concept of how strings can be compared using edit distances, consider the following two strings

String 1: Breast Cancer
String 2: Brain Cancer

Let us now suppose, we can make the following operations: deletions, substitution, and insertion to transform string 1 to string 2. Then we can proceed to transform String 1 to String 2 in the following ways listed in the table T1 below:

Table T1: step by step demonstration of comparing strings using edit distances.

Method 1	Method 1 operation	Method 2	Method 2 operation
Breast Cancer		Breast Cancer	
Brast Cancer	Delete ‘e’	Braast Cancer	Substitute ‘e’ with ‘a’
Bras Cancer	Delete ‘t’	Braist Cancer	Substitute ‘a’ with ‘i’
Bra Cancer	Delete ‘s’	Braint Cancer	Substitute ‘s’ with ‘n’
Brai Cancer	Insert ‘i’	Brain Cancer	Delete ‘t’
Brain Cancer	Insert ‘n’		

We can observe from the above table T1, that the two methods transform string 1 to string 2, but method 2 employs fewer steps to achieve this objective. Furthermore, with the given set of operations the minimum number of steps required to transform string 1 to string 2 is four. This is achieved by method 2, thus the edit distance between string 1 and string 2 is four. The set of operations that were used to transform the strings and method with which we calculated the edit distance is also known as the Levenshtein distance.