Question 3:

Doppelganger effects are not unique to solely biomedical data. It can be used in many other fields that require references to similar cases of study. It can be used in cases where the current target is compared to a similar data set from an older trend as reference. As long as there are similar markers present within the data or that there are similar references or conditions.

With the usage of machine learning, after a data has been trained, it would be good practice to double check on the data produced. Although the data would show that the two replicates may be similar, it only identifies the more prominent and similar markers. Checking the percentage of similarity based on the two models will help identify if the replicates are similar or that the trained data only assessed based on the features.

A way to check for doppelganger effects is by running the data set through different training data sets or using multiple data programmes. By using multiple programmes, it may aid in identifying the outliers amongst the different sets and help show whether there are doppelgangers present within them.

Markers within the data will be more prominent if the same values are occurring amongst the data sets. However, by this method of using multiple programmes to check, it will increase time taken to run the training data.