**Figure 2 Data:**

* **#1: barcodedNonBarcoded\_AUPRC\_AUROC\_TNR.csv**
* **#2: TNR\_plotted\_formatted.csv**

**Dataset Description:**

The dataset includes results from various doublet detection algorithms applied to multiple scRNA-seq datasets with different doublet content and algorithm parameters. While scRNA-seq aims to measure the transcriptomes of individual cells, doublets can occur when two cells are captured as one. The purpose of doublet detection algorithms is to accurately identify these doublets so they can be removed for downstream analysis. The four algorithms evaluated are DoubletFinder, hybrid, scDblFinder, and Scrublet. Key performance metrics for these algorithms include the area under the precision-recall curve (AUPRC), the area under the receiver operating characteristic curve (AUROC), and the true negative rate (TNR).

**File Description #1:**

The CSV file contains a dataset with each row representing a result and each column representing a feature. The columns in the dataset are as follows:

* "": Index number
* "X": Index number
* "dataset": The scRNA-seq dataset from which the data originated
* "sample": The specific sample within the corresponding scRNA-seq dataset from which the data originated
* "condition": The doublet detection algorithm used to generate the AUPRC, AUROC, and TNR data (options include DoubletFinder, hybrid, scDblFinder, or Scrublet)
* "auprc": Area under the precision-recall curve
* "auroc": Area under the receiver operating characteristic curve
* "dbl\_act": Actual true doublet rate of the dataset
* "isBarcoded": Indicates whether barcoding technology was used in the dataset
* "TNR": True negative rate

**File Description #2:**

The CSV file contains a dataset with each row representing a result and each column representing a feature. The columns in the dataset are as follows:

* "": Index number
* "X": Index number
* "condition": The doublet detection algorithm used to generate the AUPRC, AUROC, and TNR data (options include DoubletFinder, hybrid, scDblFinder, or Scrublet)
* "dataset": The scRNA-seq dataset from which the data originated
* "sample": The specific sample within the corresponding scRNA-seq dataset
* "TNR": True negative rate
* “dbl\_exp”: Expected true doublet rate from 10X Genomics
* "dbl\_act": Actual true doublet rate from 10X Genomics

**Research Goal:**

The research goal is to evaluate and compare the performance of four doublet detection algorithms across various scRNA-seq datasets, as indicated in the "dataset" column, using three metrics: AUPRC, AUROC, and TNR.