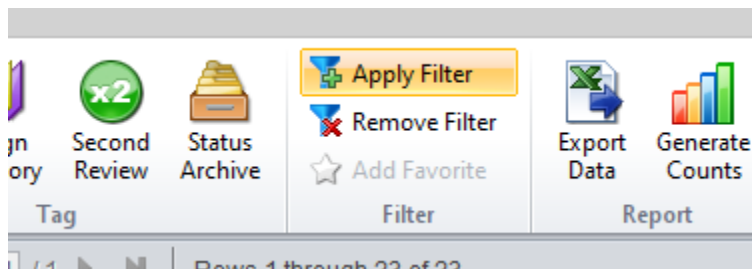


How to Create a Match Status Archive for 2024 Tumor Deduplication
Using the 2023 Tumor Deduplication Match*Pro Results File

1. Open the Match*Pro results file containing the 2023 tumor deduplication data.
2. You should see pairs that look like this (if you don't, you have the wrong file):

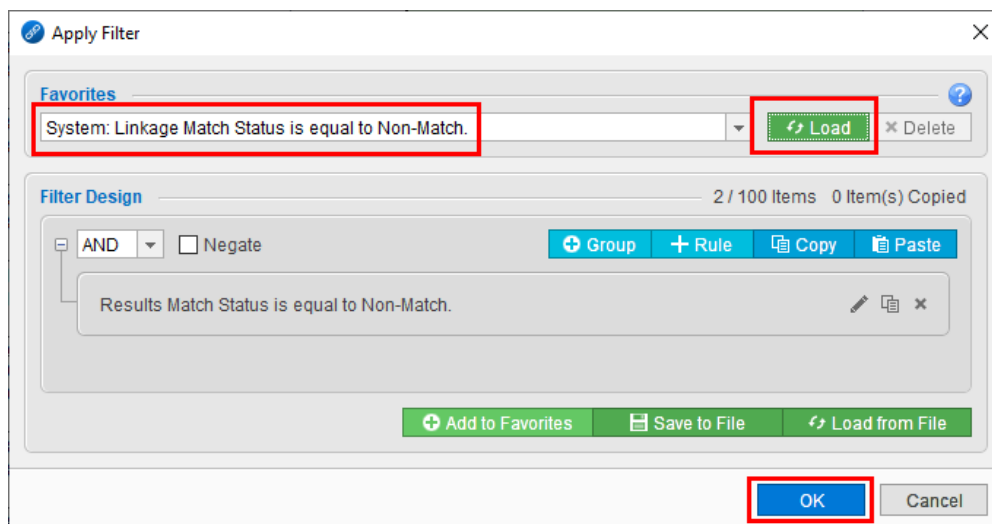
```
02,02,C170,0,,,8936,3,2021-05-03,046,3  
01,01,C171,0,,,8936,3,2021-05-03,046,3  
19.6 (1.00)
```

3. Press the **Apply Filter** button. A dialog will appear.

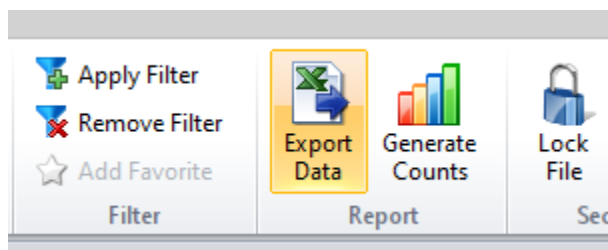


How to Create a Match Status Archive for 2024 Tumor Deduplication
Using the 2023 Tumor Deduplication Match*Pro Results File

4. Select “**System: Linkage Match Status is equal to Non-Match**” from the dropdown, press the **Load** button, then press the **OK** button.



5. The dialog will close and you'll be returned to the manual review screen. Only the non-matches should be showing at this point. **If there aren't any non-matches (i.e., the results screen is blank) you can stop at this point. You don't have any pairs from last year that need to be archived.** If the results screen is not blank, continue.
6. Press the **Export Data** button. A dialog will appear.



How to Create a Match Status Archive for 2024 Tumor Deduplication Using the 2023 Tumor Deduplication Match*Pro Results File

7. Press the **Browse** button and provide a location where the CSV file should be written.
8. Press the **Load** button. When the file dialog appears, select the export selections file that came with these instructions ([match-status-archive-export-selections.mple](#)), then press the **Open** button.
9. Press the **OK** button. This will create a csv file containing the following fields:
 - a. Match status
 - b. Line number 1
 - c. Line number 2
 - d. patientIdNumber [20] (File 1)
 - e. primarySite [400] (File 1)
 - f. laterality [410] (File 1)
 - g. histologicTypeIcdO3 [522] (File 1)
 - h. behaviorCodeIcdO3 [523] (File 1)
 - i. dateOfDiagnosis [390] (File 1)
 - j. patientIdNumber [20] (File 2)
 - k. primarySite [400] (File 2)
 - l. laterality [410] (File 2)
 - m. histologicTypeIcdO3 [522] (File 2)
 - n. behaviorCodeIcdO3 [523] (File 2)
 - o. dateOfDiagnosis [390] (File 2)

Export Data

Export Linked Data to CSV | **Export Records from File 1 or File 2**

Specify the CSV file or SAS code location(s).

CSV File Destination ... **7**

SAS Code Destination ... **x**

Select the fields to export using the checkboxes below. **8**

Search... **x** **Save** **Load**

	Field Name	Data Source
<input type="checkbox"/>	Annotation	Results
<input type="checkbox"/>	Annotation (2nd Review)	Results
<input type="checkbox"/>	Categories	Results
<input type="checkbox"/>	File 1 ID	Results
<input type="checkbox"/>	File 1 ID Mapping	Results
<input type="checkbox"/>	File 2 ID	Results
<input type="checkbox"/>	File 2 ID Mapping	Results
<input type="checkbox"/>	Line Number 1 Mapping	Results
<input type="checkbox"/>	Line Number 2 Mapping	Results
<input type="checkbox"/>	Match Status (2nd Review)	Results
<input type="checkbox"/>	Overall Similarity	Results
<input type="checkbox"/>	Recoded File 1 ID	Results
<input type="checkbox"/>	Recoded File 2 ID	Results
<input type="checkbox"/>	Set 01 File 1 Data	Results
<input type="checkbox"/>	Set 01 File 2 Data	Results
<input type="checkbox"/>	Set 01 Score	Results
<input type="checkbox"/>	Set 01 Similarity	Results

0 selected **Toggle Results Fields** **Toggle File 1 Fields** **Toggle File 2 Fields**

Which set of records should be written to the CSV file? **Records that met my filter criteria**

Task is Awaiting Execution. **OK** **Cancel** **9**

How to Create a Match Status Archive for 2024 Tumor Deduplication
Using the 2023 Tumor Deduplication Match*Pro Results File

10. I strongly suggest making a backup copy of the csv at this point, as you will now need to make several changes to it. **If you choose to use MS Excel for this work, make sure that you DO NOT let it do any sort of data conversion**, or you may lose leading zeros on the patient IDs, which will invalidate the archive.
 - a. Delete the match status column and both of the line number columns. These are the first 3 columns in the file.
 - b. Re-arrange the remaining columns so they are in the following order (you can right-click on a column header in Excel and use cut/paste to do this easily):
 - i. File 1 patientIdNumber [20]
 - ii. File 1 primarySite [400]
 - iii. File 1 laterality [410]
 - iv. File 1 histologicTypeIcdO3 [522]
 - v. File 1 behaviorCodeIcdO3 [523]
 - vi. File 1 dateOfDiagnosis [390]
 - vii. File 2 patientIdNumber [20]
 - viii. File 2 primarySite [400]
 - ix. File 2 laterality [410]
 - x. File 2 histologicTypeIcdO3 [522]
 - xi. File 2 behaviorCodeIcdO3 [523]
 - xii. File 2 dateOfDiagnosis [390]
 - c. Save the file.
 - d. Close MS Excel.

How to Create a Match Status Archive for 2024 Tumor Deduplication
Using the 2023 Tumor Deduplication Match*Pro Results File

11. Open the CSV file again, but this time use **Notepad** or some other type of text editor. You can do this by right-clicking on the file and selecting “Open With”, followed by “Notepad”. When you open the file in Notepad, it should look something like this. If it doesn’t, then you did something wrong on the previous step.

```
File Edit Format View Help
File 1 patientIdNumber [20],File 1 primarySite [400],File 1 laterality [410],
49290500,C446,1,8742,2,20200831,49290500,C446,1,8742,3,20200831
17441010,C672,0,8130,2,20200331,17441010,C672,0,8130,3,20100116
89111020,C508,2,8522,3,20100103,89111020,C505,2,8522,2,20100810
93611020,C504,1,8522,3,20100911,93611020,C504,1,8500,2,20101229
03941025,C503,2,8022,3,20211130,03941025,C509,2,8520,3,20110316
56451025,C050,2,8070,3,20200225,56451025,C050,0,8051,3,20100928
76611026,C006,0,8070,2,20100308,76611026,C001,0,8070,3,20101130
62081035,C508,1,8520,2,20200902,62081035,C502,1,8500,3,20100604
98141036,C505,2,8500,3,20100405,98141036,C504,2,8520,2,20100613
27441038,C508,1,8500,3,20100919,27441038,C504,1,8520,2,20101016
98171038,C505,1,8500,3,20100215,98171038,C508,1,8520,2,20100326
32651040,C341,1,8046,3,20200720,32651040,C342,1,8046,3,20101126
27901041,C503,1,8500,3,20200313,27901041,C508,1,8520,2,20100923
46131041,C503,1,8522,3,20200313,46131041,C504,1,8500,2,20101001
16181043,C505,1,8522,3,20211215,16181043,C504,1,8520,2,20111115
11051050,C504,1,8500,3,20200115,11051050,C504,1,8522,2,20200115
79811051,C504,1,8500,3,20211014,79811051,C503,1,8520,2,20210115
18701053,C341,1,8140,3,20200301,18701053,C343,1,8140,3,20200204
52581053,C508,1,8522,3,20200405,52581053,C508,1,8500,2,20200315
93251053,C669,1,8130,3,20200506,93251053,C678,0,8130,3,20200701
61471054,C502,2,8500,3,20200203,61471054,C502,2,8500,3,20201122
51101055,C509,1,8520,3,20211214,51101055,C504,1,8500,2,20210526
80231059,C170,0,8936,3,20200503,80231059,C171,0,8936,3,20200503
```


How to Create a Match Status Archive for 2024 Tumor Deduplication
Using the 2023 Tumor Deduplication Match*Pro Results File

- a. Replace the header row with the following (without the quotes): "id1,cancer1,id2,cancer2"
- b. The file should now look like this (the change has been highlighted):

```
File Edit Format View Help
id1,cancer1,id2,cancer2
49290500,C446,1,8742,2,20200831,49290500,C446,1,8742,3,20200831
17441010,C672,0,8130,2,20200331,17441010,C672,0,8130,3,20100116
89111020,C508,2,8522,3,20100103,89111020,C505,2,8522,2,20100810
93611020,C504,1,8522,3,20100911,93611020,C504,1,8500,2,20101229
03941025,C503,2,8022,3,20211130,03941025,C509,2,8520,3,20110316
56451025,C050,2,8070,3,20200225,56451025,C050,0,8051,3,20100928
76611026,C006,0,8070,2,20100308,76611026,C001,0,8070,3,20101130
62081035,C508,1,8520,2,20200902,62081035,C502,1,8500,3,20100604
98141036,C505,2,8500,3,20100405,98141036,C504,2,8520,2,20100613
27441038,C508,1,8500,3,20100919,27441038,C504,1,8520,2,20101016
98171038,C505,1,8500,3,20100215,98171038,C508,1,8520,2,20100326
32651040,C341,1,8046,3,20200720,32651040,C342,1,8046,3,20101126
27901041,C503,1,8500,3,20200313,27901041,C508,1,8520,2,20100923
46131041,C503,1,8522,3,20200313,46131041,C504,1,8500,2,20101001
16181043,C505,1,8522,3,20211215,16181043,C504,1,8520,2,20111115
11051050,C504,1,8500,3,20200115,11051050,C504,1,8522,2,20200115
79811051,C504,1,8500,3,20211014,79811051,C503,1,8520,2,20210115
18701053,C341,1,8140,3,20200301,18701053,C343,1,8140,3,20200204
52581053,C508,1,8522,3,20200405,52581053,C508,1,8500,2,20200315
93251053,C669,1,8130,3,20200506,93251053,C678,0,8130,3,20200701
61471054,C502,2,8500,3,20200203,61471054,C502,2,8500,3,20201122
51101055,C509,1,8520,3,20211214,51101055,C504,1,8500,2,20210526
80231059,C170,0,8936,3,20200503,80231059,C171,0,8936,3,20200503
```

How to Create a Match Status Archive for 2024 Tumor Deduplication
Using the 2023 Tumor Deduplication Match*Pro Results File

- c. Insert a double-quotation mark immediately before the values in each of the primary site columns AND immediately after the values in each of the date of diagnosis columns. Be very mindful of the locations of the commas relative to where the quotation marks are placed. The file should now look something like this (the changes have been highlighted):

```
File Edit Format View Help
id1,cancer1,id2,cancer2
49290500,"C446,1,8742,2,20200831",49290500,"C446,1,8742,3,20200831"
17441010,"C672,0,8130,2,20200331",17441010,"C672,0,8130,3,20100116"
89111020,"C508,2,8522,3,20100103",89111020,"C505,2,8522,2,20100810"
93611020,"C504,1,8522,3,20100911",93611020,"C504,1,8500,2,20101229"
03941025,"C503,2,8022,3,20211130",03941025,"C509,2,8520,3,20110316"
56451025,"C050,2,8070,3,20200225",56451025,"C050,0,8051,3,20100928"
76611026,"C006,0,8070,2,20100308",76611026,"C001,0,8070,3,20101130"
62081035,"C508,1,8520,2,20200902",62081035,"C502,1,8500,3,20100604"
98141036,"C505,2,8500,3,20100405",98141036,"C504,2,8520,2,20100613"
27441038,"C508,1,8500,3,20100919",27441038,"C504,1,8520,2,20101016"
98171038,"C505,1,8500,3,20100215",98171038,"C508,1,8520,2,20100326"
32651040,"C341,1,8046,3,20200720",32651040,"C342,1,8046,3,20101126"
27901041,"C503,1,8500,3,20200313",27901041,"C508,1,8520,2,20100923"
46131041,"C503,1,8522,3,20200313",46131041,"C504,1,8500,2,20101001"
16181043,"C505,1,8522,3,20211215",16181043,"C504,1,8520,2,20111115"
11051050,"C504,1,8500,3,20200115",11051050,"C504,1,8522,2,20200115"
79811051,"C504,1,8500,3,20211014",79811051,"C503,1,8520,2,20210115"
18701053,"C341,1,8140,3,20200301",18701053,"C343,1,8140,3,20200204"
52581053,"C508,1,8522,3,20200405",52581053,"C508,1,8500,2,20200315"
93251053,"C669,1,8130,3,20200506",93251053,"C678,0,8130,3,20200701"
61471054,"C502,2,8500,3,20200203",61471054,"C502,2,8500,3,20201122"
51101055,"C509,1,8520,3,20211214",51101055,"C504,1,8500,2,20210526"
80231059,"C170,0,8936,3,20200503",80231059,"C171,0,8936,3,20200503"
```

- d. Save the file.
- e. Close Notepad.

How to Create a Match Status Archive for 2024 Tumor Deduplication
Using the 2023 Tumor Deduplication Match*Pro Results File

12. Open Match*Pro and select the **“Match Status Archive Manager”** from the Tools menu in the upper-left corner or press the F4 key.
 - a. Select **“Delimited”** from the File Type dropdown.
 - b. Press the **Browse** button to the right of where it says: **“File Path”**.
When the file dialog appears, select the csv file you edited in step 11, then press the **Open** button.
 - c. A preview dialog will appear. Just press the **OK** button.
 - d. When you return to the Match Status Archive Manager screen, make the following selections:
 - i. Entity ID 1: **id1**
 - ii. Record ID 1: **cancer1**
 - iii. Entity ID 2: **id2**
 - iv. Record ID 2: **cancer2**
 - e. Select **“Non-Matched ID Pairs”** from the dropdown labeled **“Contents”**.
 - f. Type something in the notes field (e.g., “NAACCR tumor deduplication”).
 - g. Towards the bottom of the screen, beneath where it says: **“Provide the Output Location”**, press the **Browse** button to the right of where it says: **“File Path”**. A save dialog will appear. Specify where the archive should be written, then press the **Save** button.
 - h. When you return to the Match Status Archive Manager Dialog, press the **Create** button. The archive will be created in the location you specified.

Status Archive Manager

Specify the Management Task to Perform

Mgmt Task: Create a status archive using data from an external file

External File Setup

File Type: Delimited

File Path: C:\Users\howew\Desktop\dedup\archive-export.csv

Layout: Comma-delimited, Double Quote, File has headers

Entity ID 1	id1
Record ID 1	cancer1
Entity ID 2	id2
Record ID 2	cancer2

Contents: Non-Matched ID Pairs

Describe the Data

Notes: NAACCR tumor deduplication

Provide the Output Location

File Path: C:\Users\howew\matchpro\archive\my-archive-file.mpsa

Task is Awaiting Execution. **Create** **Cancel**

How to Create a Match Status Archive for 2024 Tumor Deduplication
Using the 2023 Tumor Deduplication Match*Pro Results File

13. Close the Match Status Archive Manager Dialog.
14. You can use the archive you created in step 12 of the Tumor Deduplication process.