# Appendix II (C) 1. Data Preparation Tutorial for Mondrian Map

Here we present the resulting data tables from each step, following the data preparation flow diagram.

## Step 1. Temporal Gene Expression Profile Selection

| Gene_symbol | baseline_R1 | baseline_R2 | baseline_TP | aggressive_R1 | aggressive_R2 | aggressive_TP | nonaggressive_R1 | nonaggressive_R2 | nonaggressive_TP |
|---|---|---|---|---|---|---|---|---|---|
| CMSS1 | 31.207177 | 35.164799 | 24.341754 | 26.446038 | 27.000944 | 32.025694 | 31.395261 | 48.859002 | 44.117700 |
| STX16 | 35.996421 | 44.218382 | 20.716331 | 99.860372 | 104.422161 | 126.156318 | 20.050733 | 21.480300 | 45.597264 |
| TLE4 | 20.813337 | 22.299578 | 10.835135 | 23.209358 | 19.561900 | 31.615145 | 72.895632 | 96.452925 | 65.074070 |
| LRRN1 | 31.667442 | 16.577943 | 53.999371 | 2.167500 | 0.433401 | 5.336610 | 121.089577 | 101.425790 | 137.481300 |
| YKT6 | 92.464768 | 82.431033 | 75.375286 | 52.902599 | 65.192139 | 59.786585 | 11.058250 | 13.540128 | 9.911100 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| UQCRFS1 | 82.462700 | 62.024200 | 65.626250 | 54.498700 | 58.399200 | 50.797400 | 68.277000 | 105.298000 | 60.563400 |
| C4orf32 | 3.717895 | 4.994350 | 2.917710 | 0.268489 | 0.327187 | 0.151905 | 4.348520 | 4.666130 | 4.126970 |
| MTRNR2L9 | 36.739885 | 82.480730 | 47.055200 | 0.430776 | 0.616574 | 0.785486 | 111.565000 | 63.872200 | 98.041800 |
| BMP2 | 5.453945 | 6.221065 | 27.504200 | 0.685032 | 0.462191 | 0.778202 | 36.265900 | 70.232900 | 40.471300 |
| EP300 | 12.412165 | 13.973550 | 12.963900 | 43.230000 | 41.312600 | 56.060900 | 48.727000 | 37.960800 | 42.483100 |

11090 rows × 9 columns

## Step 2. Differentially Expressed Gene Selection

| Gene_symbol | baseline_R1/TP | baseline_R2/TP |
|---|---|---|
| CMSS1 | 1.282043 | 1.444629 |
| STX16 | 1.737587 | 2.134470 |
| TLE4 | 1.920912 | 2.058080 |
| LRRN1 | 0.586441 | 0.307003 |
| YKT6 | 1.226725 | 1.093608 |
| ... | ... | ... |
| UQCRFS1 | 1.256551 | 0.945113 |
| C4orf32 | 1.274251 | 1.711736 |
| MTRNR2L9 | 0.780783 | 1.752850 |
| BMP2 | 0.198295 | 0.226186 |
| EP300 | 0.957441 | 1.077882 |

11090 rows × 2 columns

## Step 3. Gene-set, Network, and Pathway Analysis (GNPA)

| | GS_ID | NAME | SOURCE | GS_SIZE | ORGANISM | DESCRIPTION | LINK | TYPE | MULTI_N | OLAP | COCO_V2 | SIMILARITY_SCORE | pvalue | Rank | pFDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | WAG002718 | Neutrophil degranulation | WikiPathway_2021 | 474 | Homo sapiens | Neutrophil degranulation | https://www.wikipathways.org/index.php/Pathway... | P | 1333 | 120 | 172.7292014 | .0838279321392930300961184936183623089448 | 1.732402e-44 | 1 | 5.422417e-42 |
| 1 | WAG002628 | VEGFA-VEGFR2 signaling pathway | WikiPathway_2021 | 438 | Homo sapiens | VEGFA-VEGFR2 signaling pathway | https://www.wikipathways.org/index.php/Pathway... | P | 1333 | 97 | 980.8100106 | .0712460408468820826612487529758081271878 | 3.931521e-31 | 2 | 1.226635e-28 |
| 2 | WAG003066 | Endothelin pathway | WikiPathway_2021 | 194 | Homo sapiens | Endothelin pathway | https://www.wikipathways.org/index.php/Pathway... | P | 1333 | 56 | 968.3903924 | .0744733638450988550382938067279983973618 | 1.531451e-24 | 3 | 4.762811e-22 |
| 3 | WAG002670 | TYROBP causal network in microglia | WikiPathway_2021 | 60 | Homo sapiens | TYROBP causal network in microglia | https://www.wikipathways.org/index.php/Pathway... | P | 1333 | 32 | 221.1523998 | .116643283969862301896596518576253572356 5 | 3.132260e-24 | 4 | 9.710006e-22 |
| 4 | WAG002820 | Ebola virus pathway in host | WikiPathway_2021 | 131 | Homo sapiens | Ebola virus pathway in host | https://www.wikipathways.org/index.php/Pathway... | P | 1333 | 45 | 650.3115529 | .0824218022714578290821076149692794637 87 | 1.841119e-23 | 5 | 5.689056e-21 |

## Step 4. Pathway Networks

| | GS_A_ID | GS_A_SIZE | GS_B_ID | GS_B_SIZE | OLAP | SIMILARITY | PVALUE |
|---|---|---|---|---|---|---|---|
| 0 | WAG003212 | 430 | WAG003224 | 153 | 55 | .15522875816993464 | 149.90937335538615 |
| 1 | WAG002596 | 246 | WAG002683 | 29 | 16 | .15976567700705632 | 62.03314298178884 |
| 2 | WAG002596 | 246 | WAG002118 | 55 | 31 | .20457912457912455 | 120.66116349772959 |
| 3 | WAG002596 | 246 | WAG003300 | 33 | 19 | .17361305361305365 | 74.66521990991453 |
| 4 | WAG002414 | 90 | WAG002096 | 63 | 18 | .16380952380952382 | 73.94134202903723 |

## Step 5. Differential Pathway Analysis

| | GS_ID | wFC | pFDR |
|---|---|---|---|
| 0 | WAG001976 | 1.3186 | 0.006331 |
| 1 | WAG001977 | 1.6080 | 0.075830 |
| 2 | WAG002000 | 2.0673 | 0.000030 |
| 3 | WAG002011 | 1.0627 | 0.240331 |
| 4 | WAG002018 | 1.7749 | 0.084468 |

## Step 6. Pathway Embeddings

Utilizing language models to generate pathway embeddings described in section II A and Appendix C, we get one final data table for each profile with top 10 pathways ready for Mondrian Map Visualization. The table below is for Baseline cohort R1/TP. We have 5 similar tables for 5 other profiles.

| | GS_ID | wFC | pFDR | x | y |
|---|---|---|---|---|---|
| 0 | WAG002133 | 1.6183 | 3.724584e-43 | 607.298815 | 220.621129 |
| 1 | WAG002628 | 2.0947 | 3.882609e-43 | 200.916562 | 272.423214 |
| 2 | WAG002134 | 1.6776 | 1.419164e-41 | 704.595680 | 265.673499 |
| 3 | WAG003066 | 2.5782 | 2.237119e-30 | 185.874616 | 133.851282 |
| 4 | WAG003212 | 2.2273 | 3.933392e-30 | 180.720278 | 392.922252 |
| 5 | WAG002596 | 1.7945 | 2.100707e-29 | 287.372243 | 443.144485 |
| 6 | WAG002718 | 1.6066 | 2.283309e-27 | 775.578285 | 92.995014 |
| 7 | WAG003032 | 2.3010 | 6.165519e-26 | 446.478214 | 887.448126 |
| 8 | WAG002186 | 1.7709 | 5.848443e-23 | 93.562092 | 282.262669 |
| 9 | WAG002879 | 1.5897 | 9.164786e-23 | 70.818373 | 160.476481 |