

Here is What We'll Do

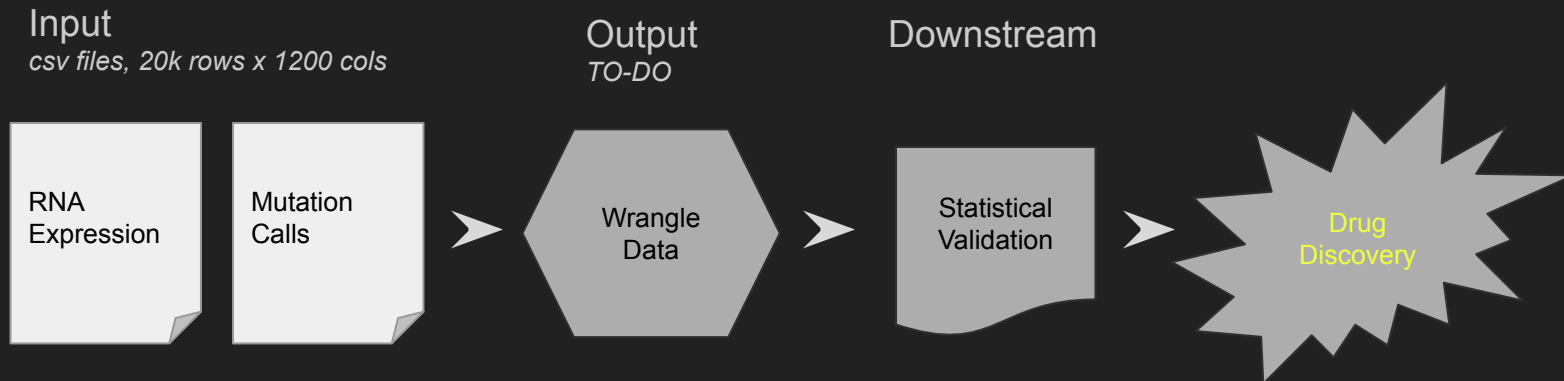
For the Programming Project in C Class

General Idea

The project encompasses: file handling and data manipulation.

The application: use in a pre-existing statistics pipeline.

The goal is: use downstream analysis to predict new drug targets in cancer



Mutation Calls Tell if a Gene is Mutated in a Cell

For Example:

RNA Expression (float)

| " | cell _a | cell _b | cell _c | cell _d |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| gene _x | 3.3 | 6.6 | 2.2 | 5.5 |
| gene _y | 8.8 | 5.5 | 4.4 | 1.1 |
| gene _z | 7.7 | 8.8 | 9.9 | 7.7 |

Gene_y is mutated in
Cell_a & Cell_b

Mutation calls (binary, int)

| " | cell _a | cell _b | cell _c | cell _d |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| gene _x | 0 | 0 | 1 | 1 |
| gene _y | 1 | 1 | 0 | 0 |
| gene _z | 1 | 0 | 1 | 1 |

Gene_x is mutated in
Cell_c & Cell_d

Cell_b is the only cell
type that has a
working Gene_z

Think of these Calls as “Context”:
A Context Loss (0) or
A Context Gain (1)

Task: All combinations of “Context” for RNA values

Take all combinations of Contexts with all Genes, storing those 0 or 1 contexts as arrays of RNA Expression values For Example:



Task: All combinations of “Context” for RNA values

Take all combinations of Contexts with all Genes, storing those 0 or 1 contexts as arrays of RNA Expression values For Example:

Mutation calls (binary, int)

| " | cell _a | cell _b | cell _c | cell _d |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| gene _x | 0 | 0 | 1 | 1 |
| gene _y | 1 | 1 | 0 | 0 |
| gene _z | 1 | 0 | 1 | 1 |

RNA Expression (float)

| " | cell _a | cell _b | cell _c | cell _d |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| gene _x | 3.3 | 6.6 | 2.2 | 5.0 |
| gene _y | 8.4 | 5.5 | 4.4 | 1.1 |
| gene _z | 7.7 | 8.8 | 9.9 | 7.0 |

gene_x-gene_x [3.3 , 6.6] [2.2, 5.0]
gene_x-gene_y [8.4 , 5.5] [4.4, 1.1]
gene_x-gene_z [7.7 , 8.8] [9.9, 7.0]
gene_y-gene_x [2.2 , 5.0] [3.3, 6.6]
gene_y-gene_y [4.4 , 1.1] [8.4, 5.5]
gene_y-gene_z [9.9 , 7.0] [7.7, 8.8]

0

1

Writing to a CSV file

Task: All combinations of “Context” for RNA values

Take all combinations of Contexts with all Genes, storing those 0 or 1 contexts as arrays of RNA Expression values For Example:

Mutation calls (binary, int)

| " | cell _a | cell _b | cell _c | cell _d |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| gene _x | 0 | 0 | 1 | 1 |
| gene _y | 1 | 1 | 0 | 0 |
| gene _z | 1 | 0 | 1 | 1 |

RNA Expression (float)

| " | cell _a | cell _b | cell _c | cell _d |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| gene _x | 3.3 | 6.6 | 2.2 | 5.0 |
| gene _y | 8.4 | 5.5 | 4.4 | 1.1 |
| gene _z | 7.7 | 8.8 | 9.9 | 7.0 |

gene_x-gene_x [3.3 , 6.6] [2.2, 5.0]
gene_x-gene_y [8.4 , 5.5] [4.4, 1.1]
gene_x-gene_z [7.7 , 8.8] [9.9, 7.0]
gene_y-gene_x [2.2 , 5.0] [3.3, 6.6]
gene_y-gene_y [4.4 , 1.1] [8.4, 5.5]
gene_y-gene_z [9.9 , 7.0] [7.7, 8.8]
gene_z-gene_x [6.6] [3.3, 2.2, 5.0]
gene_z-gene_y [5.5] [8.4, 4.4, 1.1]
gene_z-gene_z [8.5] [7.7, 9.9, 7.0]

0

1

Writing to a CSV file

All combinations need a test performed

As these combinations are created, a Statistical test must be performed between these two arrays to understand their differences, and put into a second csv file.

This test could be parametric or nonparametric, but should be saved as a different file from the original arrays:

Output CSV #1:
Gene-gene array pairs

| | |
|--|---|
| gene _x -gene _x , [3.3 , 6.6], [2.2, 5.0] | → stat_func ([3.3 , 6.6], [2.2, 5.0]) |
| gene _x -gene _y , [8.4 , 5.5], [4.4, 1.1] | |
| gene _x -gene _z , [7.7 , 8.8], [9.9, 7.0] | |
| gene _y -gene _x , [2.2 , 5.0], [3.3, 6.6] | . |
| gene _y -gene _y , [4.4 , 1.1], [8.4, 5.5] | . |
| gene _y -gene _z , [9.9 , 7.0], [7.7, 8.8] | . |
| gene _z -gene _x , [6.6], [3.3, 2.2, 5.0] | |
| gene _z -gene _y , [5.5], [8.4, 4.4, 1.1] | |
| gene _z -gene _z , [8.5], [7.7, 9.9, 7.0] | → stat_func ([8.5], [7.7, 9.9, 7.0]) |

Output CSV #2:
Gene-gene parametric or nonparametric

| |
|--|
| gene _x -gene _x , <similarity-score _{xx} >, <p-value _{xx} > |
| gene _x -gene _y , <similarity-score _{xy} >, <p-value _{xy} > |
| gene _x -gene _z , <similarity-score _{xz} >, <p-value _{xz} > |
| gene _y -gene _x , <similarity-score _{yx} >, <p-value _{yx} > |
| gene _y -gene _y , <similarity-score _{yy} >, <p-value _{yy} > |
| gene _y -gene _z , <similarity-score _{yz} >, <p-value _{yz} > |
| gene _z -gene _x , <similarity-score _{zx} >, <p-value _{zx} > |
| gene _z -gene _y , <similarity-score _{zy} >, <p-value _{zy} > |
| gene _z -gene _z , <similarity-score _{zz} >, <p-value _{zz} > |

All combinations need a test performed

So the task list is:

- 1) Use mutation calls to separate data into loss (1) or gain (0)
- 2) Apply this binary mask to RNA expression in all combinations
- 3) Save file of results line by line
- 4) Send arrays to statistical subprocess line by line

Why are we working on this?:

- 1) Finding new drug targets may lead to more patient lives saved
- 2) If we find any new targets, you'll be mentioned as collaborators on our next publication