Solutions

Datasets were generated using "CodeToGenerateDatasets.R" file.

Calculations performed using NumPy and Pandas.

Answer 1

The chemicals which are structurally similar are: -

i) Chemical C1 and C2 are similar
ii) Chemical C3 and C5 are similar

The method using which I came to this conclusion is, firstly I sorted the data in a fashion that is much easier to read and through which we can come to a conclusion.

The averages of all the 1000 genes of the transcriptome matrix of T0, which is the untreated cell matrix, is computed for all the datasets i.e. A, B, C and D which resulted into 4 values and the average of these 4 values are computed which gives the overall average about the properties of the untreated cell, i.e. the cell before being treated with the chemicals C1, C2, C3, C4 and C5.

Later, this same method is applied for all the transcriptome matrices which is resulted by the treatment of the test cell with all the chemicals and averages are computed.

Average of all the properties of the untreated cell as computed: - 558.1617183

Average of all the properties after the cell is treated with chemical C1: - 705.3640584

Average of all the properties after the cell is treated with chemical C2: - 712.8778379

Average of all the properties after the cell is treated with chemical C3: - 899.2978054

Average of all the properties after the cell is treated with chemical C4: - 560.8161931

Average of all the properties after the cell is treated with chemical C5: -908.2486711

Now the difference is calculated of the untreated cell average from the average after treatment with the cell, and the results are: -

Difference in the average after the cell with compound C1 and the average of the untreated cell: -

147.2023401

Difference in the average after the cell with compound C2 and the average of the untreated cell: -

154.7161197

Difference in the average after the cell with compound C3 and the average of the untreated cell: -

341.1359774

Difference in the average after the cell with compound C4 and the average of the untreated cell: -

2.654474865

Difference in the average after the cell with compound C5 and the average of the untreated cell: -

350.0869529

We see that the difference the compound C1 and C2 bring in the change of properties of the cell is almost similar and the same can be said with the compounds C3 and C5, that they also bring almost the similar changes in the cell as the difference calculated are almost similar.

Henceforth, we can say that the compounds C1, C2 and compounds C3, C5 are structurally similar as they brings almost the same amount of changes in the test cell.

Answer 2

As stated, the genes which were similar were C1, C2 and the other pair was C3, C5.

So, the top 10 genes in each condition which justifies the structural similarity and dissimilarity were calculated as: -

Firstly, the average of the difference of the properties of all the genes in the cell before and after reaction was calculated.

For justifying structural similarity and dissimilarity for C1 and C2, averages value of transcriptome matrix for C1 and C2 was taken. Then this average was compared with all the genes in the transcriptome matrix of both C1 and C2. The genes' average which deviated a lot from the average of C1 and C2, were responsible for the structural dissimilarity of the two chemicals, and the genes whose average value were very close to that of the average of the transcriptome of C1 and C2 were responsible for the structural similarity.

Similar logic was applied to find top 10 for C3 and C5 also.

Here are the top 10 genes in each condition: -

Supporting structural similarity between C1 and C2 (ordered rank wise): -

In C1

Gene 707

Gene 402

Gene 721

Gene 47

Gene 796

Gene 881

Gene 519

Gene 157

Gene 237

Gene 386

In C2

Gene 387

Gene 192

Gene 167

Gene 260

Gene 807

Gene 823

Gene 875

Gene 587

Gene 782

Gene 575

Supporting structural dissimilarity between C1 and C2 (ordered rank wise): -

In C1

Gene 485

Gene 798

Gene 441

Gene 749

Gene 923

Gene 135

Gene 985

Gene 602

Gene 371

Gene 407

In C2

Gene 610

Gene 77

Gene 995

Gene 756

Gene 771

Gene 505

Gene 822

Gene 694

Gene 44

Gene 825

Supporting structural similarity between C3 and C5 (ordered rank wise): -

In C3

Gene 960

Gene 712

Gene 625

Gene 828

Gene 168

Gene 450

Gene 976

Gene 675

Gene 161

Gene 877

In C5

Gene 779

Gene 625

Gene 557

Gene 239

Gene 336

Gene 81

Gene 57

Gene 500

Gene 487

Gene 860

Supporting structural dissimilarity between C1 and C2 (ordered rank wise): -

In C1

Gene 634

Gene 596

Gene 98

Gene 829

Gene 612

Gene 891

Gene 572

Gene 994

Gene 116

Gene 441

In C2

Gene 756

Gene 973

Gene 549

Gene 657

Gene 335

Gene 440

Gene 544

Gene 481

Gene 424

Gene 149

Answer 3

If the experientialist had forgotten to add chemical C4, the answers for part 1 wouldn't change as structurally C1, C2 are similar and C3, C5 are similar. On observation it is also observed that chemical C4 does not bring that amount of changes like the chemicals C1, C2, C3 and C5 bring. The difference between the average values of the properties of the cell after chemical C4 is added and the original properties is only **2.654474865.**

Chemical C4 is not structurally similar with any of the other 4 chemicals used. Therefore, the answer of part 1 wouldn't change.