



## R Session Aborted

R encountered a fatal error.

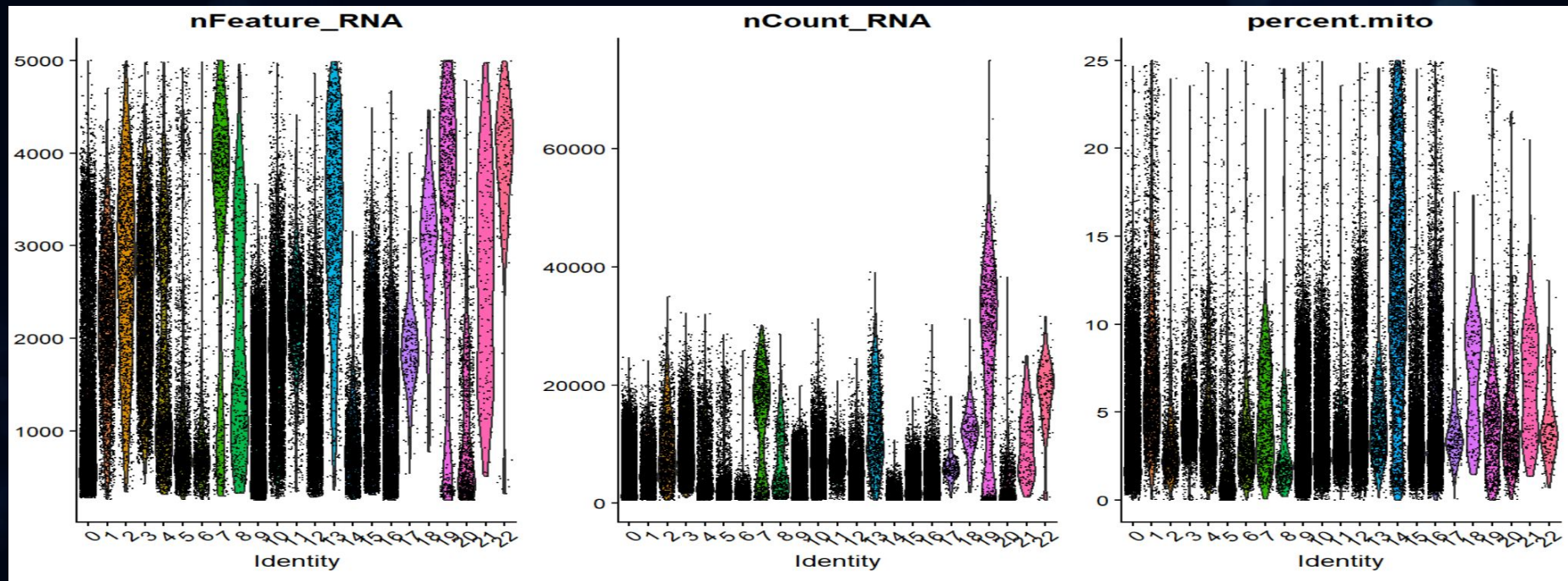
The session was terminated.

[Start New Session](#)

PharmaHacks 2024  
Holdings

# Data Exploration

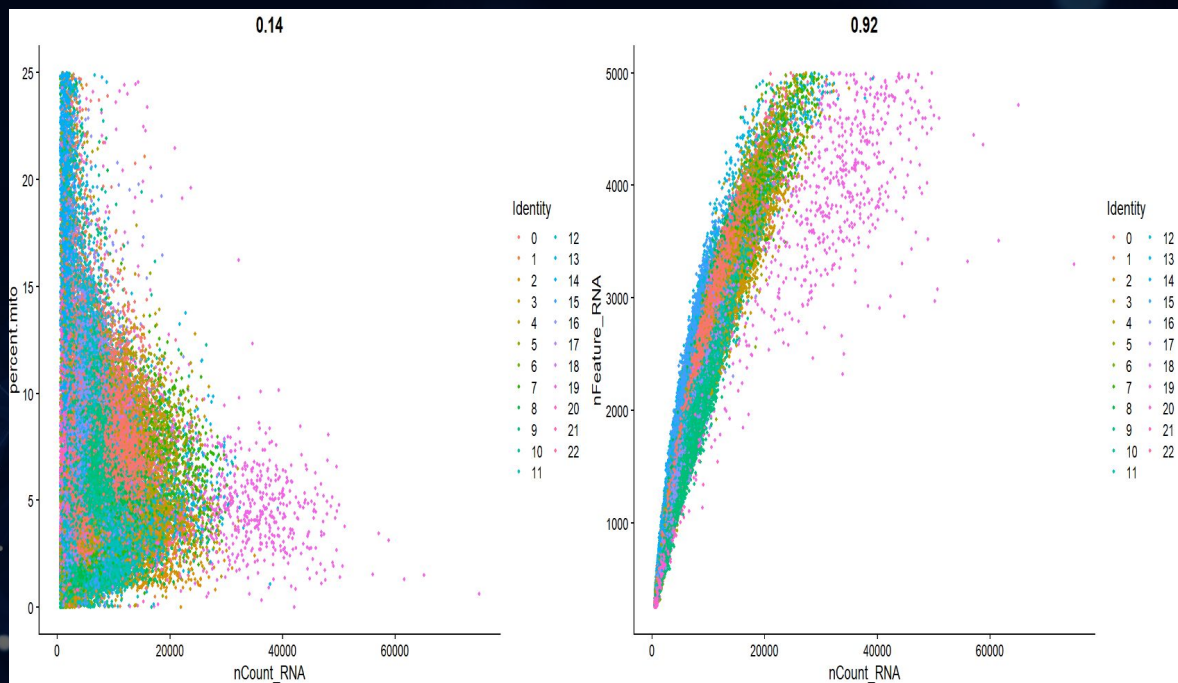
```
15 VlnPlot(cts, features = c("nFeature_RNA", "nCount_RNA", "percent.mito"), ncol = 3)
```



High Mitochondrial DNA  
means poor quality.

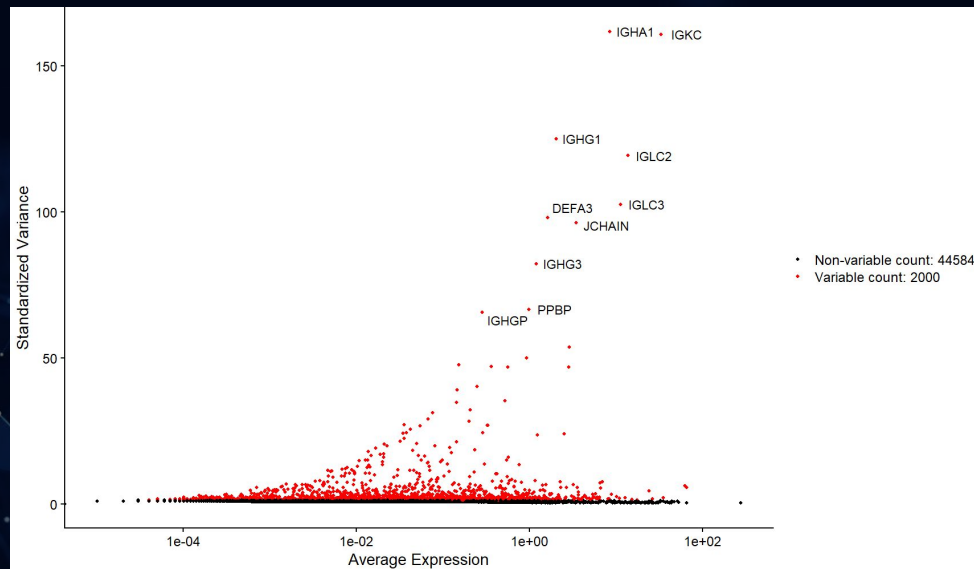
RNA features should be  
high proportional to  
count

```
16 plot1 <- FeatureScatter(cts, feature1 = "nCount_RNA", feature2 = "percent.mito")
17 plot2 <- FeatureScatter(cts, feature1 = "nCount_RNA", feature2 = "nFeature_RNA")
18 plot1 + plot2
```



## Top 10 Genes

```
20 cts <- FindVariableFeatures(cts, selection.method="vst", nfeatures =2000)
21 top10 <- head(VariableFeatures(cts), 10)
22 plot3 <- VariableFeaturePlot(cts)
23 LabelPoints(plot=plot3, points=top10, repel=TRUE)
```





# Scaling & PCA

```
> cts <- RunPCA(cts, features=variableFeatures(object=cts))
```

PC\_1

Positive: MND A, CTSS, LYZ, FCN1, VCAN, DUSP6, S100A9, CLEC7A, AIF1, CD14  
NCF2, AC020656.1, NAMPT, GRN, MAFB, SERPINA1, SPI1, KLF4, FGL2, CEBPD  
PLBD1, CST3, C5AR1, CSTA, CYBB, TYMP, MS4A6A, S100A12, CD68, IGSF6  
Negative: IL32, TRBC2, TRAC, LTB, IL7R, SYNE2, CD7, CD69, CD247, TRBC1  
CCL5, CTSW, TCF7, GZMA, ITM2A, ISG20, KLRB1, CD27, NKG7, GNLY  
LEF1, KLRD1, MAL, C12orf75, HOPX, PRF1, ARID5B, CCR7, FKBP11, GZMH

PC\_2

Positive: ATP5MC3, HSP90AA1, HIST1H1C, HSPA5, TLN1, PP1B, MTRNR2L12, HSP90B1, HIST1H1E, CALR  
CYTOR, HIST1H1D, ACTG1, JPT1, SYNE2, CD247, SELENOS, PRF1, PDIA6, GZMA  
HBB, TPM4, SPCS3, ATP5MC1, FGFBP2, ITGB1, CST7, PLEK, KIF2A, C12orf75  
Negative: ATP5E, GNB2L1, ATP5L, GPX1, C14orf2, RP11-1143G9.4, GLTSCR2, LST1, S100A8, S100A9  
FTL, LINC01272, CST3, LYZ, LGALS2, FAM26F, HLA-DRA, AIF1, TYROBP, CD74  
FCN1, FTH1, CFD, S100A12, CSTA, S100A6, AP1S2, SERPINA1, NCF1, CFP

PC\_3

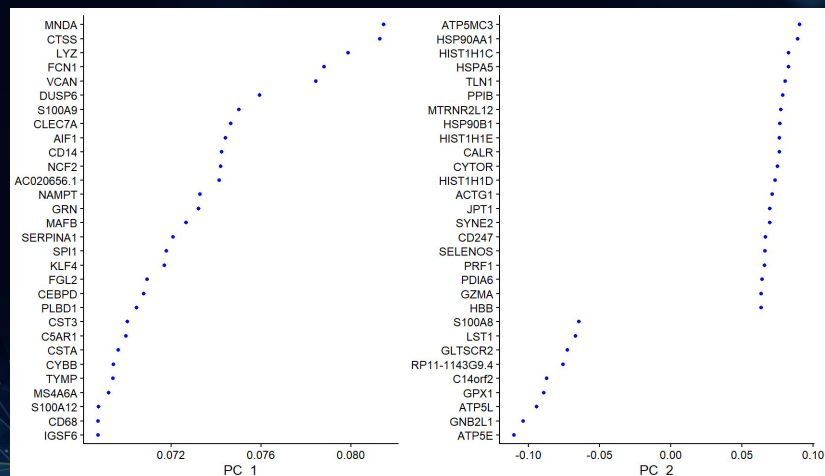
Positive: NKG7, GZMA, CST7, PRF1, S100A4, FGFBP2, CD247, KLRD1, GNLY, GZMB  
GZMH, HOPX, CTSW, CD7, SPON2, ANXA1, ID2, SYNE2, KLRF1, CCL4  
IL32, S100A6, IL2RB, MATK, ADGRG1, PP1B, HSPA5, TRDC, PSME2, MYBL1  
Negative: PF4, PPBP, CAVIN2, NGG11, TUBB1, GP9, PRKAR2B, CMTM5, TREML1, SPARC  
MYL9, ITGA2B, ACRBP, NRGN, MP1G6B, SNCA, CD9, MTURN, TSPAN33, TMEM40  
PDLIM1, PTCRA, CLEC1B, C2orf88, CTTN, F13A1, BEX3, HIST1H2AC, PTGS1, RUFY1

PC\_4

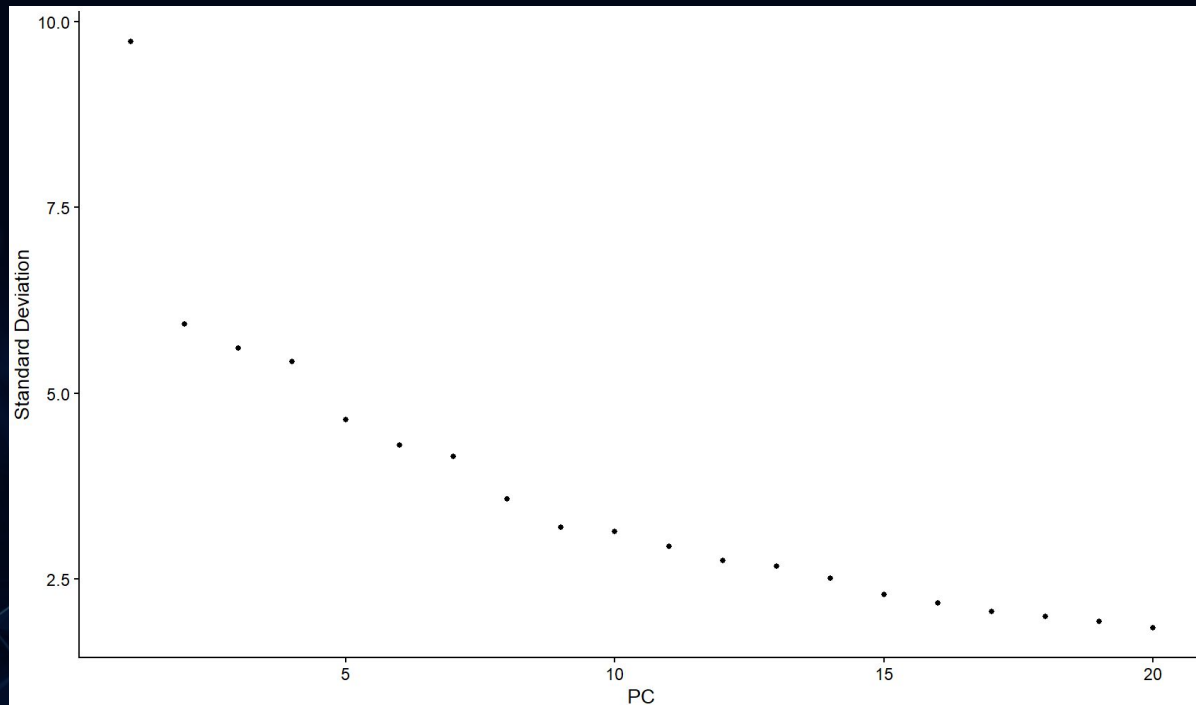
Positive: S100A4, CCL5, NKG7, S100A6, PF4, PPBP, CST7, CAVIN2, TUBB1, GP9  
NGG11, GNLY, TYROBP, CMTM5, TREML1, S100A8, PRKAR2B, SPARC, CTSW, SRGN  
MYL9, ITGA2B, NRGN, CLU, S100A9, GZMA, ACRBP, FCER1G, ATP5E, MP1G6B  
Negative: CD79A, MS4A1, IGHM, BANK1, IGHD, RALGPS2, HLA-DQA1, TNFRSF13C, LINC00926, CD79B  
CD22, FAM129C, BCL11A, HLA-DQB1, FCRL1, AFF3, COBLL1, SPIB, VPREB3, FCRL5  
TCL1A, HLA-DPA1, P2RX5, TCF4, BLK, FCER2, PAX5, HLA-DRA, HLA-DMB, ADAM28

PC\_5

Positive: S100P, SLC2A3, RFLNB, IL7R, MAL, ANXA3, TCF7, SOCS3, LTB, LEF1  
PROK2, FCGR3B, CYP4F3, PGLYRP1, BASP1, CCR7, CXCL8, NAMPT, IL1R2, CMTM2  
GOS2, TSHZ2, HP, LTF, AC245128.3, LCN2, CKAP4, SLC25A37, ARID5B, UGCG  
Negative: GZMB, GNLY, FCGR3A, NKG7, CD74, HLA-DPB1, PRF1, FGFBP2, HLA-DPA1, KLRD1  
ATP5E, GZMA, RHOC, KLRF1, HLA-DRA, GSTP1, TYROBP, CTSW, HLA-DRB1, SPON2  
GZMH, GNB2L1, CLIC3, CCL5, LGALS1, ATP5L, HOPX, CST3, CCL4, ABI3

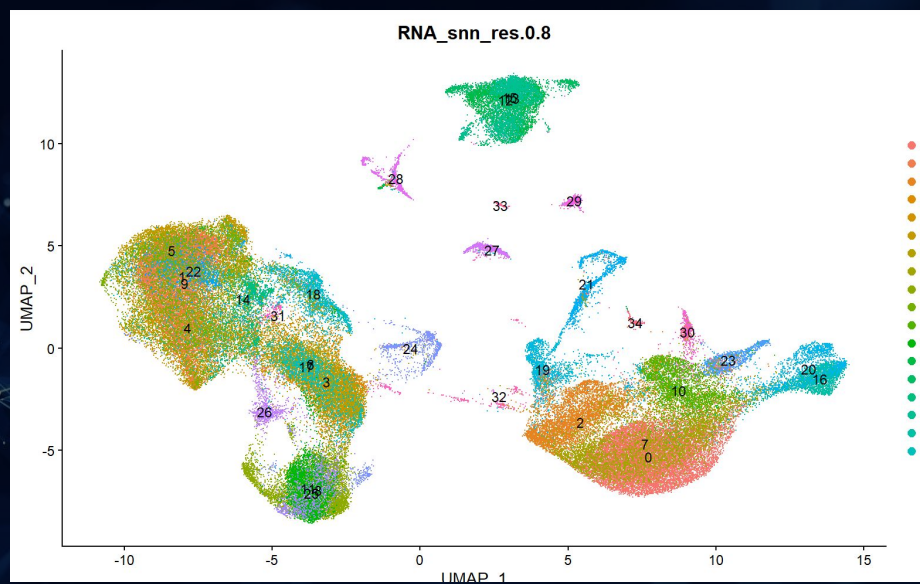
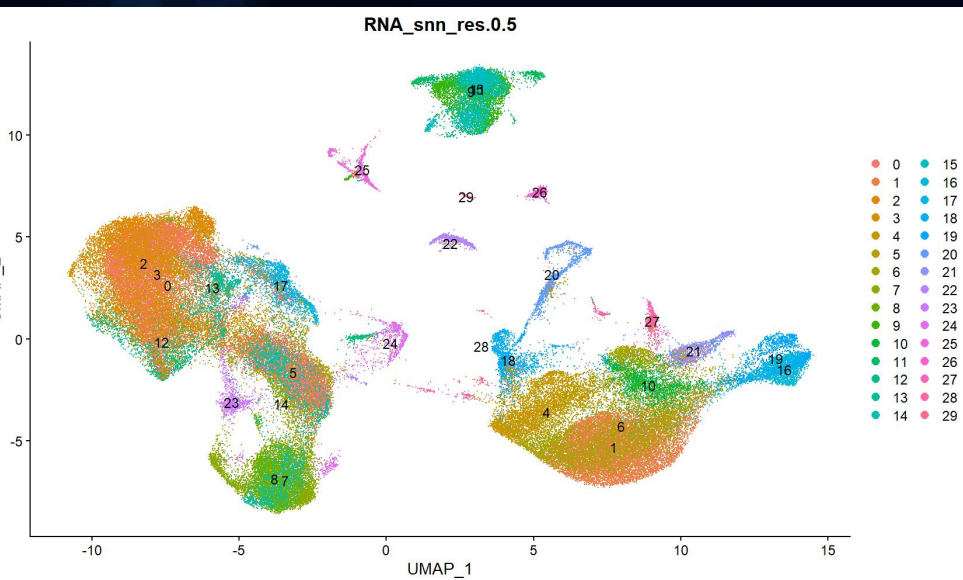


Elbow Plot, I  
think 18  
dimensions  
reasonably  
explains all  
variance

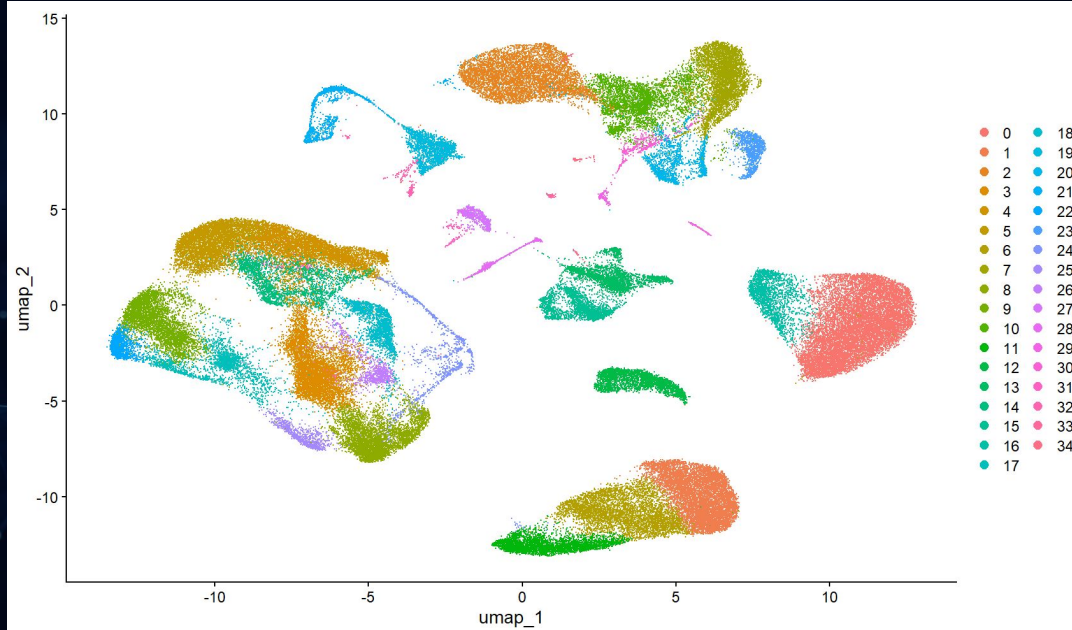


# Clustering

```
cts <- FindNeighbors(cts, dims = 1:18)
cts <- FindClusters(cts, resolution = 0.5)
DimPlot(cts, group.by = "RNA_snn_res.0.5", label = TRUE)
```



```
cts <- RunUMAP(cts, dims = 1:18)  
DimPlot(cts, reduction = "umap")
```





# SCFeatures for Aggregation

```
ffeatures <- run_gene_prop_celltype(  
  data = alldata,  
  type = "scrna",  
  num_top_gene = 100,  
  ncores = 1  
)  
ffeatures2 <- run_gene_mean(  
  alldata,  
  type = "scrna", num_top_gene = 150, ncores = 1  
)
```

"PBCM--S100A9", "PBCM--S100A8", "PBCM--LYZ", "PBCM--IFITM3", "PBCM--CST3", "PBCM--TYROBP", "PBCM--FCN1", "PBCM--S100A4", "PBCM--AIF1", "PBCM--FTL", "PBCM--IFI27", "PBCM--MMDA", "PBCM--S100A6",  
"C19-CB-0001", 0.611234889651356, 0.587437742114001, 0.58245711123409, 0.592418372993913, 0.511344770337576, 0.554786939679026, 0.491975650249032, 0.731322632400332, 0.494189263973437, 0.9619819673491886,  
"C19-CB-0003", 0.7580329461866, 0.74680603948966, 0.751451800232288, 0.605497483546234, 0.696089818041038, 0.739565931087882, 0.685636856368564, 0.818428184281843, 0.657375145180023, 0.9159898196673491886,  
"C19-CB-0002", 0.54136058767167, 0.510380070265091, 0.500479080166081, 0.36090705844778, 0.404982433727244, 0.486426061961035, 0.36346215266688, 0.739061002874481, 0.341424465027148, 0.8744809958767167,  
"C19-CB-0005", 0.904761904761905, 0.883597883597884, 0.894719894179894, 0.63265306122449, 0.8805744552003023, 0.90400604686319, 0.865457294028723, 0.928493935720786, 0.87371730914588, 0.975056666666667,  
"C19-CB-0009", 0.712142197599261, 0.729916897506925, 0.489072945521699, 0.392197599261311, 0.374192059095106, 0.561172668153389, 0.398891966759003, 0.874653739612188, 0.45848753462604, 0.980142197599261,  
"C19-CB-0012", 0.658961109600218, 0.676366603209138, 0.334239869458798, 0.0957302148490617, 0.246940440576557, 0.402502039706282, 0.222463965189013, 0.904269785150938, 0.272504759314659, 0.9820142197599261,  
"C19-CB-0013", 0.705492075193513, 0.725764835974935, 0.382602285293034, 0.290821968300774, 0.30925175082934, 0.573534832288979, 0.269812016218209, 0.929229635090306, 0.397714706966458, 0.987836666666667,  
"C19-CB-0011", 0.539699570815451, 0.56759656523605, 0.3036480686669528, 0.129828326180257, 0.190987124463519, 0.430257510729614, 0.21137339055794, 0.828326180257511, 0.273605150214592, 0.93669570815451,  
"C19-CB-0008", 0.752273285819612, 0.7621037109855, 0.443597935610715, 0.252641926763333, 0.294666994347506, 0.520275251904645, 0.387072990906857, 0.839764069796819, 0.434504792322268, 0.9778819673491886,  
"C19-CB-0020", 0.673264713644761, 0.699392979678015, 0.309316442333069, 0.109791501715492, 0.239904988123515, 0.417524412773819, 0.164951174452362, 0.834784903668514, 0.266297176035893, 0.997673285819612,  
"C19-CB-0021", 0.689137249391727, 0.70316301703163, 0.417579075425791, 0.368917274939173, 0.35948905109489, 0.504866180048662, 0.331812652068127, 0.906934306569343, 0.399330900243309, 0.9936137249391727,  
"C19-CB-0016", 0.8122028552414897, 0.817749603803486, 0.532091917591125, 0.40491283677036, 0.486529318541997, 0.755546751188589, 0.374009508716323, 0.913629160063391, 0.498811410459588, 0.998873285819612,  
"C19-CB-0198", 0.267045454545455, 0.189505347593583, 0.164104278074866, 0.058155080213904, 0.164438502673977, 0.183823529411765, 0.0718582887700535, 0.866978006625668, 0.12466577540107, 0.990973285819612,  
"C19-CB-0204", 0.266383280198371, 0.198016294712927, 0.211121291880623, 0.1775075425791, 0.155688324477506, 0.256818986893376, 0.106624158696422, 0.81339164009919, 0.233793907344315, 0.980516294712927,  
"C19-CB-0199", 0.263247863247863, 0.198290598290598, 0.1612551253561, 0.125925925925926, 0.168091168091168, 0.15042735042735, 0.086696860968666, 0.781766381766382, 0.202279202279202, 0.980673285819612,  
"C19-CB-0214", 0.111423747889702, 0.095668542487338, 0.060213843556556, 0.216657287563309, 0.128306133933596, 0.250422059651097, 0.0343275182892515, 0.795160382667417, 0.0787844682048396, 0.9613747889702,  
"C19-CB-0053", 0.10485651214128, 0.0607606471660044



# Future?

Needed to link the donors with their outcomes, and use that as a target.

Definitely dimensionality reduction.

But eventually I would take the CSV, split the dataset into 80:20, etc. then train and subsequently test to evaluate performance.

I was going to use some form on linear regression, because the benchmarking paper discussed favourable outcomes for them.