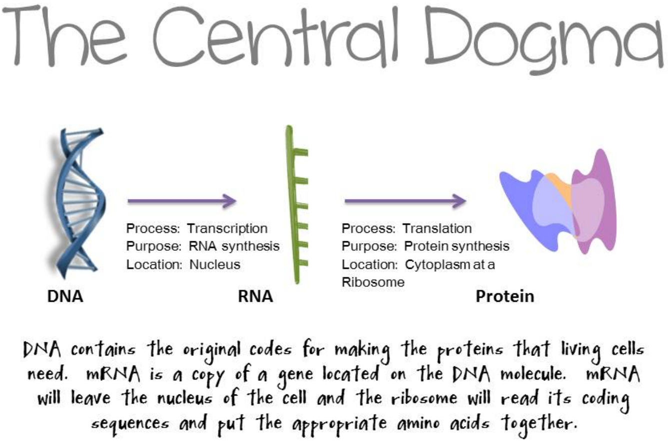
Gene Expression Data

STOR 565

Andrew Nobel

January, 2021

Genomics for Statisticians



The Central Dogma

DNA

Process: Transcription
Purpose: RNA synthesis
Location: Nucleus

RNA

Process: Translation
Purpose: Protein synthesis
Location: Cytoplasm at a Ribosome

Protein

DNA contains the original codes for making the proteins that living cells need. mRNA is a copy of a gene located on the DNA molecule. mRNA will leave the nucleus of the cell and the ribosome will read its coding sequences and put the appropriate amino acids together.

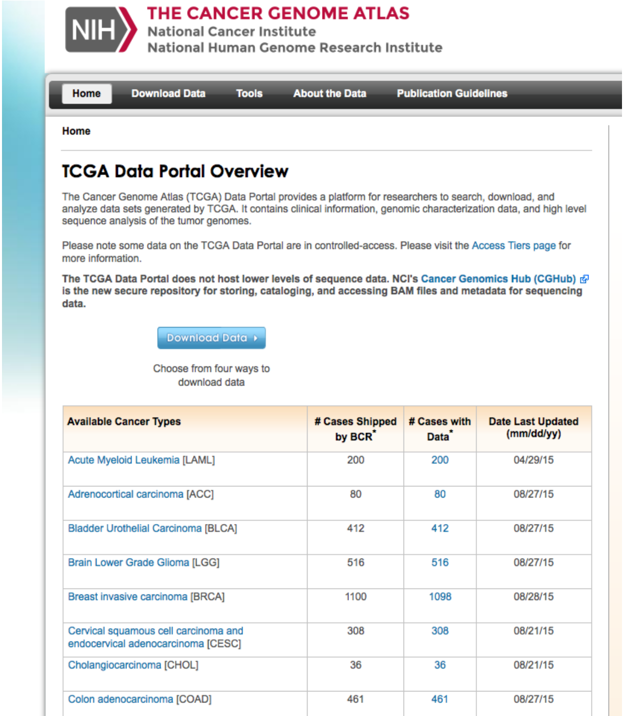
The Cancer Genome Atlas

Multi-Institution consortium supported by the National Institutes of Health.

Goal: “[T]o accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing”

Consortium collected tissue from thousands of tumors across numerous cancer types. Data derived from high-throughput technologies measuring

- Gene expression
- Micro-RNA
- DNA copy number
- Methylation



THE CANCER GENOME ATLAS
National Cancer Institute
National Human Genome Research Institute

Home

TCGA Data Portal Overview

The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA. It contains clinical information, genomic characterization data, and high level sequence analysis of the tumor genomes.

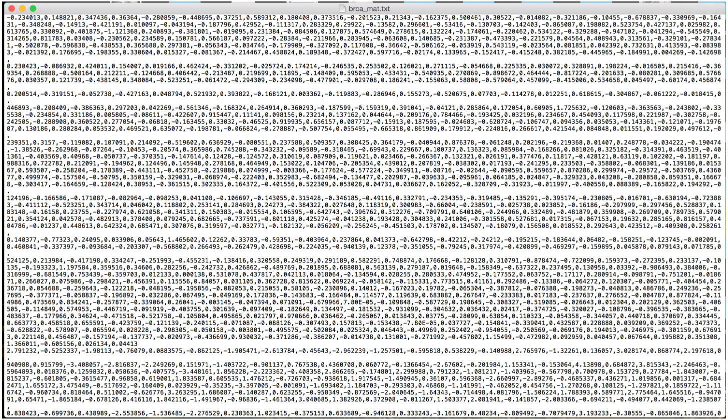
Please note some data on the TCGA Data Portal are in controlled-access. Please visit the Access Tiers page for more information.

The TCGA Data Portal does not host lower levels of sequence data. NCI's Cancer Genomics Hub (CGHub) is the new secure repository for storing, cataloging, and accessing BAM files and metadata for sequencing data.

Choose from four ways to download data

Available Cancer Types	# Cases Shipped by BCR*	# Cases with Data*	Date Last Updated (mm/dd/yy)
Acute Myeloid Leukemia [LAML]	200	200	04/29/15
Adrenocortical carcinoma [ACC]	80	80	08/27/15
Bladder Urothelial Carcinoma [BLCA]	412	412	08/27/15
Brain Lower Grade Glioma [LGG]	516	516	08/27/15
Breast invasive carcinoma [BRCA]	1100	1098	08/28/15
Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]	308	308	08/21/15
Cholangiocarcinoma [CHOL]	36	36	08/21/15
Colon adenocarcinoma [COAD]	461	461	08/27/15

Screenshot of Expression Data



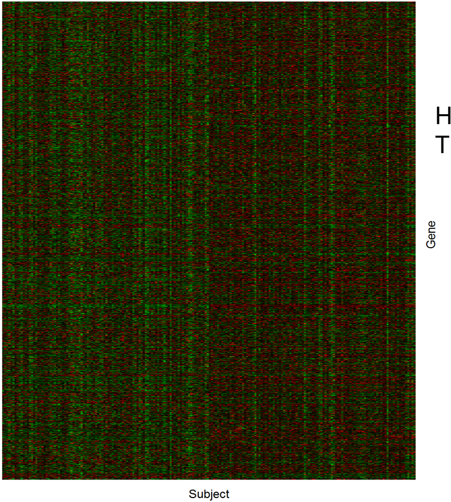
```
-0.234013,0.148821,0.347436,0.36364,-0.280859,-0.448695,0.37251,0.589312,0.180408,0.373516,-0.201523,0.21343,-0.162375,0.500461,0.30522,-0.014082,-0.321186,-0.10455,-0.678837,-0.330969,-0.164
31,-0.348248,-0.14913,-0.421191,0.010097,-0.043194,-0.187796,0.42952,-0.111317,0.283329,0.29922,-0.13582,0.296601,-0.53416,-0.130703,-0.142403,-0.865087,0.198082,0.523754,0.427137,0.025982,0.
613765,0.330092,-0.401875,-1.121368,0.240893,-0.381801,-0.019095,0.231384,-0.084506,0.127875,0.574649,0.278615,0.132224,-0.174061,-0.220462,0.534122,-0.329288,-0.947102,-0.041294,-0.545549,0.
314265,0.811783,0.03408,-0.230562,0.234589,0.150781,0.566187,0.097222,-0.28384,-0.211966,0.283945,-0.063608,0.140685,-0.231387,-0.473393,-0.221579,0.04564,0.408943,0.313561,-0.329101,-0.27834
1,-0.502078,-0.596838,-0.438553,0.365688,0.297381,-0.056343,-0.034746,-0.179909,-0.327092,0.117608,-0.36642,-0.506162,-0.053919,0.543119,-0.023564,0.081851,0.042392,0.732631,0.413593,-0.08398
3,-0.021392,0.176695,-0.198355,0.330604,0.015327,-0.081367,-0.214467,0.458824,0.189348,-0.372427,0.597716,-0.02174,0.133965,-0.152417,-0.415248,0.382185,-0.445965,-0.184991,0.004269,-0.142698
,
0.230423,-0.086932,0.424011,0.154007,0.019166,0.462424,-0.331202,-0.025724,0.174214,-0.246535,0.253502,0.126021,0.271115,-0.054668,0.225335,0.030072,0.328891,0.198224,-0.016505,0.215416,-0.36
9354,0.268888,-0.508164,0.212211,-0.124668,0.406442,-0.213487,0.219699,0.11895,-0.148409,0.595053,-0.433431,-0.540935,0.270869,-0.898672,0.464444,-0.817224,-0.201633,-0.080281,0.309685,0.5766
76,0.030357,0.121739,-0.438145,0.348084,-0.523251,-0.061472,-0.294309,-0.234098,-0.477901,-0.029708,0.186241,-0.155063,0.58808,-0.579064,0.457099,-0.415006,0.534658,0.045497,-0.60174,0.456874
,
0.200514,-0.319151,-0.052738,-0.427163,0.048794,0.521932,0.393822,-0.168121,0.003362,-0.119883,-0.286946,0.155273,-0.520675,0.07703,-0.114278,0.012251,0.618615,-0.304867,-0.061222,-0.018415,0
.
446893,-0.208409,-0.386363,0.297203,0.042269,-0.561346,-0.168324,0.264914,0.360293,-0.187599,-0.159319,0.391041,-0.04121,0.285864,0.172054,0.60905,1.725632,-0.120603,-0.363563,-0.243802,-0.33
5538,-0.234854,0.331186,0.005805,-0.08611,-0.422607,0.915447,0.11141,0.098156,0.23214,0.137162,0.044644,-0.209176,0.784466,-0.193425,0.032196,0.234667,0.454093,0.117598,0.221987,-0.302758,-0.
242505,-0.288908,0.360522,0.277054,-0.06818,-0.163455,0.33032,-0.46525,0.919935,0.656157,0.087712,-0.15913,0.187595,-0.024683,-0.628724,-0.106747,0.094356,0.035053,-0.143361,-0.121071,-0.1976
07,0.130186,0.280284,0.053532,0.469521,0.635072,-0.198781,-0.066824,-0.278887,-0.507754,0.055495,-0.665318,0.861909,0.179912,-0.224816,0.266617,0.421544,0.084848,0.011551,0.192029,0.497612,-0
.
239351,0.3157,-0.119082,0.107091,0.214092,-0.519602,0.636929,-0.080551,0.237588,0.509357,0.308425,0.364179,-0.040944,0.876378,-0.061248,0.202196,-0.219368,0.01407,0.248778,-0.034222,-0.190474
,-1.38526,-0.262968,-0.07264,-0.10453,-0.20574,0.365986,0.745288,-0.343232,-0.09589,-0.318465,-0.69943,0.229667,0.100737,0.136323,0.085984,-0.168266,0.081026,0.325182,-0.314391,0.463519,-0.40
1361,-0.403569,0.40968,-0.050737,-0.370351,-0.147614,0.12428,-0.124572,0.310619,0.087909,0.119621,0.023466,-0.266367,0.132321,0.026191,0.377476,0.11817,-0.428121,0.63119,0.102202,-0.181197,0.
988316,0.722782,0.212091,-0.194962,0.124496,0.145948,0.278168,0.464949,0.153022,0.104706,-0.205354,0.439012,0.207819,-0.038302,0.017193,-0.241295,0.233503,-0.358802,-0.068301,-0.139186,0.0153
67,0.593507,-0.258204,-0.178389,-0.443111,-0.452758,-0.219886,0.074999,-0.003366,-0.177624,-0.577224,-0.349911,-0.08716,-0.02644,-0.090595,0.559657,0.870286,0.299974,-0.29572,-0.503769,0.4360
77,0.499974,-0.157504,-0.50795,0.350159,-0.329031,-0.068974,-0.222403,0.352983,-0.682494,-0.134477,0.202987,-0.039633,-0.095961,0.064185,0.024847,-0.329323,0.043208,-0.280858,0.859351,0.16667
8,-0.303417,-0.164659,-0.128424,0.38953,-0.361515,0.302335,0.164372,-0.401556,0.522309,0.053028,0.04731,0.036627,0.162052,-0.328709,-0.31923,-0.011997,-0.400558,0.088389,-0.165822,0.194292,-0
.
124196,-0.166586,-0.171087,-0.082964,-0.098253,0.041108,-0.106697,-0.143055,0.315428,-0.346185,-0.49116,0.332791,-0.234353,-0.319485,-0.135291,-0.395174,-0.230805,-0.016701,-0.630194,-0.72388
3,-0.411112,-0.523251,0.343714,0.046042,0.118802,0.253141,0.284693,0.24273,-0.384322,0.027648,0.118319,0.300983,-0.66004,-0.238591,-0.025738,0.023852,-0.16186,-0.297999,-0.297456,0.528837,0.1
83148,-0.16158,0.23755,-0.227974,0.621058,-0.341311,0.150383,-0.015554,0.106595,-0.642743,-0.396762,0.312276,-0.709791,0.640106,-0.244966,0.332489,-0.481879,0.359908,-0.269709,0.789735,0.5790
21,0.354124,0.042578,-0.482913,0.378408,0.079245,0.682665,-0.737591,-0.801118,0.425274,-0.041238,0.193428,0.304833,0.241006,-0.301558,0.527681,0.017315,-0.067153,0.19632,0.285165,0.816157,0.4
04786,-0.01237,0.448613,0.642324,0.685471,0.307076,0.319597,-0.032771,-0.182132,-0.056209,-0.256245,-0.451503,0.178702,0.134507,-0.18079,0.156508,0.018552,0.292643,0.423512,-0.409308,0.258261
,
0.140377,-0.77323,0.24095,0.033906,0.05643,1.465602,0.12262,0.33783,-0.59351,-0.403964,0.237864,0.041373,-0.642798,-0.42212,-0.24212,-0.195215,-0.183644,0.06482,-0.158251,-0.123745,-0.002091,
0.460841,-0.337397,-0.093684,-0.203307,-0.568802,0.266493,-0.262479,0.428698,-0.224035,-0.940139,0.12378,-0.351055,-0.79245,0.317974,-0.420899,-0.469297,-0.159895,0.045878,0.079143,0.071785,0
.
524125,0.213984,-0.417198,0.334247,-0.251993,-0.455231,-0.138416,0.320558,0.249319,0.291189,0.582291,0.748874,0.176668,-0.128128,0.310791,-0.878474,-0.722099,0.159373,-0.272395,0.233137,-0.10
135,-0.193323,1.197584,0.359516,0.34606,0.282256,-0.242732,0.426862,-0.489769,0.201895,0.688081,0.563139,0.279187,0.019648,-0.158349,-0.637322,0.237495,0.130958,0.03392,-0.986493,0.384006,-0.
316999,-0.681549,0.753439,-0.359703,0.012133,0.000138,0.531078,0.437817,0.042113,0.018064,-0.134594,0.028255,0.280533,0.474952,-0.177552,0.063752,-0.17117,0.280914,-0.098791,-0.751201,-0.0186
71,0.266027,0.075986,-0.298421,-0.456391,0.115556,0.84057,0.031105,0.362728,0.815622,0.069224,-0.058142,-0.115331,0.773515,0.41161,0.292486,-0.13386,-0.064272,0.120307,-0.005771,-0.404454,0.2
36718,0.054688,-0.259643,-0.122218,-0.048195,-0.195856,-0.082053,0.215055,0.58105,-0.230896,0.14012,-0.167023,0.19782,-0.065304,-0.387812,-0.076388,-0.198273,-0.040813,0.486786,0.249236,-0.25
7695,-0.377371,-0.058837,-0.196892,-0.032286,0.067495,-0.049169,0.172836,-0.143683,-0.166484,0.114577,0.119639,0.683882,0.267647,-0.233383,0.017183,-0.237637,0.276652,-0.004787,0.877824,-0.11
4986,0.473569,0.834241,-0.257877,-0.339064,0.26041,-0.003145,-0.047394,0.071091,-0.679966,7.80E-05,-0.109848,-0.587729,0.198645,-0.308327,-0.519085,-0.026643,0.012304,0.202129,0.362503,-0.406
505,-0.114849,0.574953,-0.446719,-0.091919,-0.403755,0.301639,-0.097409,-0.182649,0.134497,-0.181532,-0.931099,-0.304632,0.036432,0.02417,-0.374725,-0.320027,-0.108796,-0.396535,-0.383665,-0.
483637,-0.177966,0.34624,-0.471518,-0.521758,-0.105804,0.495865,0.021797,0.970666,0.036462,-0.265067,0.013843,0.03775,-0.28099,0.63854,0.110323,-0.054358,-0.344057,0.440718,0.370697,0.334445,
0.663773,0.458518,0.655591,-0.423759,-0.121139,-0.240115,-0.071087,-0.088126,-0.307493,0.157813,-0.153438,-7.80E-05,0.037727,-0.154841,-0.339041,0.432587,0.228888,0.039209,0.369252,-0.347373,
-0.628822,-0.578907,-0.065594,0.028228,-0.298305,-0.050158,-0.003801,-0.495575,-0.502084,0.025324,0.046443,-0.49969,0.252402,-0.954055,-0.250569,-0.069176,0.194013,-0.246975,-0.301159,0.67691
3,0.221148,0.456487,-0.157194,-0.137737,-0.020973,-0.436699,0.930032,-0.371286,-0.386207,-0.014738,0.131001,-0.271992,-0.457802,1.15499,-0.472982,0.092959,0.040457,0.067644,0.195882,0.351308,
1.366011,-0.605156,0.026134,0.04413
2.791232,-0.5252337,-1.98113,-0.76079,0.0883575,-0.862125,-1.905471,-2.613784,-0.45643,-2.962239,-1.257501,-0.595818,0.538229,-0.140988,2.765976,-1.32261,0.136057,3.028174,0.868774,2.197618,0
.
940988,0.915799,-3.408057,-2.816837,-2.249269,0.151971,-1.403722,-0.901137,0.767538,0.4360708,0.060772,-0.1366454,-2.67602,-0.201984,1.153341,-0.153064,4.13898,0.684872,3.815343,-2.246463,-0.
5964893,0.018376,0.1259832,0.058636,-0.407575,-3.448161,1.856228,-2.223362,-0.408358,2.866265,-0.174801,2.299988,0.791232,-1.081217,-1.403963,-0.567798,0.700978,0.153729,0.27784,-1.843007,-0.
815237,-0.601805,-0.3615477,0.96858,0.619001,1.833587,0.605535,1.476212,-0.726703,-0.938618,1.917545,-1.490945,0.36107,0.596368,-2.660997,-2.89276,-0.4685337,0.436271,1.019856,0.001317,-0.684
2471,1.655172,3.475449,-0.517692,-0.160409,0.023929,-0.35235,-3.397005,-0.001091,-1.693402,1.184703,-0.293303,0.46868,-1.141991,-0.462052,0.454756,-1.270268,0.108125,-1.297821,0.1859722,-1.11
6742,-0.960734,0.818464,0.511002,-0.626776,3.263295,1.686087,-0.140287,0.623255,-0.958349,-0.872569,-2.040645,-1.64343,-0.714498,4.081796,-1.506224,1.788393,0.589771,2.775464,1.645516,-0.9470
91,0.65471,-1.865184,-0.678126,0.416116,1.842116,-1.491967,-0.96836,-1.461364,3.040685,1.382926,0.372908,-0.011267,1.503377,2.081941,-0.141857,-2.003366,-0.694211,-0.494077,0.442069,-1.238507
,
0.838423,-0.699736,0.438989,-2.553856,-1.536485,-2.276529,0.238363,1.023415,-0.375153,0.633689,-0.946128,0.333243,-3.161679,0.48234,-0.809492,-0.7079479,3.193233,-0.30555,-0.865846,1.863926,0
```

Heat Maps of Numerical Data

Heat map: Means of displaying a data matrix with numerical entries

- positive entries are red
- negative entries are green
- entries close to zero are black

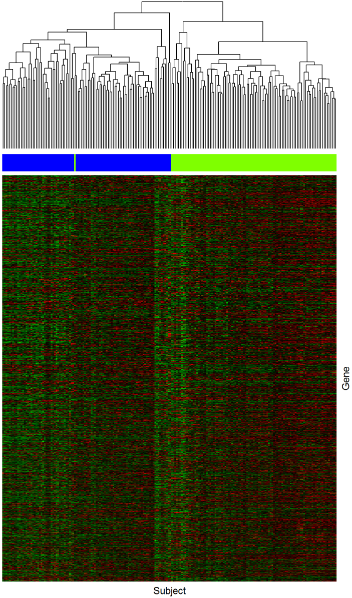
Example: Gene Expression Data from Breast Cancer



Heat map of gene expression data from The Cancer Genome Atlas (TCGA)

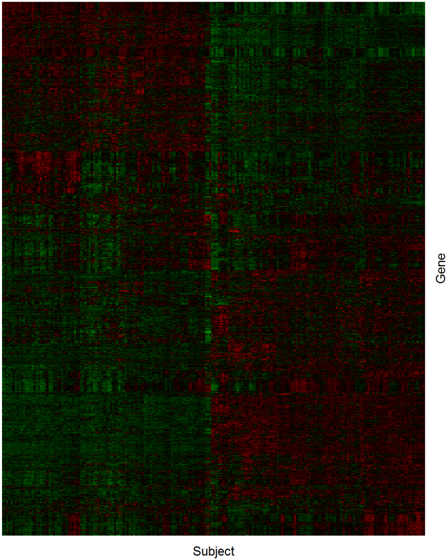
- Samples: $n = 200$ tumors from two breast cancer subtypes
  - 100 Luminal A tumors
  - 100 Basal tumors
- Variables: $p = 11,000$ genes (post-filtering)

Clustering Samples of TCGA Data

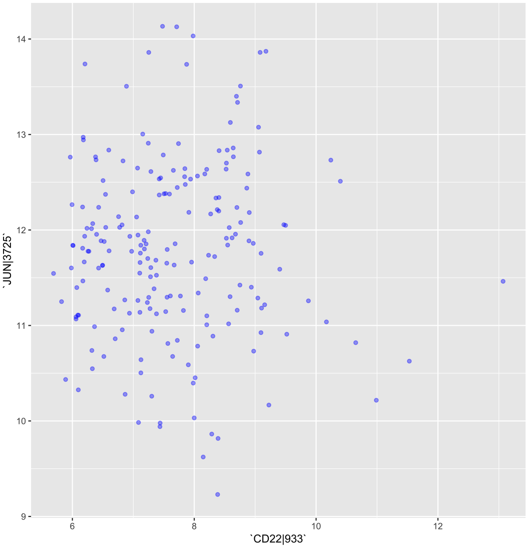


Colors: Luminal A and Basal

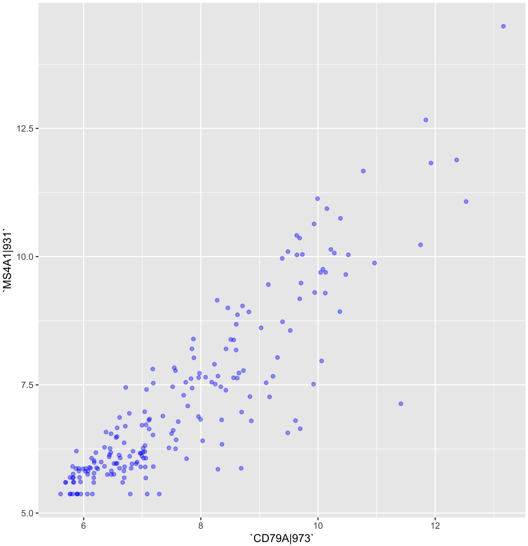
Clustering Rows and Columns



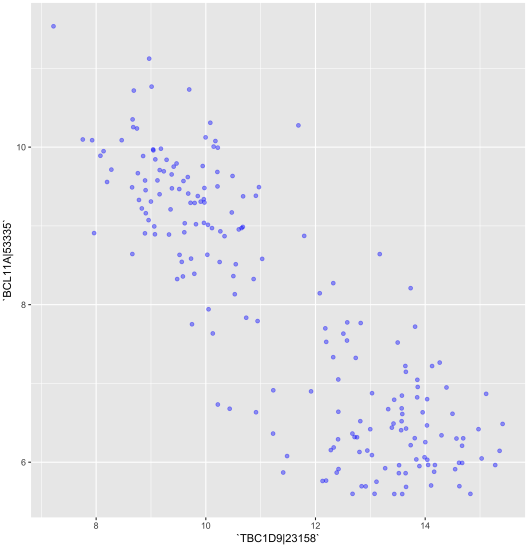
Two Uncorrelated Genes



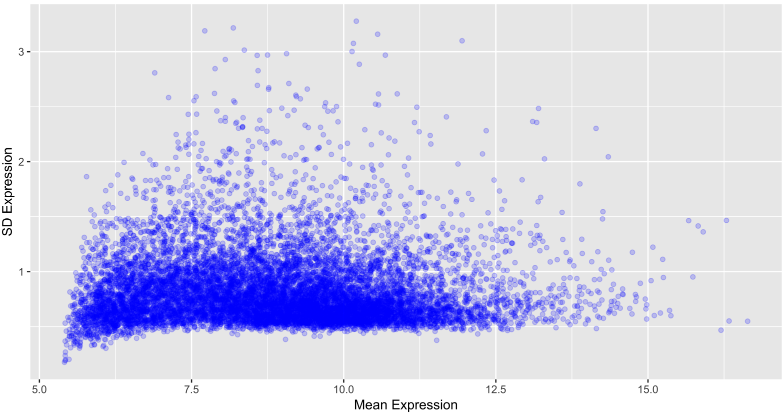
Two Positively Correlated Genes



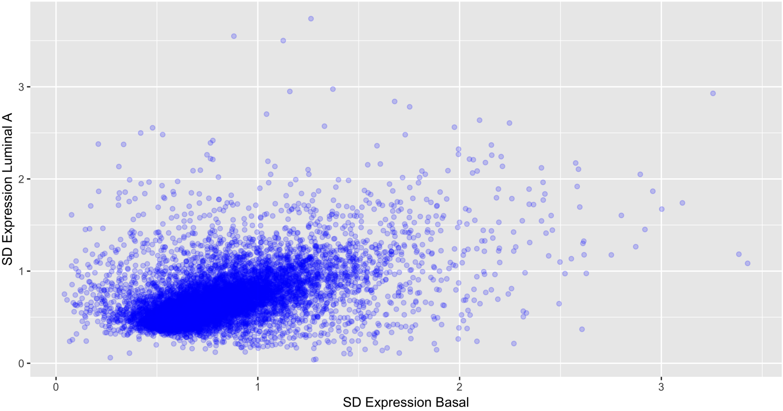
Two Negatively Correlated Genes



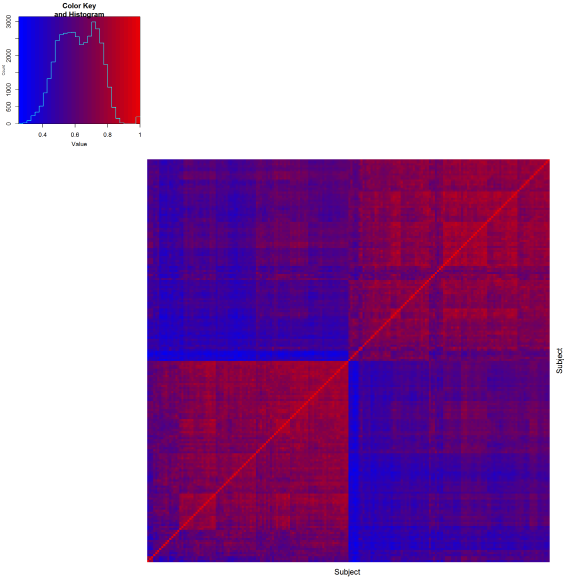
Comparison of Mean and SD of Expression for Genes



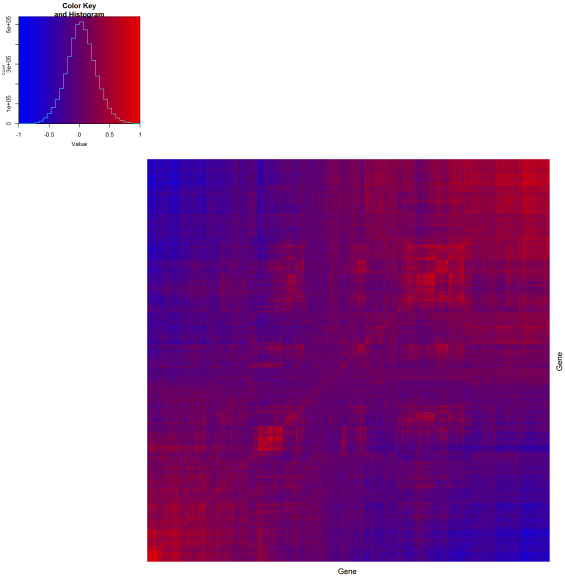
Scatterplot of SD (expression) for Two Subtypes



Heatmap: Correlation Matrix of Samples ($n \times n$)



Heatmap: Correlation Matrix of Genes ($p \times p$)


Color Key
and Histogram
Count
5e+05
3e+05
1e+05
Value
Gene
Gene