Appendix as item 10 listed as “GSE\_array\_meta.csv.”

Appendix as item 11 listed as ‘mrg5.csv’

Appendix as item 12 listed as “DE\_means\_Per\_Gene\_Chr.csv.”

Appendix as listed item 13 named “chr\_loci\_top\_genes.csv.”

Appendix as item 14 titled, “ub\_genes\_gviz.csv.”

Appendix as item 15 titled, “ensembl\_generated\_id.csv”

Appendix as item 16 titled, “ub\_genes\_ensembl.csv “

Appendix as item 17 titled, “mart\_export.txt”

Appendix as item 18 listed as “ub\_genes\_ensembl\_gviz.csv.”

Appedix as item 19 listed as “All-ggplot2-type-sample-derived.csv.”

Appendix as item 20 listed as “DE\_data\_unordered.csv.”

Appendix as item 21 listed as “MemberGviz\_130\_141.csv.”

Appendix as item 22 listed as “MemberMagnitude\_130\_142.csv.”

Appendix as item 23 listed as “top\_10\_and\_ubiq\_6.csv.” as ml ready 'TOP16\_ml\_ready.csv'

24: ubiq\_and\_top10\_samples\_only.csv for bootstrapping

25: Stats16.csv the bootstrap of 32 ul nonul

26 listed as 'most\_DE\_ml\_ready\_130.csv' top 16 mag in 130 subset of item 22

27 listed as “least\_DE16\_ml\_ready\_130.csv,”

28 listed as top 10 FC in UL+6  'FOLD16\_ml\_ready.csv'

29 listed as “majority\_ml\_ready\_10\_total.csv” the majority of 5 up, 5 down, 10 total

30. universe\_12173.csv

31. listed as 'most\_universe\_fold.csv', universe\_12173.csv

32 listed as 'most\_universe\_DE.csv', universe\_12173.csv

33 listed as 'least\_universe\_DE.csv', universe\_12173.csv

34. ML results most DE 130, “Results\_predictions\_DE16\_8\_algorithms\_used.csv”

35: ML results least DE 130, “Results\_predictions\_Least\_DE16\_8\_algorithms\_used.csv”

36: ML results top10+6 FC 130, “Results\_predictions\_FOLD16\_8\_algorithms\_used.csv”

37: ML results majority 10 from 130, “Results\_predictions\_majority10\_8\_algorithms\_used.csv”

38: ML results top 16 universe Fold, “Results\_predictions\_universe16\_fold\_8\_algorithms\_used.csv”

39: ML results most universe DE, “Results\_predictions\_universe16\_DE\_8\_algorithms\_used.csv“

40: ML results least universe DE, “Results\_predictions\_universe16\_DE\_least\_8\_algorithms\_used.csv “

In Document Tables and Figures

Files for tables: these t1,t3:t10 are the gene, HGNC name and fwd/rev strand and cytoband

Table 1. The top 10 plus six UL risk genes

Table 2: Bootstrap Simulated Results for Top 10 plus 6 Genes

Table 3: Member Majorities of Five Most Changed Up or Down

Table 4: Top 16 Genes Differentially Expressed in Subset

Table 5: Bottom 16 Genes Differentially Expressed in Subset

Table 6: Top 16 Fold Change in Subset

Table 7: Majority of 10 Most Differentially Expressed Gene Up and Down

Table 8: Top 16 Genes in Fold Change from All

Table 9: Top 16 Genes Differentially Expressed in All

Table 10: Least Expressed 16 Genes in All

Table 11: Machine Learning Results on Top 10 Plus 6

Table 12: Machine Learning Results on All Data Sets

Files for figures:

Figure 1: Histograms of UL Simulated Means for Top 10 Plus 6 Genes

Figure 2: Reverse Strand of Cytoband 11p15.5 Genes Expressed More in UL Near BET1L

Figure 3. Forward Strand of Cytoband 12q14.3 Genes Expressed Less in UL Near HMGA2

Figure 4. Gviz Map of Reverse Strand of Cytoband 17q25.3 Genes Expressed Less in UL.

Figure 5. Forward Strand of Cytoband 22q13.1 Majority of Genes Expressed More in UL

Figure 6: Heatmap of Top 10 Plus Six Genes in All Samples

Figure 7: Pairwise Comparison of All Top 10 Plus 6 Genes.

Figure 8: Comparison of Simulated Means for Non-UL and UL Top 10 Plus Six Genes

Table 1: The top 10 plus six UL risk genes

|  |  |  |
| --- | --- | --- |
| gene | GENE\_NAME | CYTOBAND |
| ASPSCR1 | alveolar soft part sarcoma chromosome region, candidate 1 | hs|17q25.3 |
| BET1L | blocked early in transport 1 homolog (S. cerevisiae)-like | hs|11p15.5 |
| CBX2 | chromobox homolog 2 | hs|17q25.3 |
| CBX7 | chromobox homolog 7 | hs|22q13.1 |
| CCDC57 | coiled-coil domain containing 57 | hs|17q25.3 |
| CYTH4 | cytohesin 4 | hs|22q13.1 |
| FASN | fatty acid synthase | hs|17q25.3 |
| GRIP1 | glutamate receptor interacting protein 1 | hs|12q14.3 |
| HMGA2 | high mobility group AT-hook 2 | hs|12q14.3 |
| KDELR3 | KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 3 | hs|22q13.1 |
| PYCR1 | pyrroline-5-carboxylate reductase 1 | hs|17q25.3 |
| RAC3 | ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3) | hs|17q25.3 |
| SOCS3 | suppressor of cytokine signaling 3 | hs|17q25.3 |
| TH | tyrosine hydroxylase | hs|11p15.5 |
| TNRC6B | trinucleotide repeat containing 6B | hs|22q13.1 |
| ZNF750 | zinc finger protein 750 | hs|17q25.3 |

Table 2: Bootstrap Simulated Results for Top 10 plus 6 Genes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Top 10 + 6 Genes | nonUL\_simulated\_mean | nonUL\_std\_dev | UL\_simulated\_mean | UL\_std\_dev | simmulated\_magnitude\_changed |
| ASPSCR1 | 5.55 | 0.12 | 6.17 | 0.11 | 0.62 |
| BET1L | 5.4 | 0.22 | 5.83 | 0.19 | 0.44 |
| CBX2 | 3.91 | 0.16 | 4.7 | 0.15 | 0.8 |
| CBX7 | 9.16 | 0.19 | 8.5 | 0.16 | 0.65 |
| CCDC57 | 4.05 | 0.16 | 4.2 | 0.14 | 0.15 |
| CYTH4 | 5.47 | 0.1 | 5.3 | 0.08 | 0.17 |
| FASN | 5.29 | 0.09 | 5.52 | 0.09 | 0.23 |
| GRIP1 | 2.64 | 0.22 | 3.32 | 0.2 | 0.68 |
| HMGA2 | 3.64 | 0.13 | 3.74 | 0.2 | 0.09 |
| KDELR3 | 6.76 | 0.1 | 7.49 | 0.12 | 0.72 |
| PYCR1 | 6.21 | 0.14 | 6.89 | 0.16 | 0.68 |
| RAC3 | 2.04 | 0.2 | 2.78 | 0.23 | 0.74 |
| SOCS3 | 6.01 | 0.26 | 5.24 | 0.17 | 0.77 |
| TH | 2.76 | 0.25 | 3.41 | 0.26 | 0.65 |
| TNRC6B | 6.85 | 0.14 | 6.92 | 0.13 | 0.07 |
| ZNF750 | 1.35 | 0.24 | 0.53 | 0.22 | 0.82 |

|  |  |  |  |
| --- | --- | --- | --- |
| A screenshot of a video game  Description automatically generated  ASPSCR1 | A screenshot of a video game  Description automatically generated  BET1L | A screenshot of a video game  Description automatically generated  CBX2 | A picture containing screenshot  Description automatically generated  CBX7 |
| A picture containing screenshot, text  Description automatically generated  CCDC57 | A screenshot of a video game  Description automatically generated  CYTH4 | A screenshot of a video game  Description automatically generated  FASN | A screenshot of a video game  Description automatically generated  GRIP1 |
| A picture containing screenshot, text  Description automatically generated  HMGA2 | A screenshot of a video game  Description automatically generated  KDELR3 | A picture containing screenshot  Description automatically generated  PYCR1 | A picture containing screenshot, text  Description automatically generated  RAC3 |
| A picture containing screenshot  Description automatically generated  SOCS3 | A screenshot of a video game  Description automatically generated  TH | A picture containing screenshot  Description automatically generated  TNRC6B | A picture containing screenshot  Description automatically generated  ZNF750 |

Figure 5: Histograms of UL Simulated Means for Top 10 Plus 6 Genes

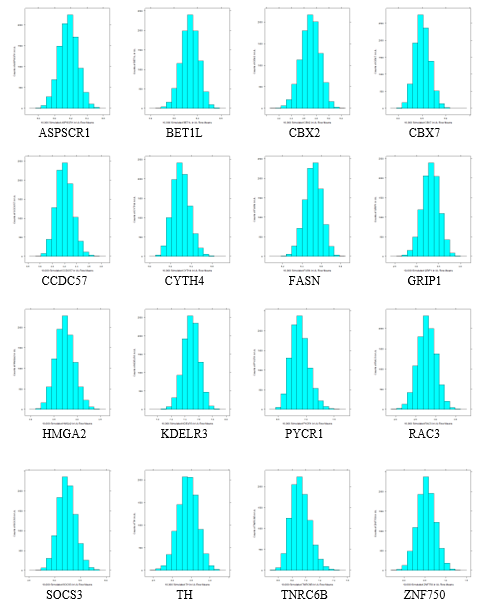


Figure 5: Histograms of UL Simulated Means for Top 10 Plus 6 Genes

**Table 3**: Member Majorities of Five Most Changed Up or Down

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **genes** | **type** | **all** | **up** | **down** | **majority** | **strand** | **cytoband** | **diff\_expr** |
| EPS8L2 | down | 33 | 16 | 17 | TRUE | + | hs|11p15.5 | -0.5 |
| TNNI2 | down | 33 | 16 | 17 | TRUE | + | hs|11p15.5 | -0.48 |
| SCT | down | 33 | 16 | 17 | TRUE | - | hs|11p15.5 | -0.36 |
| INS | down | 33 | 16 | 17 | TRUE | - | hs|11p15.5 | -0.32 |
| RPLP2 | down | 33 | 16 | 17 | TRUE | + | hs|11p15.5 | -0.32 |
| KDELR3 | up | 43 | 25 | 18 | TRUE | + | hs|22q13.1 | 0.72 |
| GRIP1 | up | 6 | 5 | 1 | TRUE | - | hs|12q14.3 | 0.68 |
| MICALL1 | up | 43 | 25 | 18 | TRUE | + | hs|22q13.1 | 0.58 |
| ADSL | up | 43 | 25 | 18 | TRUE | + | hs|22q13.1 | 0.43 |
| MGAT3 | up | 43 | 25 | 18 | TRUE | + | hs|22q13.1 | 0.42 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SampleID | **RF** | **RF2** | **LDA** | **GBM** | **KNN** | **RPART** | **GLM** | **Combined** | **Type** |
| gsm1667145 | nonUL | UL | nonUL | UL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm336254 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm336258 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm336260 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm336270 | UL | UL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm336273 | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm336276 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm52662 | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm52663 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm52665 | UL | UL | nonUL | UL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm52667 | UL | UL | UL | UL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm52669 | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm9099 | UL | UL | nonUL | UL | UL | UL | nonUL | UL | nonUL |
| gsm569425 | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm569427 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm336202ul | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | UL |
| gsm336208ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336209ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336214ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336215ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336218ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336220ul | nonUL | nonUL | UL | UL | nonUL | nonUL | nonUL | UL | UL |
| gsm336229ul | UL | UL | UL | UL | UL | nonUL | UL | UL | UL |
| gsm336232ul | UL | UL | UL | UL | UL | nonUL | UL | UL | UL |
| gsm336234ul | UL | UL | UL | UL | nonUL | UL | UL | UL | UL |
| gsm336238ul | UL | UL | nonUL | UL | nonUL | UL | nonUL | nonUL | UL |
| gsm336239ul | UL | UL | nonUL | nonUL | UL | UL | nonUL | UL | UL |
| gsm336240ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336241ul | nonUL | nonUL | UL | nonUL | nonUL | nonUL | UL | UL | UL |
| gsm336245ul | nonUL | nonUL | UL | nonUL | UL | nonUL | UL | UL | UL |
| gsm336248ul | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | UL |
| gsm38689ul | nonUL | nonUL | nonUL | UL | UL | UL | nonUL | UL | UL |
| gsm38692ul | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | UL |
| gsm9094ul | UL | UL | UL | nonUL | UL | UL | nonUL | UL | UL |
| gsm569429ul | UL | UL | nonUL | UL | UL | UL | nonUL | UL | UL |
| **results** | **0.69** | **0.66** | **0.77** | **0.69** | **0.77** | **0.54** | **0.74** | **0.86** | **100** |

**Table 10**: Machine Learning Results on Top 10 Plus 6

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SampleID** | **RF** | **RF2** | **LDA** | **GBM** | **KNN** | **RPART** | **GLM** | **Combined** | **Type** |
| gsm1667145 | nonUL | UL | nonUL | UL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm336254 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm336258 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm336260 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm336270 | UL | UL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm336273 | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm336276 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm52662 | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm52663 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm52665 | UL | UL | nonUL | UL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm52667 | UL | UL | UL | UL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm52669 | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm9099 | UL | UL | nonUL | UL | UL | UL | nonUL | UL | nonUL |
| gsm569425 | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | nonUL |
| gsm569427 | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL |
| gsm336202ul | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | UL |
| gsm336208ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336209ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336214ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336215ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336218ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336220ul | nonUL | nonUL | UL | UL | nonUL | nonUL | nonUL | UL | UL |
| gsm336229ul | UL | UL | UL | UL | UL | nonUL | UL | UL | UL |
| gsm336232ul | UL | UL | UL | UL | UL | nonUL | UL | UL | UL |
| gsm336234ul | UL | UL | UL | UL | nonUL | UL | UL | UL | UL |
| gsm336238ul | UL | UL | nonUL | UL | nonUL | UL | nonUL | nonUL | UL |
| gsm336239ul | UL | UL | nonUL | nonUL | UL | UL | nonUL | UL | UL |
| gsm336240ul | UL | UL | UL | UL | UL | UL | UL | UL | UL |
| gsm336241ul | nonUL | nonUL | UL | nonUL | nonUL | nonUL | UL | UL | UL |
| gsm336245ul | nonUL | nonUL | UL | nonUL | UL | nonUL | UL | UL | UL |
| gsm336248ul | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | nonUL | UL |
| gsm38689ul | nonUL | nonUL | nonUL | UL | UL | UL | nonUL | UL | UL |
| gsm38692ul | nonUL | nonUL | nonUL | nonUL | nonUL | UL | nonUL | nonUL | UL |
| gsm9094ul | UL | UL | UL | nonUL | UL | UL | nonUL | UL | UL |
| gsm569429ul | UL | UL | nonUL | UL | UL | UL | nonUL | UL | UL |
| **results** | **0.69** | **0.66** | **0.77** | **0.69** | **0.77** | **0.54** | **0.74** | **0.86** | **100** |