

META-ANALYSIS OF THE GENES UBIQUITOUSLY
ASSOCIATED WITH HUMAN UTERINE LEIOMYOMA
DEVELOPMENT IN HEALTHY HUMANS USING
THE GENE EXPRESSION OMNIBUS DATA

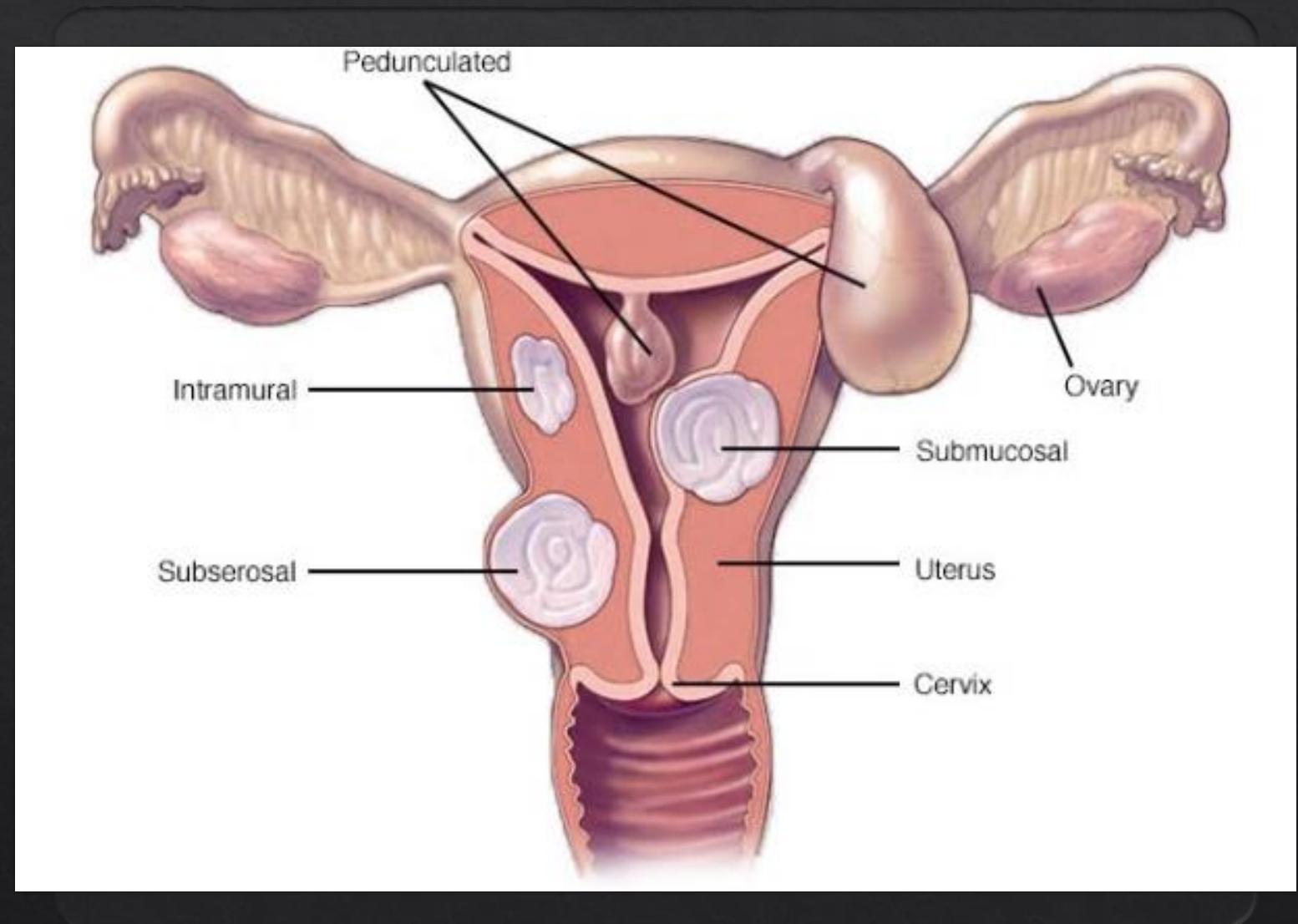
Janis Corona

2019

Description of Uterine Leiomyoma

What is a
uterine
leiomyoma
(UL)?

Uterine Leiomyoma in Different Uterine Layers



Description of Uterine Leiomyoma

What are the
symptoms of
UL?

Description of Uterine Leiomyoma



Who can get UL?



What are risk factors of UL?

Description of Uterine Leiomyoma

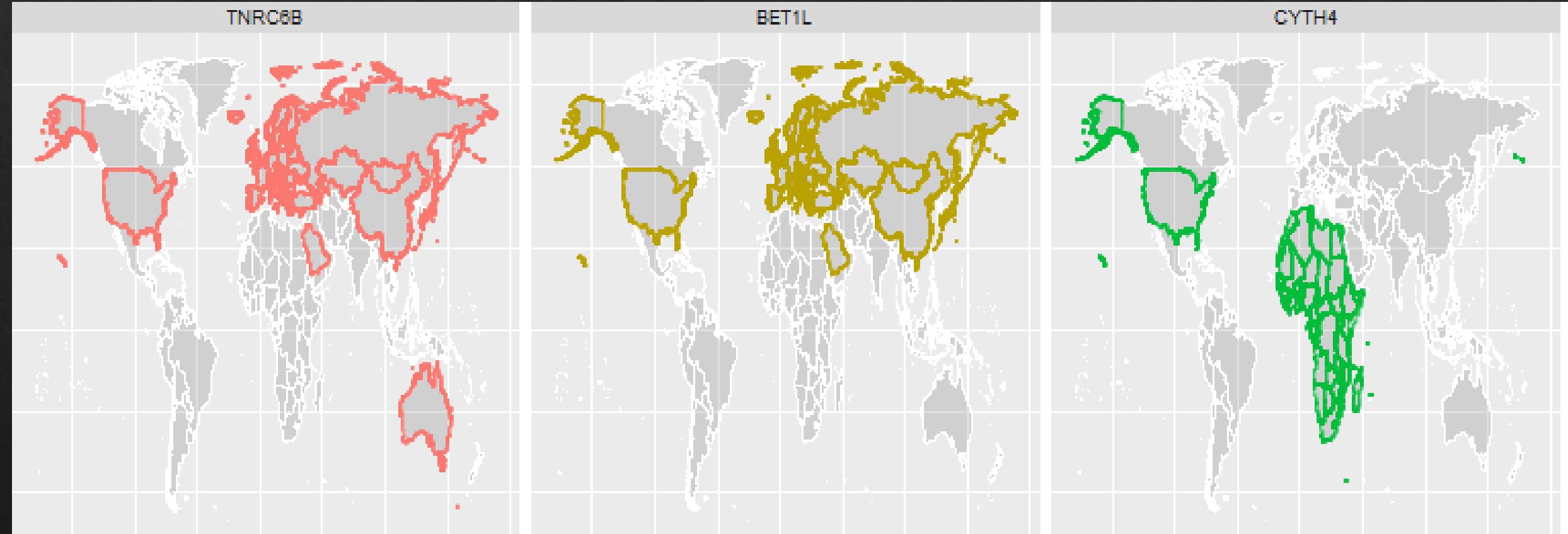


What is the treatment for UL?



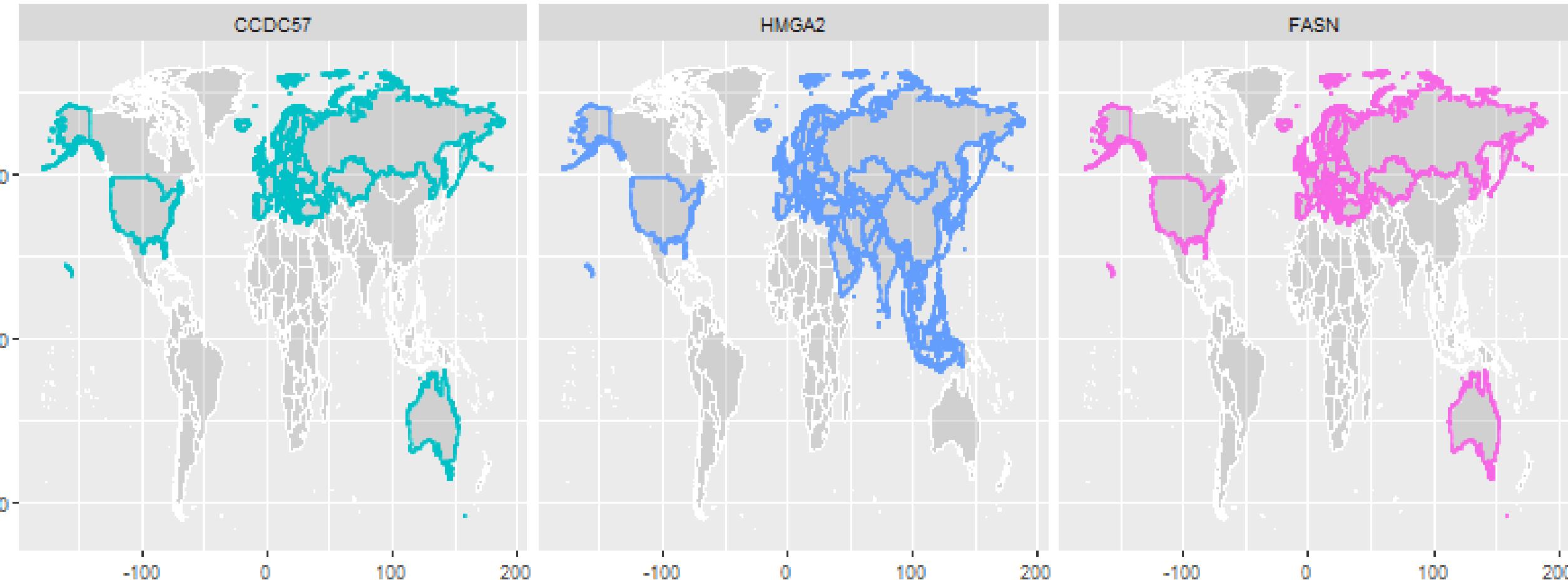
How do UL develop?

UL Risk in Population Studies: TNRC6B, BET1L, and CYTH4 as Top UL Risk Genes



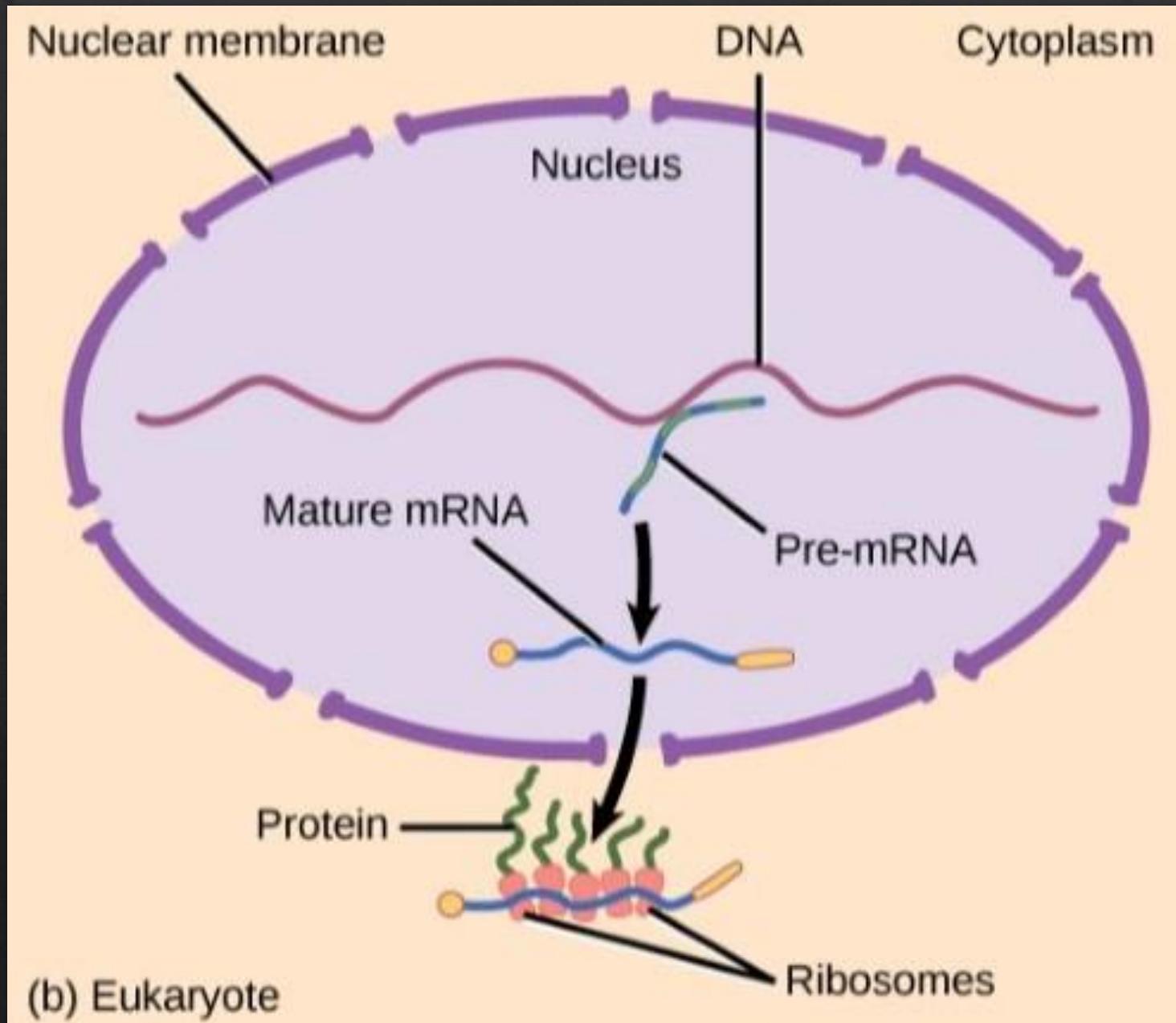
R, version 3.6, 2019

UL Risk in Population Studies: CCDC57, HMGA2, and FASN as Next Best UL Risk Genes



Gene Regulation: Transcription and Translation

Rye, Wise, Jurukovski, DeSai,
Choi, & Avissair (2017) *Biology:*
OpenStax



Gene Regulation: Gene Expression



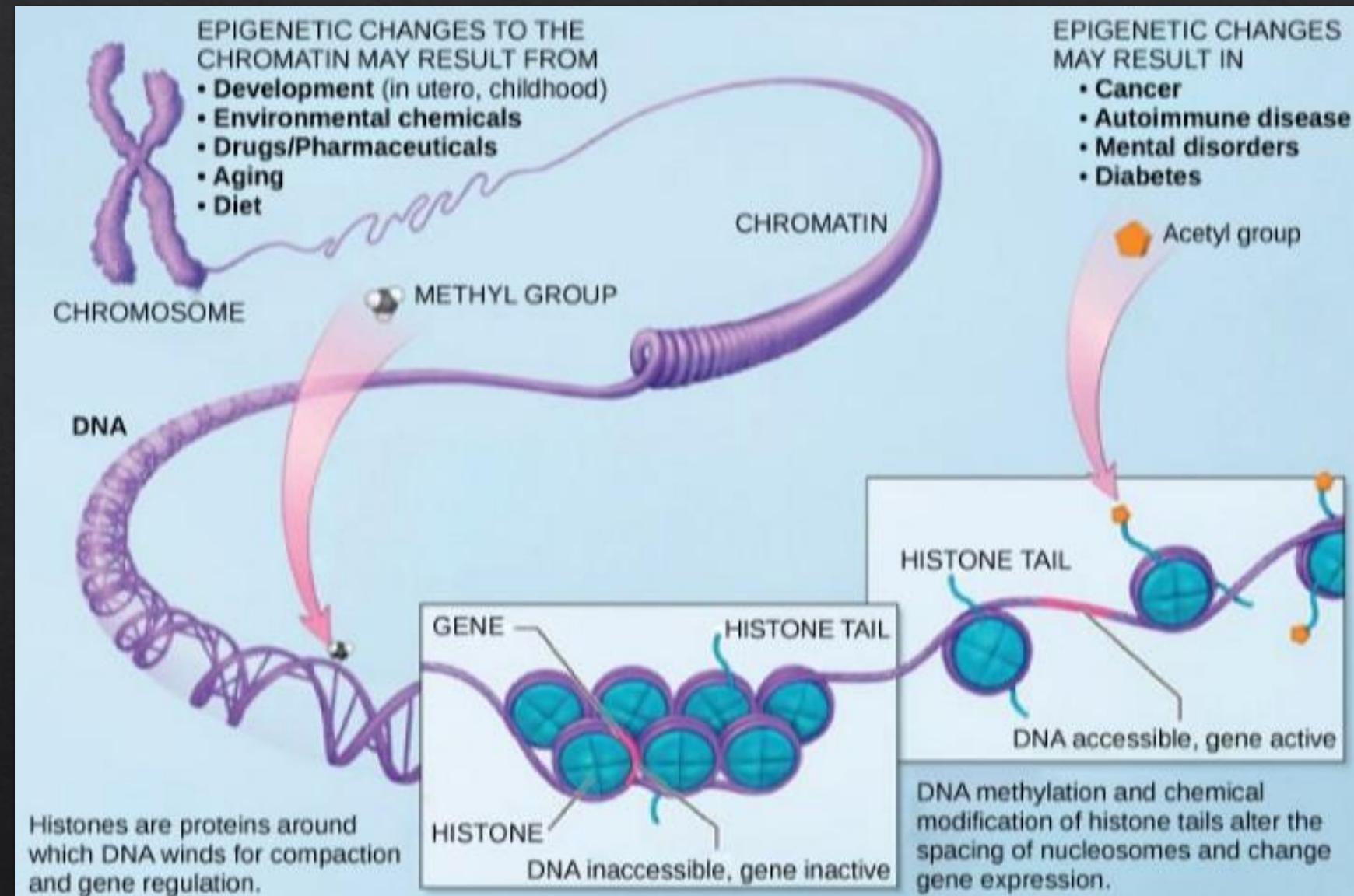
What affects gene expression
of DNA?



Where can the gene expression
of DNA be influenced?

Gene Regulation: Transcription

Rye, Wise, Jurukovski, DeSai,
Choi, & Avissair (2017) *Biology:*
OpenStax



Objective of this Research

Does gene expression data lead to gene targets in UL risk and UL pathogenesis?

- ❖ Seek out a way to find a relationship between genes originating from the same cytobands as the UL risk genes and expand to the universe of all genes
- ❖ Use machine learning on the data derived from these scenarios to test UL prediction of gene targets discovered
- ❖ Confirm if gene expression data of the six UL risk genes is enough to predict UL in a UL sample that is not from a specific population study

Methods: Genes Belonging to Same Cytoband Locations as Six Genes Ubiquitous to UL Risk Studies Isolated



All GEO data combined from five studies in R



There were a total of 12,173 genes in common after removing duplicates

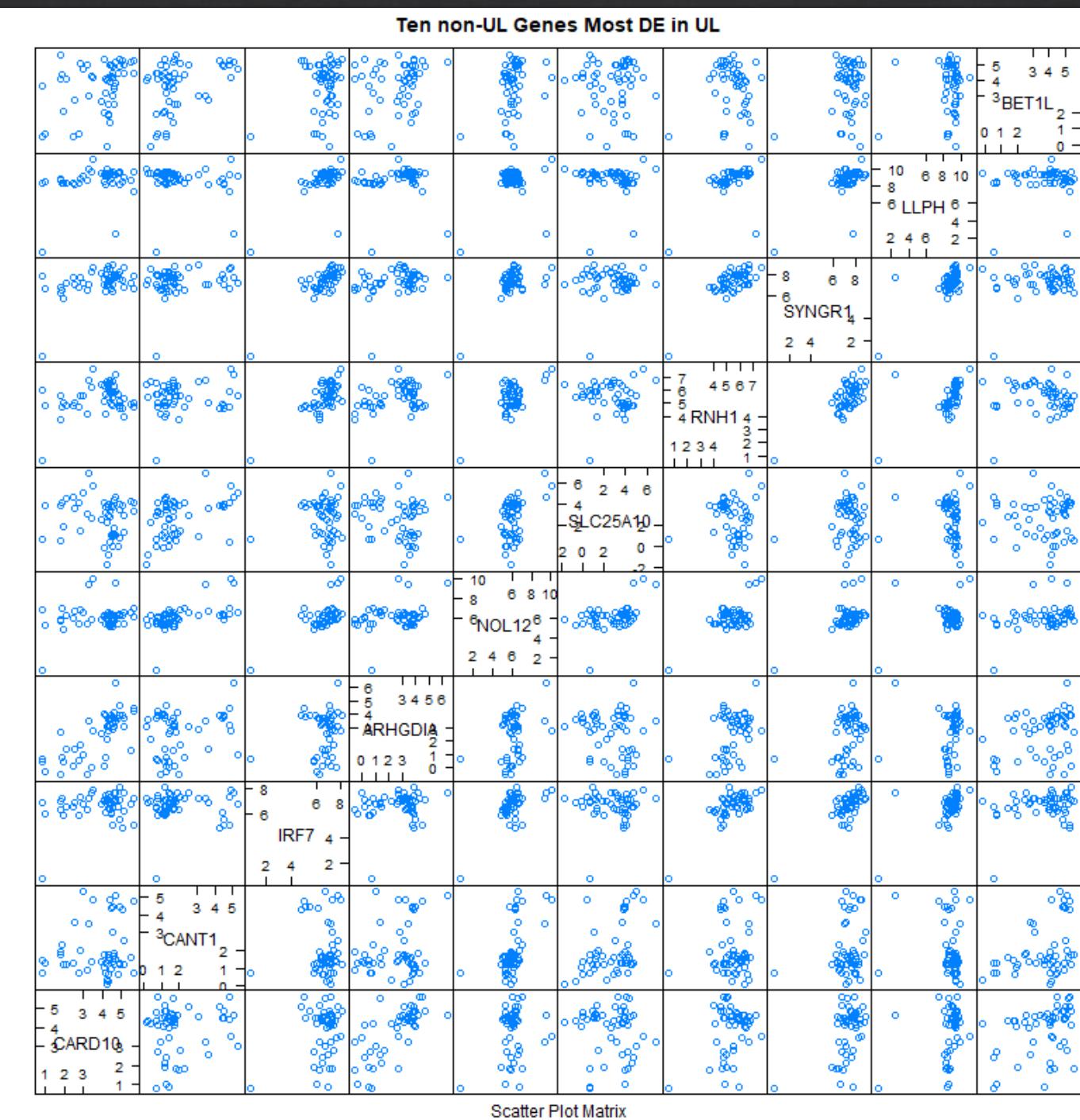


This became 130 genes after removing those not in neighborhood of UL risk genes

Methods: A Lattice Pairwise Comparison of Genes Expressed the Most in UL in Subset of Six UL Risk Genes

- ◊ The R package lattice was used for the splom() on data belonging to those genes having the most change in expression in UL from the subset of genes only along the same cytoband as the six genes ubiquitous to UL risk studies.

Results: A Lattice Pairwise Comparison of Ten Genes



Method: TOP16 Genes Data Preparation

- ❖ In R, the universe of all genes in common were filtered for those genes only along the same cytoband location as the six ubiquitous genes,

Method: TOP16 Genes Data Preparation



UL Means



Non-UL Means



Difference in Means



Magnitude of Difference in Means.

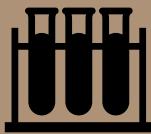
Method: TOP16 Genes Data Preparation



The ten most differentially expressed in magnitude were selected



The six genes ubiquitous to UL risk added



The data table was then prepared for machine learning

Results: TOP16; Top 10 Genes of Highest Magnitude of Change in UL and the 6 Genes Ubiquitous to Current UL Risk Studies

▲	GENE_NAME
ARHGDI	Rho GDP dissociation inhibitor (GDI) alpha
BET1L	blocked early in transport 1 homolog (<i>S. cerevisiae</i>)-like
CANT1	calcium activated nucleotidase 1
CARD10	caspase recruitment domain family, member 10
CCDC57	coiled-coil domain containing 57
CYTH4	cytohesin 4
FASN	fatty acid synthase
FSCN2	fascin homolog 2, actin-bundling protein, retinal (<i>Strongylo...</i>
GRIP1	glutamate receptor interacting protein 1
HMGA2	high mobility group AT-hook 2
IRF7	interferon regulatory factor 7
NOL12	nucleolar protein 12
RNH1	ribonuclease/angiogenin inhibitor 1
SLC25A10	solute carrier family 25 (mitochondrial carrier; dicarboxylate ...
SLC38A10	solute carrier family 38, member 10
TNRC6B	trinucleotide repeat containing 6B

Method: TOP16 Genes with Simulated Means Data Preparation

- ❖ Bootstrap simulated results were done to each of these 16 genes

Results: Table of Simulated Means of TOP16, Differential Expression, and Standard Errors

	GENE	DE	nonUL_means	nonUL_sd	ul_means	ul_sd
1	ARHGDIA	0.67960725	2.641037	0.22302368	3.3206441	0.19550963
2	BET1L	0.50336845	3.728317	0.19786562	4.2316854	0.22174494
3	CANT1	0.74089735	2.039291	0.19571653	2.7801885	0.22669439
4	CARD10	0.79536548	3.907386	0.16439952	4.7027510	0.15190607
5	CCDC57	-0.32291559	5.541802	0.44041553	5.2188861	0.39960207
6	CYTH4	0.34285382	9.132156	0.20108231	9.4750098	0.16203686
7	FASN	-0.08060559	3.568928	0.14433342	3.4883228	0.13786381
8	FSCN2	-0.82095400	1.353323	0.24191907	0.5323693	0.21577234
9	GRIP1	-0.77192382	6.009125	0.25920759	5.2372015	0.17186176
10	HMGAA2	-0.48103471	2.013780	0.25947227	1.5327457	0.19534709
11	IRF7	0.72158037	6.763880	0.10061099	7.4854601	0.11669587
12	NOL12	0.68422306	6.206984	0.14204988	6.8912070	0.15764669
13	RNH1	0.62227484	5.546361	0.12016566	6.1686359	0.11400604
14	SLC25A10	0.64881624	2.757531	0.25160843	3.4063468	0.25704965
15	SLC38A10	-0.65092224	9.155826	0.18822490	8.5049034	0.15824961
16	TNRC6B	0.05655061	7.353408	0.05615986	7.4099588	0.07051176

Method: TOP16 Genes with Simulated Means Histograms Produced Each

- ❖ Histograms of each of the TOP16 Genes Simulated Means was made

Results: Histograms for Each TOP16 Gene's Simulated Means for the Population

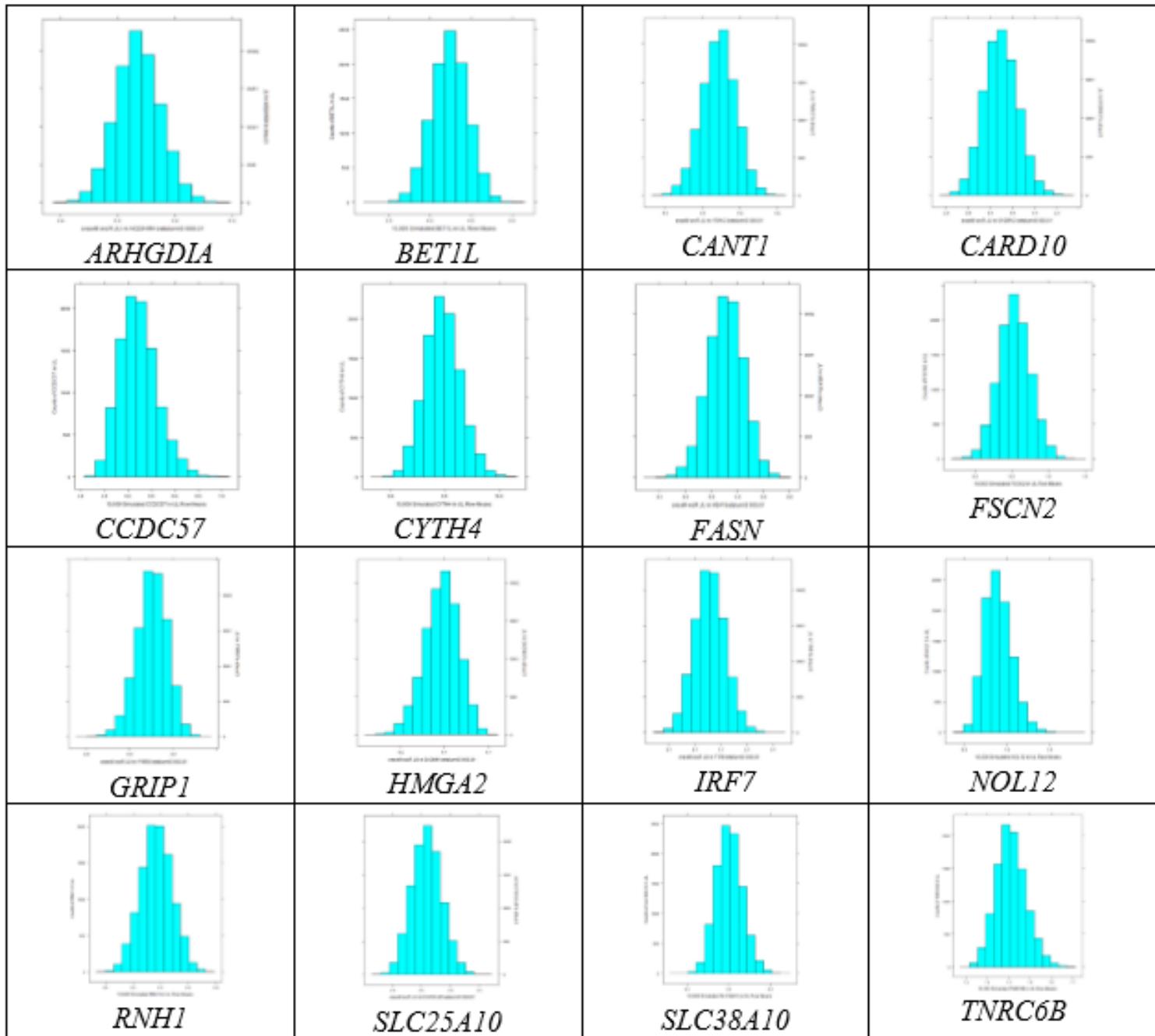
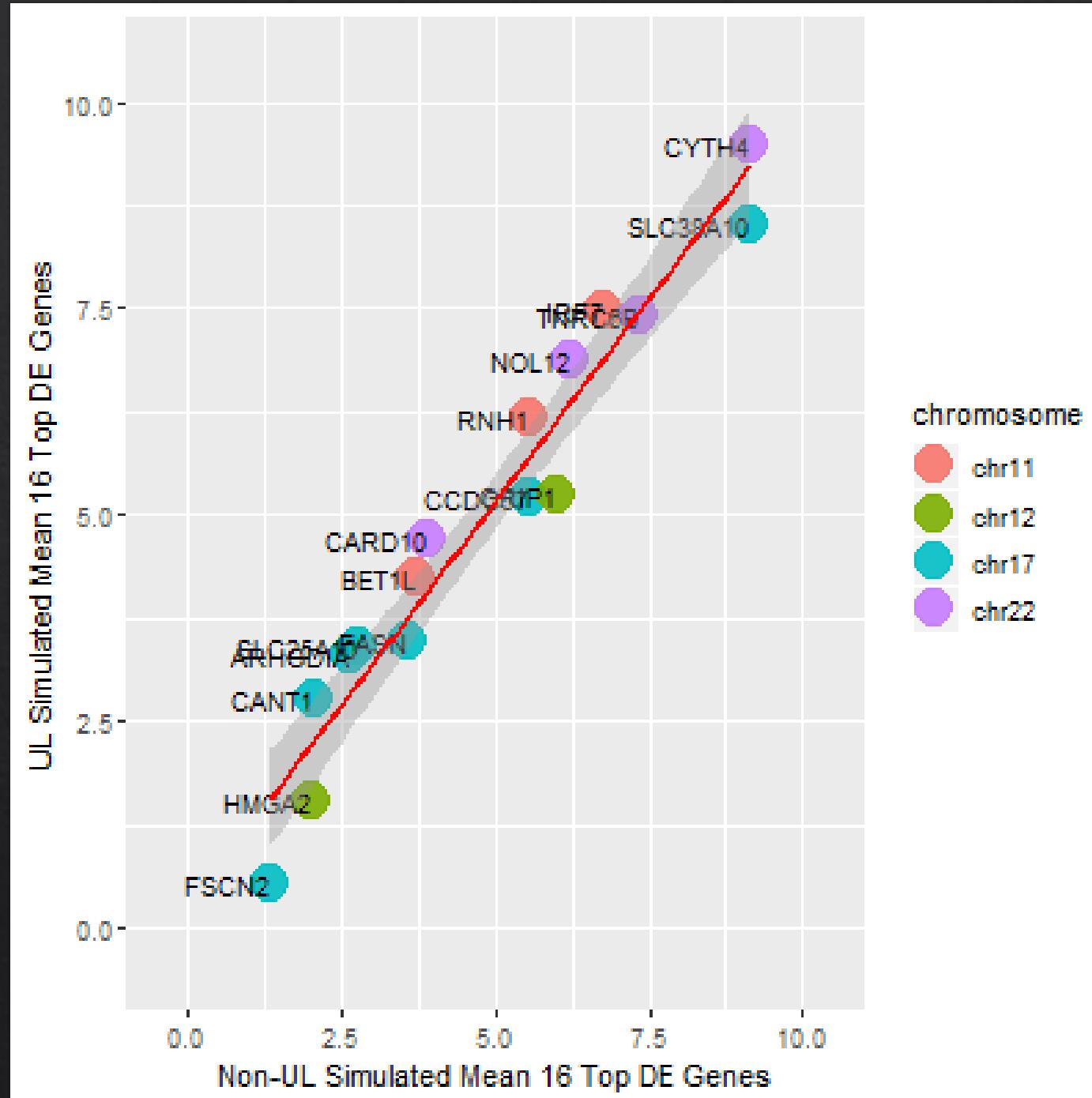


Figure 8. Histogram of TOP16 Simulated Means of 10K Samplings per Gene

- ❖ A visualization would be great to see how these TOP16 genes simulated means compare by chromosomal location.

Method:
Visualize How
the UL and
Non-UL Means
Compare for
TOP16 genes

Results: Plot Using ggplot2 showing TOP16 Genes and the Chromosome Each Lives Comparing Simulated UL and Non-UL Means of Each Gene



Methods – Machine Learning Algorithms

- ❖ The data sets that were derived were then each trained (70 per cent of samples) and tested (30 per cent of samples) using seven machine learning algorithms in R, and a combined model
 - ❖ Latent Dirichlet Allocation method of caret package in R (LDA)
 - ❖ Random Forest method of caret package in R (RF)
 - ❖ Generalized Boosted Regression Models method of caret package in R (GBM)
 - ❖ Random Forest package in R called randomForest (RF2)
 - ❖ K Nearest Neighbor (KNN) method of caret package in R
 - ❖ Recursive Partitioning and Regression Trees (Rpart) from the rpart package in R
 - ❖ Generalized Linear Regression Model (GLM) from the MASS package in R
 - ❖ Combined Model using the ‘gam’ method in caret package of R on all seven algorithms above

Results: Machine Learning TOP16 Outcomes

Table5. Table of the Combined Model

	<i>predRF</i>	<i>predRF2</i>	<i>predIlda</i>	<i>predGbm</i>	<i>predKNN</i>	<i>predRPART</i>	<i>predGLM</i>	<i>CombinedPredictions2</i>	<i>TYPE</i>
<i>gsm1667145</i>	UL	UL	nonUL	UL	UL	UL	nonUL	nonUL	nonUL
<i>gsm336254</i>	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	nonUL	nonUL
<i>gsm336258</i>	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL
<i>gsm336260</i>	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL
<i>gsm336270</i>	nonUL	UL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	nonUL
<i>gsm336273</i>	nonUL	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	nonUL
<i>gsm336276</i>	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL
<i>gsm52662</i>	nonUL	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	nonUL
<i>gsm52663</i>	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL
<i>gsm52665</i>	UL	UL	nonUL	UL	nonUL	UL	nonUL	nonUL	nonUL
<i>gsm52667</i>	UL	UL	nonUL	UL	nonUL	UL	nonUL	nonUL	nonUL
<i>gsm52669</i>	UL	nonUL	nonUL	nonUL	nonUL	UL	nonUL	UL	nonUL
<i>gsm9099</i>	UL	UL	nonUL	UL	nonUL	UL	nonUL	nonUL	nonUL
<i>gsm569425</i>	nonUL	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	nonUL
<i>gsm569427</i>	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL
<i>gsm336202ul</i>	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL
<i>gsm336208ul</i>	UL	UL	UL	UL	UL	UL	UL	UL	UL
<i>gsm336209ul</i>	UL	UL	UL	UL	UL	UL	UL	UL	UL
<i>gsm336214ul</i>	UL	UL	UL	UL	UL	UL	UL	UL	UL
<i>gsm336215ul</i>	UL	UL	UL	UL	UL	UL	UL	UL	UL
<i>gsm336218ul</i>	UL	UL	UL	UL	UL	UL	UL	UL	UL
<i>gsm336220ul</i>	nonUL	nonUL	UL	UL	UL	nonUL	UL	UL	UL
<i>gsm336229ul</i>	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
<i>gsm336232ul</i>	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
<i>gsm336234ul</i>	UL	UL	UL	UL	nonUL	UL	UL	UL	UL
<i>gsm336238ul</i>	UL	UL	nonUL	nonUL	UL	UL	nonUL	UL	UL
<i>gsm336239ul</i>	UL	UL	nonUL	nonUL	UL	UL	nonUL	UL	UL
<i>gsm336240ul</i>	UL	UL	UL	UL	UL	UL	UL	UL	UL
<i>gsm336241ul</i>	nonUL	nonUL	nonUL	UL	nonUL	nonUL	UL	UL	UL
<i>gsm336245ul</i>	nonUL	nonUL	UL	nonUL	UL	nonUL	nonUL	UL	UL
<i>gsm336248ul</i>	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL
<i>gsm38689ul</i>	UL	nonUL	nonUL	UL	nonUL	UL	nonUL	nonUL	UL
<i>gsm38692ul</i>	nonUL	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	UL
<i>gsm9094ul</i>	UL	UL	nonUL	nonUL	nonUL	UL	nonUL	UL	UL
<i>gsm569429ul</i>	UL	UL	nonUL	UL	nonUL	UL	nonUL	nonUL	UL
<i>results</i>	0.69	0.66	0.74	0.69	0.71	0.54	0.74	0.83	100

Methods – Data Sets Derived from 130 Genes Common to Cytobands of UL Risk Studies

- ❖ DE16_most_130_results
 - ❖ Highest magnitude of change in expression/inhibition in UL from cytobands of six genes ubiquitous to UL risk studies
- ❖ Machine learning

Results –Data Table: DE16_most_130_results

	TYPE	FSCN2	CARD10	GRIP1	CANT1	IRF7	ARHGDIA	NOL12	SLC25A10	SLC38A10	RNH1	SYNGR1	TALDO1	FN3K	POLR2F	LLPH	SMCR7L
gsm52670	nonUL	0.14	1.85	3.46	2.29	7.29	0.77	6.99	4.19	6.40	4.70	6.51	5.31	7.16	4.99	8.31	0.85
gsm52671	nonUL	-1.32	3.23	4.76	0.93	6.28	-0.51	5.35	3.15	8.70	4.18	6.51	6.00	7.49	4.28	8.57	1.26
gsm9098	nonUL	1.85	1.81	7.47	1.81	7.03	2.68	5.73	0.58	8.87	5.51	6.27	7.01	7.30	6.01	8.74	2.29
gsm9099	nonUL	5.28	2.98	6.81	5.29	6.78	1.00	6.60	5.07	7.87	5.56	7.18	6.40	7.46	5.30	9.13	2.61
gsm9100	nonUL	-0.51	2.81	7.69	2.04	7.36	1.68	6.83	1.49	7.95	5.51	7.70	6.80	7.97	5.50	8.73	1.38
gsm9101	nonUL	1.43	4.11	5.08	1.77	7.14	0.14	6.10	4.19	8.11	5.16	7.02	6.72	7.37	6.27	8.50	2.41
gsm9102	nonUL	0.14	5.52	5.59	1.63	6.76	1.32	6.57	4.31	8.24	4.93	7.25	6.92	7.73	6.02	8.58	2.17
gsm569424	nonUL	3.45	5.61	5.24	4.78	5.07	4.49	7.04	4.08	8.05	4.73	6.71	6.98	6.17	5.96	7.41	6.08
gsm569425	nonUL	3.29	4.87	5.73	4.46	5.51	3.98	6.27	3.96	7.78	4.58	7.56	7.56	7.47	5.39	8.10	5.17
gsm569426	nonUL	3.32	5.07	4.69	4.28	4.84	3.91	6.29	3.87	7.25	4.59	7.40	6.44	7.56	5.38	8.52	5.07
gsm569427	nonUL	3.26	4.50	4.79	4.46	5.12	3.80	5.82	3.80	8.85	4.80	7.01	7.06	7.84	5.87	8.36	5.01
gsm569428	nonUL	3.23	4.47	4.35	4.39	6.11	3.74	6.40	3.94	7.69	4.88	7.90	6.57	7.87	5.64	8.82	4.65
gsm1667147ul	UL	2.46	2.29	8.34	5.64	8.11	5.21	12.21	5.22	12.54	7.82	8.70	6.16	11.85	10.90	14.03	2.96
gsm1667148ul	UL	2.94	3.78	7.48	7.45	8.39	5.29	12.61	6.34	11.84	6.06	7.72	4.49	11.10	10.55	13.76	3.43
gsm1667149ul	UL	3.14	5.38	5.94	3.14	9.75	6.42	6.78	6.92	12.76	7.91	8.57	6.05	11.50	11.70	3.14	3.95
gsm336202ul	UL	-1.32	4.10	7.31	3.57	7.47	2.86	6.09	-0.32	9.35	6.37	8.17	5.67	5.90	6.30	9.48	1.38
gsm336203ul	UL	-1.74	5.00	6.03	3.86	7.84	3.48	7.07	3.68	8.68	6.18	8.20	5.16	6.09	5.22	9.40	3.71
gsm336204ul	UL	-0.51	5.45	5.02	3.45	8.30	3.41	7.24	4.36	8.26	6.09	8.59	5.30	5.69	5.01	9.77	1.32
gsm336205ul	UL	-0.15	5.44	4.10	5.70	6.98	4.09	7.11	5.06	8.35	6.86	9.02	4.70	5.06	5.03	9.85	0.26
gsm336206ul	UL	-1.74	6.08	3.79	2.20	7.55	4.64	7.72	4.36	8.47	6.70	8.98	5.39	5.29	5.36	9.85	0.93
gsm336207ul	UL	-0.74	5.56	4.37	1.38	7.42	3.46	6.69	4.04	8.90	5.94	8.41	5.22	5.19	4.68	9.82	1.96
gsm336208ul	UL	0.77	5.68	4.64	5.70	8.66	4.03	7.70	5.20	7.71	6.53	8.86	5.69	5.75	5.17	9.82	0.77
gsm336209ul	UL	2.00	5.59	6.01	4.49	8.18	3.60	7.58	5.48	8.40	6.50	8.74	5.69	5.50	5.27	9.49	2.14
gsm336210ul	UL	-1.74	5.88	4.71	5.48	8.53	4.10	7.57	4.58	8.15	6.26	8.69	5.44	5.71	5.30	9.95	0.85

	predRF	predRF2	predIlda	predGbm	predKNN	predRPART	predGLM	CombinedPredictions2	TYPE
gsm569428	UL	UL	UL	UL	nonUL	UL	UL	nonUL	nonUL
gsm1667147ul	nonUL	nonUL	UL	nonUL	UL	nonUL	nonUL	UL	UL
gsm1667148ul	nonUL	nonUL	UL	UL	UL	nonUL	nonUL	UL	UL
gsm336202ul	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL
gsm336203ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336205ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336209ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336216ul	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
gsm336217ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336220ul	nonUL	nonUL	UL	nonUL	UL	nonUL	UL	UL	UL
gsm336225ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336226ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336227ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336232ul	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
gsm336245ul	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	UL	UL
gsm336246ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336251ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm38695ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm9094ul	UL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	UL	UL
gsm9095ul	UL	UL	nonUL	UL	nonUL	nonUL	nonUL	nonUL	UL
gsm9097ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm569430ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
results	0.69	0.67	0.78	0.67	0.92	0.61	0.72	0.92	100

Results –Machine Learning results:
DE16_most_130_results

Methods – Overview of Data Sets Derived from 130 Genes Common to Cytobands of UL Risk Studies

- ❖ DE16_least_130_results
 - ❖ Use as a limit of machine learning algorithms to see if the genes expressed/inhibited **the least** in UL also do well in predicting UL with the machine learning algorithms used
- ❖ Machine learning

Results –Data Table:

DE16_least_130_results

	TYPE	PSMD13	DNAL4	ATHL1	CDHRS	SIGIRR	MRPL23	FOXK2	RASSF7	SIRT3	GRAP2	AATK	TH	TMEM184B	WDR45L	KDELR3	GCGR
gsm1667144	nonUL	9.72	16.90	7.15	10.96	5.30	9.87	13.29	6.30	15.26	11.30	9.50	6.18	10.63	8.39	8.13	9.16
gsm1667145	nonUL	3.74	16.41	6.71	9.97	7.05	11.30	12.24	5.03	15.74	7.48	9.74	6.52	10.02	7.98	8.04	8.36
gsm1667146	nonUL	11.23	16.24	8.33	9.66	5.60	11.14	13.50	4.49	15.78	10.72	9.08	7.12	10.08	7.83	9.28	7.96
gsm336252	nonUL	8.67	12.84	4.88	8.19	3.69	9.91	7.70	5.10	12.48	5.79	6.12	4.67	4.61	5.08	5.79	6.48
gsm336253	nonUL	6.98	13.18	6.02	8.40	3.34	10.27	7.74	4.77	12.38	6.00	6.35	1.93	4.89	4.80	6.20	6.45
gsm336254	nonUL	7.22	12.92	5.56	7.87	2.47	10.08	7.60	4.87	12.36	5.75	6.13	4.52	4.13	5.29	6.26	6.10
gsm336255	nonUL	7.42	13.21	6.20	8.18	4.22	10.27	7.83	6.03	12.43	6.22	6.31	3.64	5.10	5.08	6.32	6.96
gsm336256	nonUL	7.69	13.33	6.05	8.38	3.77	10.23	8.29	6.19	12.75	6.78	6.21	1.85	5.39	5.90	6.42	6.92
gsm336257	nonUL	7.72	13.18	6.33	8.45	5.17	10.22	7.91	4.64	12.55	6.71	6.47	3.38	6.48	5.07	6.50	6.57
gsm336258	nonUL	7.47	12.95	5.75	8.22	3.04	10.16	8.40	5.58	12.32	6.66	6.27	3.71	5.49	5.63	6.61	7.09
gsm336259	nonUL	7.62	13.03	6.57	8.46	3.95	10.07	8.24	5.50	12.69	6.46	6.27	2.04	5.24	5.58	6.75	7.03
gsm336260	nonUL	6.90	13.08	6.16	8.35	3.44	10.58	6.98	4.99	12.23	5.74	6.22	1.26	4.87	4.59	6.17	6.13
gsm336261	nonUL	7.75	13.17	6.10	8.81	3.42	10.15	7.97	5.98	12.88	6.47	5.75	1.58	5.21	5.51	6.20	6.55
gsm336262	nonUL	7.88	13.14	6.53	8.67	2.19	10.23	8.16	6.40	12.58	6.90	6.65	2.17	5.80	5.45	6.85	6.95
gsm336263	nonUL	7.46	12.94	6.13	8.16	2.49	10.21	8.07	5.94	12.17	6.62	6.32	4.04	5.48	4.94	6.89	6.96
gsm336264	nonUL	8.05	12.97	6.40	8.32	3.68	10.31	7.99	6.07	12.58	6.62	6.39	2.51	5.93	5.21	6.22	6.84
gsm336265	nonUL	7.60	13.25	6.47	7.91	4.06	10.24	8.00	5.82	12.18	6.08	6.22	4.18	4.11	5.02	6.18	6.78
gsm336266	nonUL	8.08	13.23	5.77	8.13	3.98	10.12	8.21	5.31	12.75	6.32	6.44	2.79	4.31	5.21	6.38	6.67
gsm336267	nonUL	7.62	13.04	5.86	7.90	3.58	10.22	7.97	6.11	11.70	6.27	6.45	2.17	5.51	4.56	6.60	7.13
gsm336268	nonUL	7.68	13.04	6.22	8.05	4.36	10.07	7.85	5.09	12.36	6.39	6.16	4.09	5.66	6.05	6.79	7.01
gsm336269	nonUL	8.84	13.36	6.85	8.25	3.29	10.31	7.68	5.65	12.68	6.44	6.59	2.17	5.45	5.06	6.61	7.17
gsm336270	nonUL	6.19	12.92	6.00	8.15	2.85	10.35	7.58	5.52	11.85	4.95	6.77	3.04	3.81	4.72	6.57	6.00
gsm336271	nonUL	5.53	13.38	6.44	8.34	2.94	10.56	7.07	6.18	12.38	5.33	6.50	2.46	4.54	4.83	6.43	6.15
gsm336272	nonUL	6.48	12.45	5.97	8.46	3.52	9.82	7.46	6.03	11.93	4.34	6.22	3.92	2.29	4.29	6.39	6.62
gsm336273	nonUL	5.50	12.83	5.75	7.95	1.79	10.15	7.15	5.77	11.28	4.97	6.57	2.91	4.13	4.93	6.36	6.04

Results –Machine Learning DE16_least_130_results

	predRF	predRF2	predIlda	predGbm	predKNN	predRPART	predGLM	CombinedPredictions2	TYPE
gsm569428	UL	UL	nonUL	nonUL	UL	UL	nonUL	UL	nonUL
gsm1667147ul	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	UL	UL
gsm1667148ul	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	UL	UL
gsm336202ul	UL	UL	nonUL	nonUL	UL	nonUL	nonUL	UL	UL
gsm336203ul	nonUL	nonUL	UL	nonUL	UL	nonUL	UL	UL	UL
gsm336205ul	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
gsm336209ul	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
gsm336216ul	UL	UL	nonUL	nonUL	UL	UL	nonUL	UL	UL
gsm336217ul	nonUL	nonUL	UL	UL	UL	UL	UL	UL	UL
gsm336220ul	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
gsm336225ul	UL	UL	nonUL	UL	UL	UL	nonUL	UL	UL
gsm336226ul	UL	UL	UL	UL	UL	nonUL	UL	nonUL	UL
gsm336227ul	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
gsm336232ul	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
gsm336245ul	nonUL	nonUL	nonUL	nonUL	UL	nonUL	nonUL	UL	UL
gsm336246ul	UL	UL	UL	UL	nonUL	UL	UL	UL	UL
gsm336251ul	UL	UL	nonUL	nonUL	UL	UL	nonUL	UL	UL
gsm38695ul	UL	UL	UL	nonUL	UL	UL	UL	nonUL	UL
gsm9094ul	nonUL	nonUL	nonUL	nonUL	nonUL	UL	nonUL	UL	UL
gsm9095ul	UL	UL	nonUL	UL	nonUL	UL	nonUL	UL	UL
gsm9097ul	UL	UL	nonUL	nonUL	nonUL	UL	nonUL	UL	UL
gsm569430ul	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
results	0.47	0.53	0.42	0.47	0.5	0.58	0.33	0.72	100

Methods – Overview of Data Sets Derived from 130 Genes Common to Cytobands of UL Risk Studies

- ❖ FOLD16_130_results
 - ❖ Fold change added to get the 16 genes with the most fold change in the cytobands of the six ubiquitous genes, including those six genes
- ❖ Machine learning

Results – Data Table: FOLD16_130_results

Fold16_130 x Filter

	TYPE	CANT1	ARHGDIA	SLC25A10	ATF4	CARD10	MRPL12	BET1L	RNH1	NOL12	IRF7	DRD4	TNRC6B	CYTH4	CCDC57	FASN	HMGA2
gsm52664	nonUL	1.20	0.49	1.93	-1.32	1.93	-1.32	2.04	5.01	4.87	5.84	1.32	7.15	7.55	4.29	2.88	1.63
gsm52665	nonUL	3.60	1.38	3.66	-0.51	2.54	3.75	2.81	5.00	6.68	6.93	3.22	7.47	7.00	5.17	3.96	3.55
gsm52666	nonUL	1.20	1.43	4.56	-1.74	2.10	1.38	4.09	4.81	6.24	6.54	3.12	7.22	7.59	4.55	2.39	2.87
gsm52667	nonUL	1.43	0.49	0.68	-0.51	0.85	3.46	3.67	5.97	6.84	6.92	0.00	6.57	8.57	3.90	3.17	1.85
gsm52668	nonUL	1.26	0.77	3.70	-1.32	4.19	0.14	3.19	4.13	5.46	6.05	-1.00	6.80	8.17	4.02	3.08	0.49
gsm52669	nonUL	1.26	0.14	4.70	2.00	2.43	3.29	0.49	4.78	5.91	6.52	3.58	7.07	8.24	4.15	3.59	0.26
gsm52670	nonUL	2.29	0.77	4.19	-3.32	1.85	-0.51	4.36	4.70	6.99	7.29	0.26	6.83	7.44	3.39	2.86	0.77
gsm52671	nonUL	0.93	-0.51	3.15	-1.32	3.23	2.61	4.94	4.18	5.35	6.28	2.54	6.75	7.71	3.79	2.97	0.49
gsm9098	nonUL	1.81	2.68	0.58	-0.74	1.81	0.77	5.59	5.51	5.73	7.03	1.14	7.55	7.90	2.29	3.62	1.58
gsm9099	nonUL	5.29	1.00	5.07	2.26	2.98	0.93	4.68	5.56	6.60	6.78	3.70	7.23	8.11	2.79	2.84	2.49
gsm9100	nonUL	2.04	1.68	1.49	-2.32	2.81	1.20	5.11	5.51	6.83	7.36	2.49	7.00	8.67	4.01	4.30	2.32
gsm9101	nonUL	1.77	0.14	4.19	-1.00	4.11	0.00	5.36	5.16	6.10	7.14	1.20	7.24	8.26	3.81	2.91	3.19
gsm9102	nonUL	1.63	1.32	4.31	-2.32	5.52	-1.00	5.17	4.93	6.57	6.76	1.58	6.99	8.01	3.38	4.72	1.43
gsm569424	nonUL	4.78	4.49	4.08	3.46	5.61	4.15	5.25	4.73	7.04	5.07	5.53	8.16	6.60	4.22	6.19	5.23
gsm569425	nonUL	4.46	3.98	3.96	3.36	4.87	4.01	4.87	4.58	6.27	5.51	5.04	7.33	7.83	3.94	5.44	4.41
gsm569426	nonUL	4.28	3.91	3.87	3.42	5.07	3.95	5.28	4.59	6.29	4.84	4.34	7.51	8.37	4.02	5.55	4.61
gsm569427	nonUL	4.46	3.80	3.80	3.45	4.50	3.88	4.92	4.80	5.82	5.12	4.49	7.10	8.15	3.82	5.29	4.56
gsm569428	nonUL	4.39	3.74	3.94	3.50	4.47	3.94	4.98	4.88	6.40	6.11	4.26	6.98	7.93	4.09	5.04	4.34
gsm1667147ul	UL	5.64	5.21	5.22	6.74	2.29	4.26	7.52	7.82	12.21	8.11	8.33	8.73	12.55	17.49	3.14	5.74
gsm1667148ul	UL	7.45	5.29	6.34	6.22	3.78	4.05	6.60	6.06	12.61	8.39	8.44	8.23	13.48	17.71	4.16	4.43
gsm1667149ul	UL	3.14	6.42	6.92	7.89	5.38	7.98	6.90	7.91	6.78	9.75	9.32	10.00	13.43	17.51	6.21	7.14
gsm336202ul	UL	3.57	2.86	-0.32	2.97	4.10	-0.15	4.60	6.37	6.09	7.47	4.92	7.21	10.00	5.50	3.26	1.14
gsm336203ul	UL	3.86	3.48	3.68	3.89	5.00	3.54	5.22	6.18	7.07	7.84	4.40	7.32	9.67	5.09	3.28	1.72
gsm336204ul	UL	3.45	3.41	4.36	2.95	5.45	3.88	4.24	6.09	7.24	8.30	5.16	7.50	9.80	6.37	3.41	1.14
gsm336205ul	UL	5.70	4.20	5.20	3.24	5.11	4.10	4.20	5.00	7.44	6.00	4.70	7.10	9.10	6.60	4.20	2.20

Showing 35 to 62 of 121 entries, 17 total columns

Results –Machine Learning FOLD16_130_results

	predRF	predRF2	predIlda	predGbm	predKNN	predRPART	predGLM	CombinedPredictions2	TYPE
gsm569428	UL	UL	UL	nonUL	nonUL	nonUL	UL	nonUL	nonUL
gsm1667147ul	nonUL	nonUL	nonUL	nonUL	UL	nonUL	UL	UL	UL
gsm1667148ul	nonUL	nonUL	nonUL	nonUL	UL	nonUL	UL	UL	UL
gsm336202ul	nonUL	nonUL	nonUL	UL	nonUL	nonUL	nonUL	UL	UL
gsm336203ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336205ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336209ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336216ul	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
gsm336217ul	UL	UL	nonUL	UL	nonUL	UL	nonUL	nonUL	UL
gsm336220ul	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
gsm336225ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336226ul	UL	UL	UL	UL	UL	UL	nonUL	nonUL	UL
gsm336227ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336232ul	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
gsm336245ul	nonUL	nonUL	UL	nonUL	nonUL	nonUL	UL	nonUL	UL
gsm336246ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm336251ul	UL	UL	UL	UL	UL	UL	UL	UL	UL
gsm38695ul	UL	UL	nonUL	UL	nonUL	nonUL	UL	UL	UL
gsm9094ul	nonUL	UL	UL	nonUL	nonUL	nonUL	UL	UL	UL
gsm9095ul	nonUL	nonUL	UL	nonUL	nonUL	nonUL	UL	nonUL	UL
gsm9097ul	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
gsm569430ul	nonUL	nonUL	nonUL	nonUL	UL	UL	nonUL	nonUL	UL
results	0.64	0.69	0.67	0.72	0.69	0.58	0.72	0.78	100

Methods – Data Sets Derived from the Universe of All Genes in Common

- ❖ The data sets that were derived to test the best machine learning results for UL prediction on the 12,173 universe of genes in common (none include the six UL risk genes):
 - ❖ universe16_fold_results
 - ❖ universe16_DE_most_results
 - ❖ universe16_DE_least_results

Methods – Data Sets Derived from the Universe of All Genes in Common

- ❖ universe16_fold_results

Results –Data Table: universe16_fold_results

	TYPE	HSPB1	DSTN	S100A6	CNN1	ACTG2	VIM	SPARCL1	TPM2	ACTA2	PCP4	TAGLN	DES	RAMP1	CYR61	UBC	ACTB
GSM52667	nonUL	9.6	6.85	10.6	40.60	14.45	3.966667	20.00000	58.5	47.2	3.4	2.9	2.7	9.5	1.500000	10.666667	6.40
GSM52668	nonUL	10.4	6.10	19.4	23.00	18.85	17.433333	37.00000	73.1	36.6	1.6	10.3	3.8	27.4	10.200000	9.100000	11.40
GSM52669	nonUL	15.4	1.90	23.8	19.20	14.50	49.333333	42.30000	35.8	34.8	3.3	2.3	3.8	21.1	224.500000	12.066667	27.20
GSM52670	nonUL	29.9	15.35	23.4	14.50	49.30	51.166667	13.70000	49.3	26.3	10.4	1.7	2.2	16.5	73.000000	4.600000	16.00
GSM52671	nonUL	17.7	16.35	10.6	14.30	6.40	31.400000	120.90000	58.9	15.8	1.8	1.8	31.8	10.9	3.600000	15.666667	50.20
GSM9098	nonUL	9.2	5.45	35.9	18.60	25.60	9.000000	29.60000	54.6	39.0	2.4	10.5	8.6	32.8	3.200000	10.833333	46.00
GSM9099	nonUL	18.6	9.80	7.4	4.20	24.40	100.500000	58.20000	36.7	6.0	21.1	3.1	7.3	56.5	209.100000	27.300000	24.00
GSM9100	nonUL	29.1	17.20	32.9	12.70	187.25	245.066667	196.00000	229.5	33.1	72.9	3.0	119.6	36.8	61.500000	7.900000	38.80
GSM9101	nonUL	7.3	3.65	43.7	23.70	22.00	15.233333	51.80000	50.8	60.7	3.8	2.6	28.0	27.2	3.300000	25.033333	38.20
GSM9102	nonUL	14.2	5.40	14.7	20.80	14.10	7.266667	24.50000	34.5	27.0	2.0	2.2	4.8	25.2	1.000000	7.166667	26.10
GSM569424	nonUL	21.0	23.55	17.5	27.50	30.45	20.700000	29.20000	25.4	42.2	43.1	29.0	44.0	55.3	34.600000	22.000000	21.60
GSM569425	nonUL	14.5	16.85	15.1	19.70	21.85	19.633333	25.50000	87.2	18.2	47.5	20.7	35.5	31.4	27.800000	18.233333	17.70
GSM569426	nonUL	17.7	36.35	18.8	21.60	1011.50	343.866667	22.60000	28.3	244.2	37.0	25.3	27.2	34.0	626.000000	22.000000	17.90
GSM569427	nonUL	14.8	13.70	14.3	20.30	23.75	16.933333	21.30000	30.6	57.9	32.2	15.6	30.2	36.9	18.900000	18.433333	16.10
GSM569428	nonUL	17.8	13.70	14.9	19.90	63.10	54.500000	21.70000	43.9	19.2	42.7	16.6	22.6	34.0	87.300000	14.900000	16.00
GSM1667147UL	UL	55.5	853.95	271.1	5.40	1441.30	10442.800000	8369.73333	4040.3	1219.9	3858.8	323.9	94.7	5.0	1221.100000	24.850000	89.10
GSM1667148UL	UL	17.5	813.85	3098.5	5.95	12735.20	37059.400000	2648.46667	9134.7	3621.9	984.6	139.1	9.4	10.2	2579.433333	109.400000	826.60
GSM1667149UL	UL	87675.1	18548.70	13167.2	8.80	8.80	81.800000	1512.66667	8.8	743.5	542.5	667.9	126.0	8.8	950.866667	8.800000	8.80
GSM336202UL	UL	2.9	4.85	29.3	18.80	21.95	35.666667	10.80000	68.2	38.7	11.6	1.7	13.7	6.3	89.900000	7.566667	18.05
GSM336203UL	UL	1.1	73.20	13.6	6.00	146.05	316.300000	204.60000	344.8	37.4	138.3	44.7	28.0	72.3	367.366667	15.133333	62.50
GSM336204UL	UL	18.4	128.75	249.8	3.30	763.05	1164.533333	445.90000	812.4	49.6	331.8	97.7	235.6	1020.1	389.933333	10.833333	41.80
GSM336205UL	UL	19.3	82.30	87.4	6.00	1676.70	411.133333	281.80000	3274.3	197.4	222.9	64.2	15.1	51.8	311.133333	29.833333	54.65
GSM336206UL	UL	21.8	31.65	92.7	8.80	884.35	909.300000	327.95000	1913.8	251.7	549.4	140.8	5.0	11.9	526.300000	9.233333	81.45
GSM336207UL	UL	1.3	86.40	840.8	13.10	565.10	516.900000	254.00000	2743.9	221.1	195.0	62.4	2.7	41.4	305.100000	18.100000	73.90

Showing 39 to 66 of 121 entries, 17 total columns

Results - Machine Learning: universe16_fol d_results

	predRF	predRF2	predIda	predGbm	predKNN	predRPART	predGLM	CombinedPredictions2	TYPE
GSM569428	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL
GSM1667147UL	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
GSM1667148UL	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
GSM336202UL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL
GSM336203UL	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
GSM336205UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336209UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336216UL	UL	UL	nonUL	UL	nonUL	UL	UL	UL	UL
GSM336217UL	nonUL	nonUL	UL	nonUL	UL	nonUL	UL	UL	UL
GSM336220UL	UL	UL	nonUL	UL	nonUL	UL	UL	UL	UL
GSM336225UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336226UL	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
GSM336227UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336232UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336245UL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL
GSM336246UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336251UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM38695UL	nonUL	nonUL	UL	nonUL	UL	nonUL	UL	UL	UL
GSM9094UL	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
GSM9095UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM9097UL	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
GSM569430UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
results	0.86	0.86	0.86	0.86	0.86	0.86	0.72	0.92	100

Methods – Data Sets Derived from the Universe of All Genes in Common

- ❖ universe16_DE_most_results

Results – Data Table: universe16_DE_most_results

universe16_DE_most_results

Filter

	TYPE	HSPB1	DSTN	S100A6	CNN1	ACTG2	VIM	SPARCL1	TPM2	ACTA2	PCP4	TAGLN	DES	RAMP1	CYR61	UBC	ACTB
GSM1667144	nonUL	140415.20	138391.900	137714.40	321295.2	210342.40	131806.00	141254.8	171722.900	200859.900	12035.6	146782.600	175906.60	154054.4	24100.867	177613.08	153537.775
GSM1667145	nonUL	116491.67	136927.200	74720.70	284928.4	274898.60	132901.45	135557.9	138482.150	210342.400	18874.2	178790.850	102849.80	143577.5	101269.800	141931.85	107229.825
GSM1667146	nonUL	202443.03	134994.400	90894.20	250314.1	342010.20	103499.45	172181.4	142301.250	186641.500	7183.6	129726.950	127199.20	180022.8	17953.367	114247.57	101239.575
GSM336252	nonUL	8104.50	4686.867	3546.75	7410.8	4806.25	3097.25	7621.5	1728.800	3128.667	1035.2	6633.933	3240.70	2831.0	5008.850	6814.70	8891.033
GSM336253	nonUL	7473.60	4529.133	2838.80	8247.0	5377.85	3690.45	8458.8	1557.950	3772.467	1340.7	7476.600	2466.10	1946.0	3107.850	7600.90	10473.283
GSM336254	nonUL	8198.30	5661.733	2737.95	7237.3	5258.55	3407.75	8294.1	1506.375	3844.033	2215.0	6682.233	3413.45	2241.3	5991.300	6661.55	8766.683
GSM336255	nonUL	8035.60	4614.167	2695.00	9162.4	5497.85	4024.05	9235.7	1610.800	4285.333	2960.7	8085.300	3298.35	1905.8	5454.600	8639.90	11175.600
GSM336256	nonUL	10245.60	5589.200	2473.30	11289.9	6920.20	4156.70	9879.5	1982.200	4507.467	1448.8	9752.867	5200.70	2984.3	6240.200	9636.95	13330.450
GSM336257	nonUL	12170.90	5455.233	4053.45	10963.2	8209.15	3954.85	11058.0	2172.525	4744.967	2761.6	9771.933	4529.85	1803.5	4579.000	8630.50	12937.500
GSM336258	nonUL	5050.40	3737.967	2970.05	7779.7	3705.35	3458.55	8450.0	1405.775	3267.600	1711.1	6061.200	2517.10	1604.5	5315.800	7435.15	8606.700
GSM336259	nonUL	10227.60	4989.600	4566.00	9363.8	5530.80	3615.10	8536.4	1923.700	3385.333	755.9	8388.333	4206.60	3247.6	1462.250	8210.55	10383.467
GSM336260	nonUL	8253.20	5179.967	2250.95	9470.1	5909.10	3240.95	10493.5	1300.400	4648.267	3720.6	7394.200	3404.85	1338.6	2936.900	8236.20	11319.717
GSM336261	nonUL	10915.50	5233.433	5299.45	10347.2	6298.70	3937.60	9672.6	2223.725	3809.433	980.7	9519.133	4276.50	3801.5	1349.050	8060.15	11461.683
GSM336262	nonUL	6533.40	3552.133	2582.80	6491.1	2849.40	3843.20	7466.3	1345.850	3454.433	2238.6	6150.233	1518.85	2459.9	945.850	7193.20	8715.050
GSM336263	nonUL	4528.80	3680.767	1394.80	3975.7	1840.65	3520.75	7406.9	1031.100	3170.933	1675.2	4679.067	1105.90	1581.9	5250.150	7309.35	8540.100
GSM336264	nonUL	6758.80	3935.167	3315.50	6935.7	3101.60	4116.65	7746.8	1340.525	3503.667	1809.4	6900.033	2352.95	1858.0	4817.950	6863.20	9044.133
GSM336265	nonUL	7195.00	4392.467	2114.15	6579.2	5131.85	4255.70	8832.2	1683.775	3954.700	648.0	6429.933	2085.55	1391.6	3251.650	7407.40	10642.083
GSM336266	nonUL	9961.00	4845.967	2579.80	8875.3	5836.25	4494.65	9152.9	1796.525	3986.200	1058.9	8111.133	2929.40	1942.9	4485.700	8348.05	11477.833
GSM336267	nonUL	3368.40	3937.600	977.90	3536.3	1567.85	3710.85	7573.9	802.425	3040.833	423.3	3854.933	657.65	949.7	5945.550	8119.05	8742.150
GSM336268	nonUL	8292.10	5043.467	3538.20	8862.2	5551.90	4323.35	8661.6	1924.075	3582.300	1228.6	7718.833	4363.20	1620.5	5119.850	8784.60	11338.000
GSM336269	nonUL	10828.20	5312.033	2968.50	9876.2	6746.75	4500.30	10829.6	2533.100	4800.933	1556.4	9372.167	2865.05	2562.6	1913.250	7876.35	12341.700
GSM336270	nonUL	4587.10	4820.167	1875.05	6364.2	5052.55	3329.40	9514.4	1107.425	4897.467	3933.0	6773.133	2126.70	905.3	7432.550	7593.65	11111.000
GSM336271	nonUL	6210.50	5307.933	1252.45	6798.2	6528.40	4129.95	10375.4	1244.875	4988.867	1795.0	7386.867	1197.95	552.7	2287.100	7617.20	11990.000
GSM336272	nonUL	5128.00	4427.567	2497.65	6838.0	4740.60	2415.30	7001.3	1862.250	3258.300	3825.6	6047.333	3195.70	993.9	2692.650	6293.80	7687.333
GSM336273	nonUL	3595.30	4792.600	2397.15	5405.6	3294.20	3093.65	7432.0	873.275	3785.033	2912.8	4449.500	1675.25	1000.8	581.600	6876.20	8870.383

Showing 1 to 27 of 121 entries. 17 total columns.

Results - Machine Learning: universe16_DE_ most_results

	predRF	predRF2	predIlda	predGbm	predKNN	predRPART	predGLM	CombinedPredictions2	TYPE
GSM569428	UL	UL	nonUL	nonUL	nonUL	UL	UL	nonUL	nonUL
GSM1667147UL	nonUL	nonUL	UL	nonUL	nonUL	nonUL	UL	UL	UL
GSM1667148UL	nonUL	nonUL	UL	nonUL	nonUL	nonUL	UL	UL	UL
GSM336202UL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL
GSM336203UL	UL	nonUL	UL	nonUL	UL	nonUL	UL	UL	UL
GSM336205UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336209UL	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
GSM336216UL	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
GSM336217UL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL
GSM336220UL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL	UL	UL
GSM336225UL	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
GSM336226UL	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
GSM336227UL	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
GSM336232UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336245UL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	nonUL	UL
GSM336246UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM336251UL	UL	UL	UL	UL	UL	nonUL	UL	UL	UL
GSM38695UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM9094UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM9095UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM9097UL	UL	UL	UL	UL	UL	UL	UL	UL	UL
GSM569430UL	UL	UL	UL	UL	UL	UL	nonUL	UL	UL
results	0.67	0.64	0.81	0.69	0.72	0.53	0.78	0.86	100

Methods – Data Sets Derived from the Universe of All Genes in Common

- ❖ universe16_DE_least_results

Results – Data Table: universe16_DE_least_results

universe16_DE_least_results

Filter

	TYPE	USP32P2	RCVRN	SYNGR3	MORC1	KLK2	SUV39H1	LIG4	KLHDC4	GRIK4	FABP1	TLX3	LAMB4	DNTT	VN1R1	LEFTY1	C7orf64
GSM1667144	nonUL	111.30	10.2	52.6	13.8	18.333333	104.7	28.80000	169.000	5.10	6.10	10.2	10.750000	6.100000	15.0	47.0	129.40000
GSM1667145	nonUL	5.70	15.0	162.6	77.2	45.233333	50.8	99.06667	29.000	5.70	5.70	5.7	16.700000	6.900000	5.7	8.8	117.40000
GSM1667146	nonUL	59.10	14.6	121.6	113.9	52.766667	31.9	57.00000	28.000	5.55	4.90	6.5	19.300000	5.600000	5.3	21.3	133.50000
GSM336252	nonUL	372.85	6.3	2.7	0.5	6.875000	34.2	61.65000	48.525	0.80	5.55	11.9	2.5666667	5.933333	16.0	14.4	35.27500
GSM336253	nonUL	242.45	3.1	26.2	13.6	7.100000	34.4	72.30000	53.175	1.80	0.70	2.1	4.5666667	13.033333	25.7	17.5	30.57500
GSM336254	nonUL	309.65	8.3	2.6	1.2	9.100000	31.4	92.70000	40.500	0.60	2.25	0.8	7.8333333	10.933333	8.9	12.5	38.60000
GSM336255	nonUL	451.80	10.3	2.0	0.9	7.900000	43.6	66.35000	45.825	12.00	2.15	1.7	6.2333333	8.033333	8.9	24.0	49.05000
GSM336256	nonUL	358.60	5.2	1.8	0.4	8.600000	43.9	66.70000	47.575	2.90	2.50	1.5	1.4333333	6.333333	25.8	10.8	32.37500
GSM336257	nonUL	230.00	14.0	8.4	1.9	18.700000	48.2	48.25000	44.450	13.10	9.50	2.3	17.5666667	21.200000	31.9	71.9	34.65000
GSM336258	nonUL	159.20	3.5	1.7	1.9	2.500000	39.5	60.25000	67.275	1.70	1.50	1.6	5.8666667	5.633333	9.1	12.0	47.87500
GSM336259	nonUL	227.00	22.2	1.6	0.4	7.825000	11.7	56.25000	63.525	4.90	1.15	10.0	6.9666667	3.666667	36.1	15.5	37.97500
GSM336260	nonUL	306.30	15.1	1.7	0.8	7.500000	35.4	73.40000	40.350	1.30	1.10	0.9	4.2000000	5.533333	15.4	67.6	60.35000
GSM336261	nonUL	194.70	17.6	5.2	1.2	10.375000	43.6	54.65000	47.750	1.00	2.30	0.5	6.0000000	5.933333	21.3	20.6	50.55000
GSM336262	nonUL	273.25	11.2	2.2	2.9	7.675000	42.9	58.70000	75.900	1.70	2.65	2.5	8.3666667	9.300000	16.9	24.7	39.07500
GSM336263	nonUL	386.00	15.9	13.6	0.7	5.325000	49.9	63.70000	75.725	1.40	6.60	2.8	4.8666667	5.400000	17.9	19.4	55.15000
GSM336264	nonUL	269.00	16.9	1.8	4.9	8.550000	47.6	40.05000	59.425	2.90	1.95	0.4	1.4666667	2.633333	11.6	17.4	38.12500
GSM336265	nonUL	233.10	3.9	3.1	4.4	10.625000	36.5	64.75000	54.950	1.60	1.40	2.4	5.2666667	7.500000	24.9	4.1	49.22500
GSM336266	nonUL	141.10	2.8	4.0	0.9	10.025000	56.8	54.30000	48.200	1.60	3.65	0.9	13.3333333	11.900000	25.2	20.9	35.27500
GSM336267	nonUL	384.65	2.8	2.1	0.3	6.375000	33.4	72.55000	69.500	1.10	2.15	0.9	8.5666667	9.100000	41.5	17.2	72.20000
GSM336268	nonUL	135.05	8.6	1.3	0.4	8.150000	49.9	66.65000	56.700	2.00	3.50	1.7	6.3333333	9.933333	28.5	23.5	45.42500
GSM336269	nonUL	265.05	11.4	1.1	1.0	6.025000	49.9	44.40000	53.800	1.70	3.75	9.1	4.9000000	7.300000	28.3	13.4	41.50000
GSM336270	nonUL	320.50	7.1	1.6	1.8	19.900000	84.3	92.35000	52.725	1.70	2.25	8.8	8.2666667	25.000000	27.3	26.1	45.77500
GSM336271	nonUL	524.75	19.9	6.3	3.1	16.725000	74.4	57.35000	61.625	9.70	2.90	5.7	9.6333333	11.400000	36.0	29.8	34.85000
GSM336272	nonUL	407.05	8.0	5.0	0.7	15.825000	66.7	74.15000	62.850	4.30	3.65	3.1	6.3333333	21.466667	27.8	37.8	45.50000
GSM336273	nonUL	322.20	4.6	4.2	1.9	10.000000	29.7	83.80000	59.075	9.90	1.10	18.6	5.1333333	7.400000	22.8	18.0	49.35000

Showing 1 to 27 of 121 entries, 17 total columns

Results - Machine Learning: universe16_D E_least_results

	predRF	predRF2	predIlda	predGbm	predKNN	predRPART	predGLM	CombinedPredictions2	TYPE
GSM569428	UL	UL	UL	nonUL	nonUL	UL	UL	nonUL	nonUL
GSM1667147UL	nonUL	nonUL	nonUL	nonUL	UL	UL	nonUL	UL	UL
GSM1667148UL	nonUL	nonUL	nonUL	nonUL	UL	UL	nonUL	UL	UL
GSM336202UL	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
GSM336203UL	UL	UL	UL	nonUL	UL	UL	UL	nonUL	UL
GSM336205UL	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
GSM336209UL	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
GSM336216UL	UL	nonUL	nonUL	UL	UL	UL	UL	UL	UL
GSM336217UL	nonUL	nonUL	nonUL	nonUL	UL	UL	nonUL	UL	UL
GSM336220UL	nonUL	UL	UL	nonUL	UL	UL	UL	UL	UL
GSM336225UL	UL	UL	UL	UL	UL	UL	nonUL	nonUL	UL
GSM336226UL	nonUL	nonUL	nonUL	nonUL	UL	UL	nonUL	UL	UL
GSM336227UL	UL	UL	nonUL	UL	nonUL	UL	nonUL	UL	UL
GSM336232UL	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
GSM336245UL	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
GSM336246UL	UL	UL	UL	UL	UL	UL	UL	nonUL	UL
GSM336251UL	nonUL	nonUL	UL	nonUL	UL	UL	UL	UL	UL
GSM38695UL	nonUL	nonUL	nonUL	UL	nonUL	UL	UL	UL	UL
GSM9094UL	UL	UL	nonUL	UL	nonUL	UL	nonUL	UL	UL
GSM9095UL	nonUL	nonUL	nonUL	UL	UL	UL	nonUL	UL	UL
GSM9097UL	nonUL	UL	nonUL	UL	nonUL	UL	nonUL	UL	UL
GSM569430UL	nonUL	nonUL	UL	UL	UL	UL	UL	UL	UL
results	0.31	0.33	0.33	0.47	0.56	0.58	0.42	0.75	100

Methods: Machine Learning All Data Set Outcomes



The next analysis method used a table to compare the various data sets used in each of the machine learning algorithms used on the TOP16 gene results earlier



To do this the results of each of the eight data tables were extracted in R and added to one table of solely results for each algorithm

Results: Machine Learning All Data Set Outcomes

	predRF	predRF2	predIlda	predGbm	predKNN	predRPART	predGLM	CombinedPredictions2	TYPE
TOP16_results	0.69	0.66	0.74	0.69	0.71	0.54	0.74	0.83	100
DE16_most_130_results	0.69	0.67	0.78	0.67	0.92	0.61	0.72	0.92	100
DE16_least_130_results	0.47	0.53	0.42	0.47	0.5	0.58	0.33	0.72	100
FOLD16_130_results	0.64	0.69	0.67	0.72	0.69	0.58	0.72	0.78	100
universe16_fold_results	0.86	0.86	0.86	0.86	0.86	0.86	0.72	0.92	100
universe16_DE_most_results	0.67	0.64	0.81	0.69	0.72	0.53	0.78	0.86	100
universe16_DE_least_results	0.31	0.33	0.33	0.47	0.56	0.58	0.42	0.75	100

Conclusions

- ❖ Five studies from GEO using microarray gene expression data was analyzed to see if the six genes ubiquitous to current UL risk studies make good predictors of UL by first sub-setting the 12,173 gene data to only those 130 genes in the same cytobands as the six UL risk genes
- ❖ A data set of the top10 highest change genes plus the six genes ubiquitous to UL risk was made called TOP16
- ❖ Bootstrap simulations of TOP16 genes made to confirm these genes represent the population well

Conclusions

- ❖ Seven total data sets were built and used to test whether some genes in different categories of expression were good predictors of UL or if the machine learning algorithms were better on the genes in the samples with the most change in fold change or magnitude.
- ❖ Results also showed 92 per cent accuracy for the most expressed genes in DE or fold change out of all genes the five studies had in common
- ❖ Results also showed the least expressed genes out of all genes in DE and fold change made the worst predictors
- ❖ This could mean there is a need to keep looking at the most expressed genes and to keep the six genes ubiquitous to UL risk studies as gene targets for UL pathogenesis, as only including the genes found on the same cytoband locations as the six genes ubiquitous to UL risk studies scored worse than the top genes in all chromosomes

Conclusions

- ❖ Implications of this study:
 - ❖ This could mean there is a need to keep looking at the most expressed genes and to keep the six genes ubiquitous to UL risk studies as gene targets for UL pathogenesis. However, exclusively looking only at the genes that reside on the same cytoband location as the six UL risk genes is not better than looking at the most expressed genes in magnitude or fold change of all genes in UL.
- ❖ Limitations of this study:
 - ❖ Some genes in the middle range of changes in gene expression in UL compared to non-UL samples were ignored, some of these genes could offer clues into UL risk.

Conclusions

- ❖ Future extensions to this study:
 - ❖ These genes that were targeted in the data sets of all genes should be further evaluated and described in a way that could connect how these genes fill the role of UL pathogenesis.

Remaining Questions

- ❖ It is apparent that gene expression data was able to predict up to 92 per cent accuracy of a sample being UL or not.
- ❖ What specific role does each gene play in UL pathogenesis if any when expressed less or more in UL compared to non-UL samples?

References

- Aissani, B., Zhang, K., and Wiener, H. (2015). Evaluation of GWAS candidate susceptibility loci for uterine leiomyoma in the multi-ethnic NIEHS uterine fibroid study. *Frontiers in Genetics*, 6, 241. DOI:10.3389/fgene.2015.00241
- Bondagji, N., Morad, F., Al-Nefaei, A., Khan, I., Elango, R., Abdullah, L., ..., Shaik, N. (2017). Replication of GWAS loci revealed the moderate effect of TNRC6B locus on susceptibility of Saudi women to develop uterine leiomyomas. *Journal of Obstetrics and Gynaecology*, 43(2):330-338. DOI:10.1111/jog.13217
- Cha, P, Takahashi, A., Hosono, N., Low, S., Kamatani, N., Kubo, M., & Nakamura, Y. (2011) A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids. *Nature Genetics* 43(5).
- Dvorská, D., Braný, D., Danková, Z., Halašová, E., & Višňovský, J. (2017). Molecular and clinical treatment of uterine leiomyomas. *Tumor Biology*, 39(6). DOI: 10.1177/1010428317710226.

References

- Edwards, T., Hartmann, K., & Edwards, D. (2013). Variants in BET1L and TNRC6B associate with increasing fibroid volume and fibroid type among European Americans. *Human Genetics*, 132(12). DOI:10.1007/s00439-013-1340-1
- Eggert, S., Huyck, K., Somasundaram, P., Kavalla, R., Stewart, E., Lu, A., ... Morton, C. (2012) Genome-wide linkage and association analyses implicate FASN in predisposition to uterine leiomyomata. *American Journal of Human Genetics*, 91(4), 621–628. DOI: 10.1016/j.ajhg.2012.08.009
- Hellwege, et al. (2017) A multi-stage genome-wide association study of uterine fibroids in African Americans. *Human Genetics*, 136(10), 1363–1373. DOI:10.1007/s00439-017-1836-1
- Hodge, J.C., Kim, T., Dreyfuss, J.M., Somasundaram, P., Christacos, N.C., Rouselle, M., ... Morton, C.C. (2012). Expression profiling of uterine leiomyomata cytogenetic subgroups reveals distinct signatures in matched myometrium: transcriptional profiling of the t (12;14) and evidence in support of predisposing genetic heterogeneity. *Human Molecular Genetics*, 21, 102312–2329. DOI:10.1093/hmg/dds051

References

- Liu, B., Wang, T., Jiang, J., Li, M., Ma, W., Wu, H., & Zhou, Q. (2018). Association of BET1L and TNRC6B with uterine leiomyoma risk and its relevant clinical features in Han Chinese population. *Scientific Reports*, 8,7401. DOI:10.1038/s41598-018-25792-z
- Mayo Clinic (2019) Uterine Fibroids: Overview. Retrieved July 15, 2019 from <https://www.mayoclinic.org/diseases-conditions/uterine-fibroids/symptoms-causes/syc-20354288>
- R, version 3.6 (2019) Ggplot2 Package for world map plotting of genes
- Rafnar et al. (2018) Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits. *Nature Communications*, 9:3636. DOI:10.1038/s41467-018-05428-6
- Rye, Wise, Jurukovski, DeSai, Choi, & Avissair (2017) ‘Eukaryotic Epigenetic Gene Regulation.’ *Biology: OpenStax*. Retrieved from: <https://cnx.org/contents/jVCgr5SL@15.43:5cz8fb2@10/16-3-Eukaryotic-Epigenetic-Gene-Regulation>