Discovery of multi-disease biomarkers in CVD and DM affected HIV cohort using machine learning approaches.

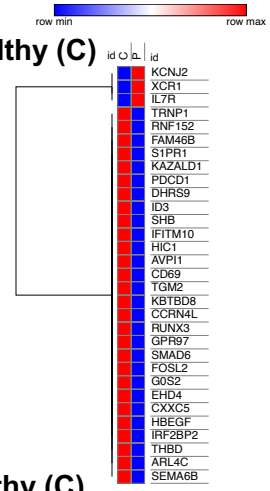
Introduction

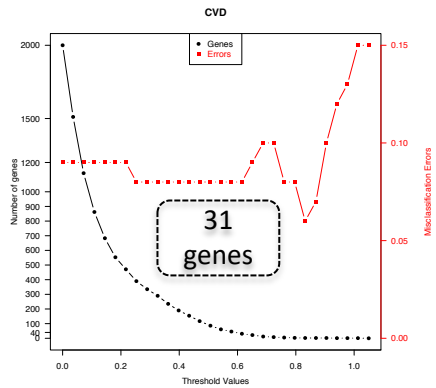
- WIHS RNA Seq data from 92 patients
- Filtered for highly variable genes (2000)
- Voom DDA biomarker prediction
- WGCNA (Module) and PRESTO (Cluster) analysis
- Lasso regression on the PRESTO clusters with diseases
- WGCNA Module disease status and biomarker trait associations
- Common Biomarkers (Voom DDA, WGCNA and PRESTO)
- ROC analysis
- GSEA
- IPA analysis
- Kaplan Meier survival Analysis
- Network Analysis
- Conclusive Biomarkers from three methods

VoomDDA- Biomarker Prediction

A

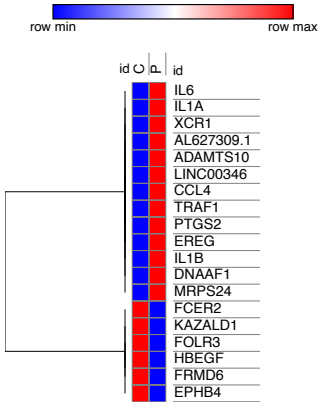
CVD (P) vs Healthy (C)

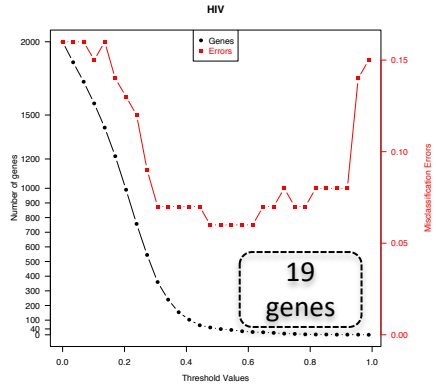




B

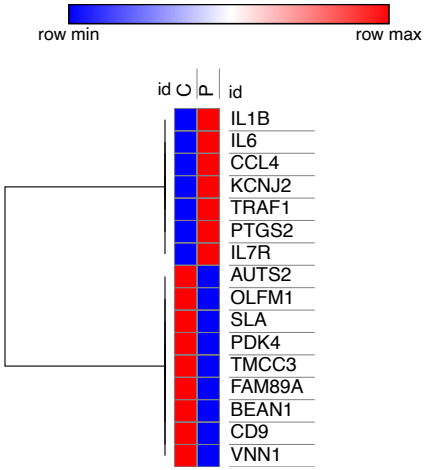
HIV (P) vs Healthy (C)

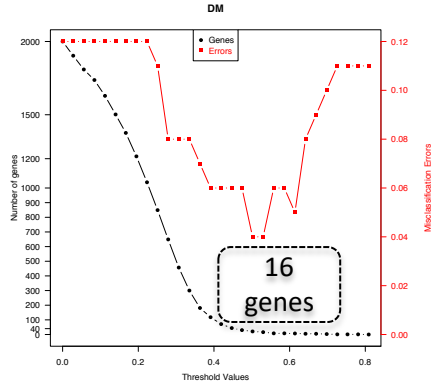




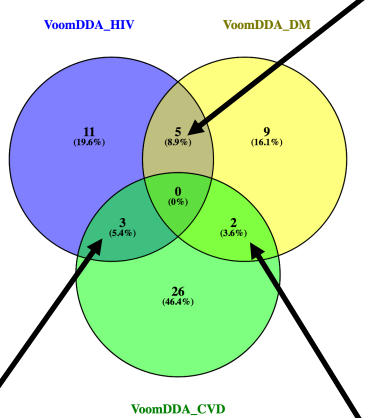
C

DM (P) vs Healthy (C)





HIV vs CVD vs DM



HIV

L1A
FCER2
LINC00346
EREG
FOLR3
EPHB4
AL627309.1
MRPS24
ADAMTS10
FRMD6
DNAAF1

DM

AUTS2
OLFM1
SLA
PDK4
TMCC3
FAM89A
BEAN1
CD9
VNN1

CVD

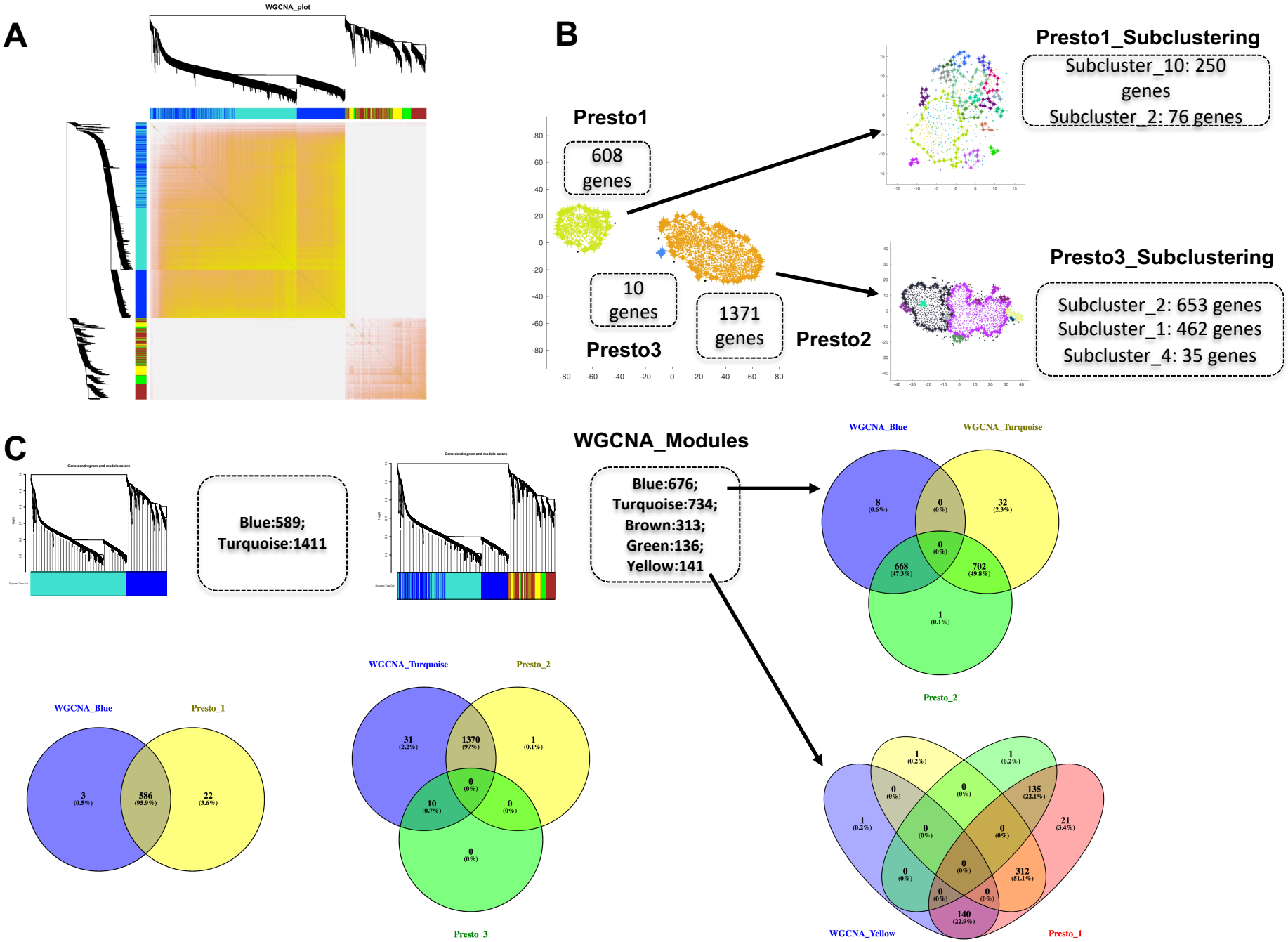
SEMA6B
PDCD1
GPR97
IFITM10
HIC1
S1PR1
SHB
FAM46B
KBTBD8
CXXC5
ARL4C
THBD
RNF152
TRNP1
TGM2
SMAD6
CCRN4L
EHD4
AVPI1
DHRS9
RUNX3
G0S2
IRF2BP2
ID3
CD69
FOSL2

IL6
CCL4
IL1B
PTGS2
TRAF1

XCR1
HBEGF
KAZALD1

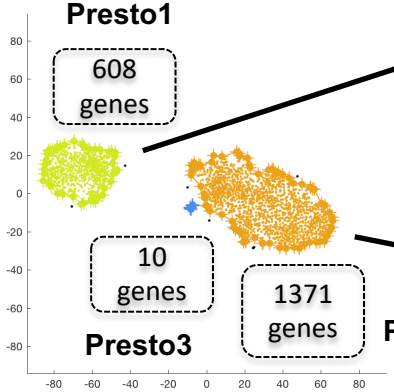
KCNJ2
IL7R

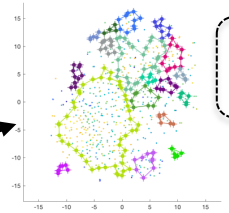
WGCNA –Presto: Disease Correlated Genes



Presto: Lasso Disease Correlated Genes

A

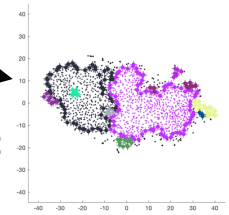




Presto1_Subclustering

Subcluster_10: 250 genes
Subcluster_2: 76 genes

→ Lasso modeling for HIV, CVD and DM



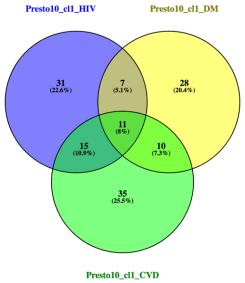
Presto3_Subclustering

Subcluster_2: 653 genes
Subcluster_1: 462 genes
Subcluster_4: 35 genes

→ Lasso modeling for HIV, CVD and DM

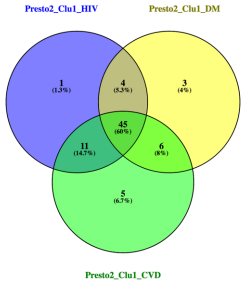
B

Lasso_Subcluster10_P1



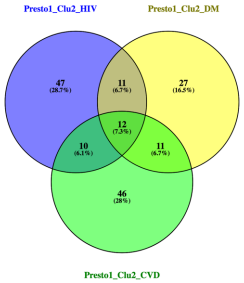
CCR5 CD69 GPX3 HLA.G LAIR2 MSR1 OLFM1 PRUNE2 RPS28 TMEM88 TNNT1

Lasso_Subcluster2_P1



PDCD1 FN1 SH3RF3 TKTL1 CD226 GPNMB BIN1 KREMEN1 DFNA5 ESPL1 AXL DKFZP667F0711 SLC2A5 SLC6A8 MB21D2 FAM115C TNFRSF11A DNAJB5 PAX8 PRKCQ CCL2 INHBA CTSL SLC30A4 SYN1 HILPDA SDS ERRFI1 MTRNR2L8 RASGRP3 HES1 CTD.3088G3.8 HES4 MTHFD1L DHRS9 SPSB1 TMEM51 RORA TNFSF14 CA2 TRIB2 ITGA6 BAI1 SLC41A2 FAM46C

Lasso_Subcluster1_P2

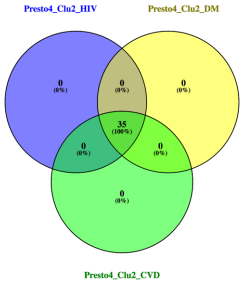


PTPN13 CLCN4 WBP2NL HPGD CSGALNACT1 TBC1D24 CCNE2 SEMA4G TUBD1 C8orf44 BMP8B STAMBPL1

Lasso_Subcluster2_P2

ERAP2 CLEC4D C10orf105 CES1 CCDC125 LGALS2 FAM89A VNN1 TMTC1

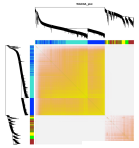
Lasso_Subcluster4_P2



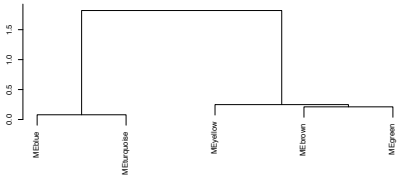
SIGLEC1 IFITM1 SERPING1 TIMMDC1 RSAD2 TFEC IFI44 LGALS3BP LY6E IFI35 ITGA1 XAF1 ATF7IP CMPK2 PARP9 C2 IFIT2 SP110 OAS1 IFI44L ETV7 OAS2 IFIT3 RNASE1 IFI16 MX1 OAS3 SOS1 TDRD7 ISG15 IFITM3 GBP3 ACO1 IFI6 EPSTI1

WGCNA – Module/Gene trait relationship

A

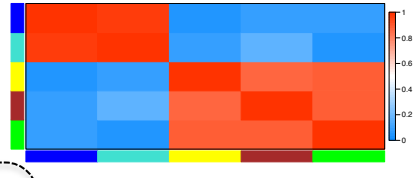


Module_dendrogram



WGCNA_Modules

Blue:676;
Turquoise:734;
Brown:313;
Green:136;
Yellow:141



B

PTGS2
IL6 FLT1
SLC39A8
INHBA
HS3ST3B1
PLK2
TNFRSF11A
DNAAF1
CCL3
MESDC1
NR4A3 IL1A
MT1E
PRKCDBP
ADORA2A
RHOH
PRDM8
SPAG5 IL1B
CCL4 LRP5
TMEM88
STAT4 CDK1
WASF1
NRARP
CCL20 FFAR3
C17orf96
HOXA5
FAM65C
MRPS24
THBS1 PKIG
GLTPD2
EREG CXCL1
EHD1 CEBPA

Gene_CVD relationship

Genes	GS.CVD	p.GS.CVD
ATP6V1C1	0.26776709	0.00986514
DYSF	0.24797759	0.01715585
HPGD	0.23486935	0.02422074
ZNF438	0.22117942	0.0341083
RBPMS	0.20999384	0.04452523
BET1	0.20861745	0.04597258
CCNE2	0.20793522	0.0467043
ZNF780A	0.20679083	0.04795332
MINPP1	0.19828745	0.05812391
TDRD9	0.19804335	0.05844007
ADAM19	0.19748585	0.05916739
C10orf25	0.19731415	0.05939288
RPA4	0.19697864	0.0598355
SYCP2	0.19473716	0.06286209
ANGPT2	0.19343551	0.06467604
HAL	0.19091468	0.06830969
UNC119B	0.18830205	0.07224802
C16orf52	0.1881072	0.07254891
TBCCD1	0.18808097	0.0725895
C1orf115	0.18749661	0.07349836
GBP3	0.18614502	0.07563558
ZNF544	0.18604577	0.07579446
C10orf10	0.18602422	0.07582898
COPB2	0.18170345	0.08301208
GSKIP	0.18007518	0.08585566
NCOA5	0.17962142	0.08666175
CD99L2	0.17942799	0.08700718
IFT80	0.17776307	0.09002588
C10orf88	0.17754061	0.09043541
LTA4H	0.17627567	0.09279225
HOMEZ	0.17616025	0.09300968
SLFN5	0.17527904	0.09468309
C9orf64	0.17507279	0.09507816
ITGA1	0.17402789	0.09709964

C

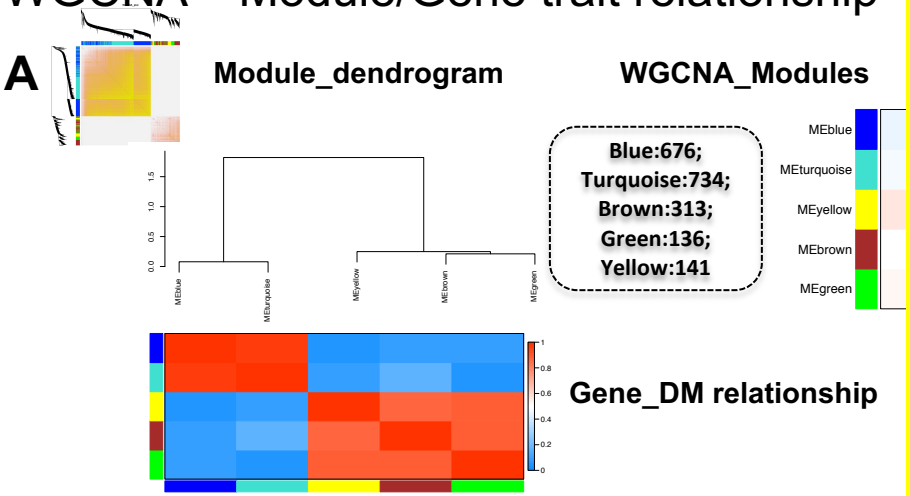
Gene_HIV relationship

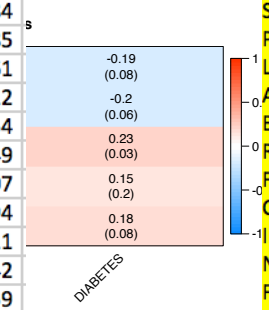
Genes	GS.HIV	p.GS.HIV
PTGS2	0.280331	0.006798
IL6	0.258474	0.012856
FLT1	0.249124	0.016633
SLC39A8	0.238803	0.021878
INHBA	0.229356	0.027861
HS3ST3B1	0.228386	0.028548
PLK2	0.22714	0.02945
TNFRSF11A	0.225481	0.030689
DNAAF1	0.222716	0.032852
CCL3	0.218417	0.036468
MESDC1	0.211501	0.042984
NR4A3	0.210667	0.043831
IL1A	0.210336	0.044171
MT1E	0.20952	0.045019
PRKCDBP	0.207585	0.047084
ADORA2A	0.206263	0.048538
RHOH	0.201336	0.054292
PRDM8	0.199298	0.056829
SPAG5	0.198122	0.058338
IL1B	0.196205	0.060866
CCL4	0.194754	0.062839
LRP5	0.193013	0.065274
TMEM88	0.192473	0.066045
STAT4	0.192451	0.066077
CDK1	0.190881	0.068359
WASF1	0.190626	0.068737
NRARP	0.189862	0.069875
CCL20	0.189298	0.070725
FFAR3	0.187472	0.073537
C17orf96	0.185582	0.076541
HOXA5	0.180323	0.085417
FAM65C	0.179559	0.086773
MRPS24	0.179277	0.087277
THBS1	0.177593	0.090339
PKIG	0.177473	0.090559
GLTPD2	0.175943	0.093419
EREG	0.175923	0.093458
CXCL1	0.174975	0.095266
EHD1	0.173428	0.098276
CEBPA	0.172829	0.099461
CXCL3	0.170484	0.104209

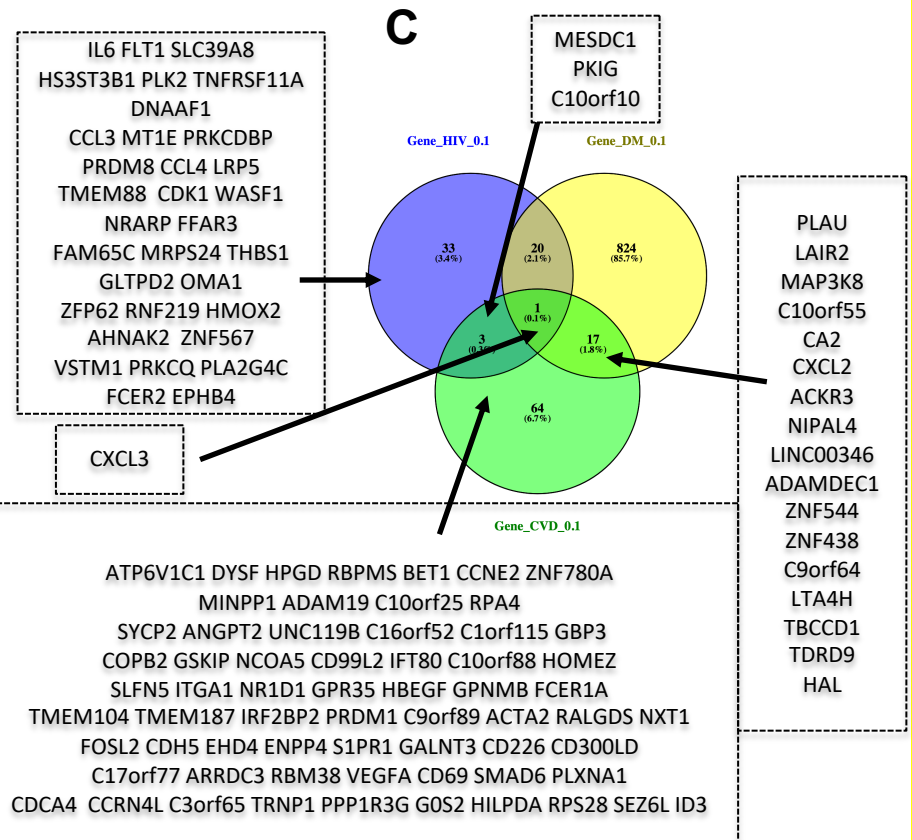
OMA1	-0.172969	0.099182
PRUNE2	-0.173837	0.097473
ZFP62	-0.175075	0.095073
WDR89	-0.175781	0.093727
RNF219	-0.178522	0.08864
PDIK1L	-0.178829	0.088084
HMOX2	-0.1798	0.086344
AHNAK2	-0.185261	0.07706
C10orf10	-0.186096	0.075715
ZNF567	-0.189952	0.06974
VSTM1	-0.190328	0.069178
POLR3K	-0.192966	0.06534
PRKCQ	-0.195666	0.061593
PLA2G4C	-0.238504	0.022049
FCER2	-0.274401	0.008121
EPHB4	-0.289057	0.005197

NR1D1	-0.172673	0.0997
CXCL3	-0.173424	0.09828
GPR35	-0.17355	0.09803
HBEGF	-0.173687	0.09776
GPNMB	-0.174497	0.09618
FCER1A	-0.174575	0.09603
TMEM104	-0.175942	0.09342
CXCL2	-0.176624	0.09213
MESDC1	-0.176783	0.09184
TMEM187	-0.177074	0.09129
CA2	-0.177692	0.09015
NIPAL4	-0.179693	0.08653
IRF2BP2	-0.18079	0.08459
PRDM1	-0.180895	0.08441
C9orf89	-0.181115	0.08403
ACTA2	-0.181649	0.08310
ADAMDEC1	-0.181864	0.08273
RALGDS	-0.182471	0.08169
NXT1	-0.183131	0.08058
FOSL2	-0.184145	0.07888
CDH5	-0.184184	0.07882
EHD4	-0.184959	0.07755
MAP3K8	-0.185088	0.07734
ENPP4	-0.186291	0.07540
S1PR1	-0.186484	0.07509
PKIG	-0.188615	0.07176
GALNT3	-0.189527	0.07037
CD226	-0.189666	0.07016
LINC00346	-0.190156	0.06943
CD300LD	-0.191272	0.06778
C17orf77	-0.191338	0.06768
ARRDC3	-0.192244	0.06637
RBM38	-0.193085	0.06517
VEGFA	-0.193132	0.06510
CD69	-0.198894	0.05734
SMAD6	-0.199863	0.05611
PLXNA1	-0.200651	0.05513
CDCA4	-0.203919	0.05120
CCRN4L	-0.206388	0.048
C3orf65	-0.211637	0.04284
TRNP1	-0.212819	0.04167
PPP1R3G	-0.212852	0.0416
ACKR3	-0.213057	0.04143
LAIR2	-0.214125	0.04040
G0S2	-0.214388	0.04015
HILPDA	-0.216621	0.03807
RPS28	-0.219909	0.03517
SEZ6L	-0.229014	0.02810
PLAU	-0.248307	0.01700
C10orf55	-0.284264	0.00602
ID3	-0.30494	0.0031

WGCNA – Module/Gene trait relationship





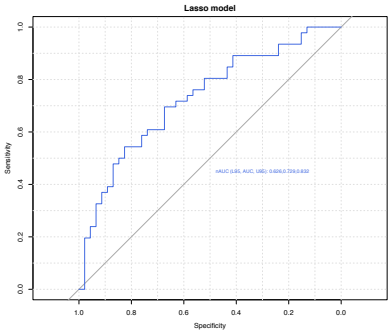


SPTLC2	-0.229776	0.027568
GPR162	-0.229788	0.027561
SCIMP	-0.230467	0.027093
MRPS26	-0.230644	0.026972
TNFSF10	-0.23092	0.026784
GPR34	-0.231215	0.026585
SCRN3	-0.231251	0.026561
CUTC	-0.231324	0.026512
SLC35C1	-0.23141	0.026454
ZNF589	-0.231717	0.026249
KIAA0513	-0.23178	0.026207
ICA1L	-0.232254	0.025894
KLHDC8B	-0.232379	0.025811
ST6GALNAC3	-0.232946	0.025442
CDPF1	-0.233074	0.025359
ZSCAN2	-0.233502	0.025083
CLEC6A	-0.233804	0.024891
CKLF	-0.234336	0.024554
DOK2	-0.234548	0.024421
LEPR	-0.234591	0.024395
FAM198B	-0.234605	0.024386
ZNF33B	-0.235248	0.023987
FBXL5	-0.236629	0.023148
CALHM2	-0.237441	0.022667
POMK	-0.237508	0.022627
JKAMP	-0.237915	0.022389
MSRB2	-0.238795	0.021883
LRRCC1	-0.238962	0.021788
ZFP36L2	-0.239071	0.021726
IVD	-0.239133	0.021691
FNTB	-0.239539	0.021462
ICT1	-0.24	0.021205
ZNF527	-0.241688	0.020286
RCBTB2	-0.241986	0.020127
SIGLEC5	-0.242514	0.019849
C4orf27	-0.243494	0.01934
TRIM5	-0.245249	0.018459
COMMD8	-0.245915	0.018133
PAAF1	-0.247576	0.017343
SUOX	-0.248601	0.016869
CXCL9	-0.249288	0.016559
MRVI1	-0.251912	0.015417
RINL	-0.256583	0.013553
KCTD12	-0.259106	0.01263
1-Mar	-0.259408	0.012524
VNN1	-0.259682	0.012427
SLC46A2	-0.262012	0.011635
FRMD4B	-0.272225	0.008661
RASA4	-0.276568	0.007614
BEAN1	-0.301133	0.003535

Genes	GS.DM	p.GS.DM
NTSR1	0.335638	0.001073
PNPLA1	0.294124	0.00443
BRPF3	0.263576	0.011129
SPHK1	0.262224	0.011566
PLAU	0.258914	0.012698
LYPD3	0.257712	0.013133
ALAS1	0.256201	0.013698
BTG1	0.25061	0.015975
RASGRP3	0.244546	0.018808
FFAR2	0.243267	0.019457
C15orf48	0.239771	0.021333
IL1RN	0.23923	0.021636
NCR3LG1	0.238929	0.021806
F3	0.238593	0.021998
SYN1	0.238214	0.022216
IER3	0.238171	0.022241
IRAK2	0.237189	0.022815
ETS2	0.236799	0.023046
LAIR2	0.236458	0.02325
EIF2AK3	0.236267	0.023365
BCAS1	0.236119	0.023454
CHST7	0.235761	0.023672
AMZ1	0.235688	0.023717
NR4A3	0.234511	0.024445
ATP13A3	0.234121	0.02469
SPAG5	0.233964	0.024789
HMGA1	0.233148	0.025311
NINJ1	0.233088	0.02535
EHD1	0.23306	0.025368
TNFRSF9	0.232708	0.025597
SLC30A4	0.232698	0.025603
SYNJ2	0.231684	0.026271
FRMD6	0.230869	0.026819
ADNP2	0.23074	0.026906
B3GNT5	0.229498	0.027762
GPX3	0.228662	0.028351
IL7R	0.22769	0.029049
ZBTB43	0.227004	0.02955
CCL2	0.226765	0.029727
INHBA	0.2265	0.029923
NKX3-1	0.226471	0.029944
AIFM2	0.225723	0.030506
NFAT5	0.225476	0.030693
CCDC71L	0.225311	0.030819
PHF10	0.224989	0.031065
PTPN7	0.223842	0.031956
NRIP3	0.223453	0.032263
CCL20	0.222841	0.032752
PANX2	0.222138	0.03332
TIMP1	0.222007	0.033426

CVD Common Biomarkers VoomDDA- WGCNA-Presto_Lasso

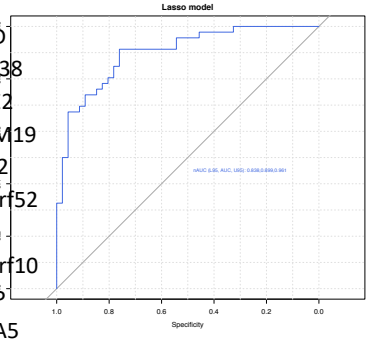
HBEGF
IRF2BP2
FOSL2
EHD4
S1PR1
CCRN4L
TRNP1
G0S2



AUC:0
.729

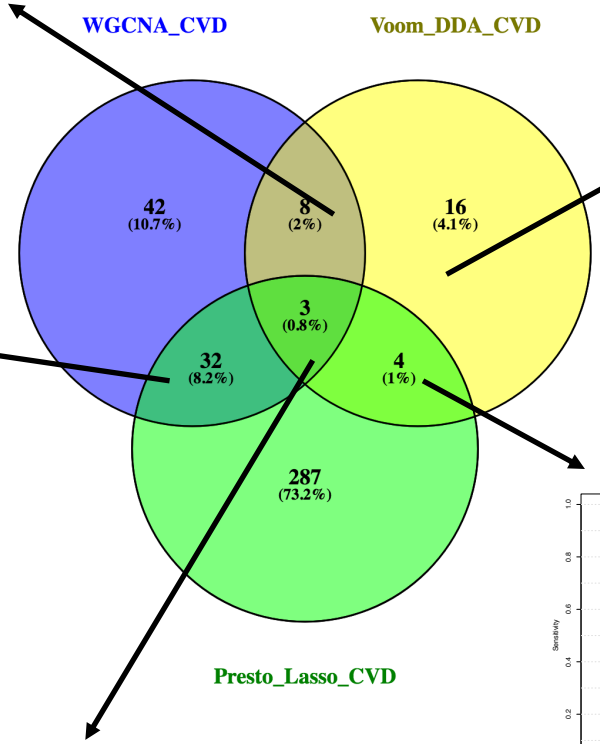
CD69
SMAD6
ID3

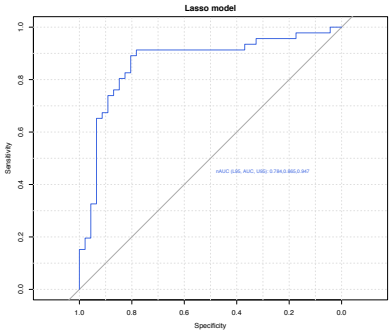
ATP6V1C1
DYSF
HPGD
ZNF438
CCNE2
ADAM19
SYCP2
C16orf52
GBP3
C10orf10
GSKIP
NCOA5
IFT80
C10orf88
LTA4H
HOMEZ
C9orf64
ITGA1
GPR35
GPNMB
CXCL2
MESDC1
CA2
PRDM1
C9orf89
CD226
RBM38
VEGFA
CDCA4
LAIR2
HILPDA
RPS28



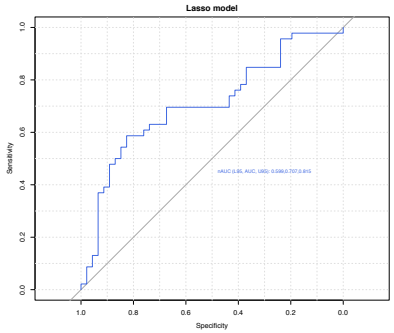
CD69
SMAD6
ID3

AUC:0
.899



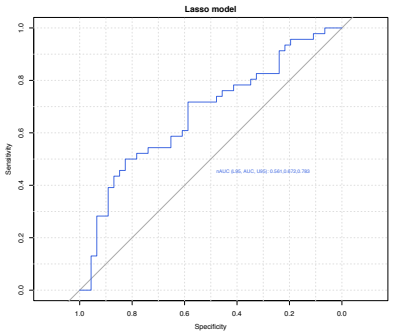


AUC:0
.865



PDCD1
TGM2
AVPI1
DHRS9
CD69
SMAD6
ID3

AUC:0
.707



AUC:0
.672

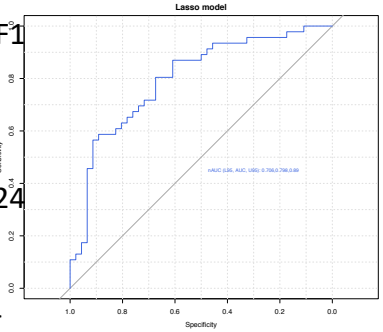
CD69
SMAD6
ID3

CD69
SMAD6
ID3
SEMA6B
GPR97
IFITM10
HIC1
SHB
FAM46B
KBTBD8
CXXC5
ARL4C
KCNJ2
THBD
RNF152
XCR1
IL7R
RUNX3
KAZALD1
PDCD1
TGM2
AVPI1
DHRS9

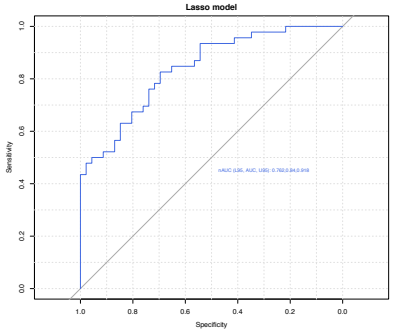
HBEGF
IRF2BP2
FOSL2
EHD4
S1PR1
CCRN4L
TRNP1
G0S2

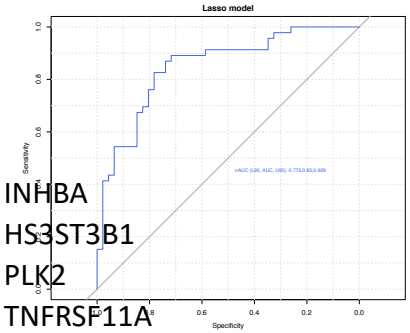
HIV Common Biomarkers VoomDDA- WGCNA-Presto_Lasso

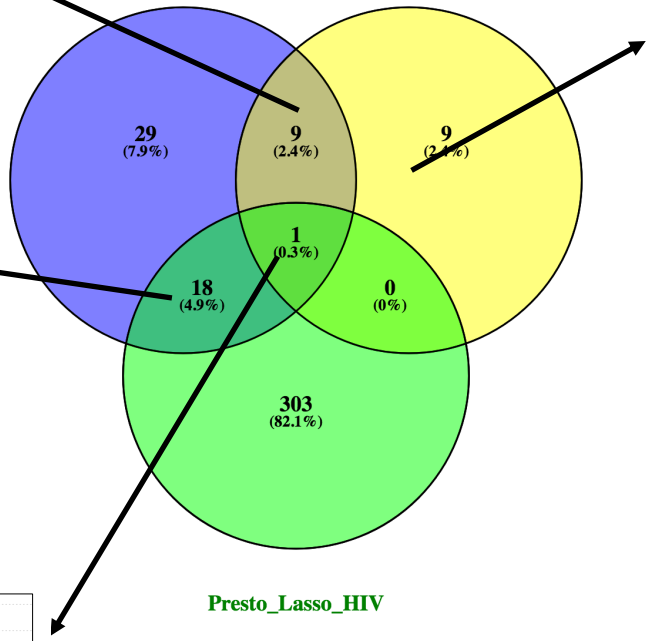
PTGS2
IL6
DNAAF1
IL1A
IL1B
CCL4
MRPS24
EREG
FCER2
EPHB4



AUC:0
.798





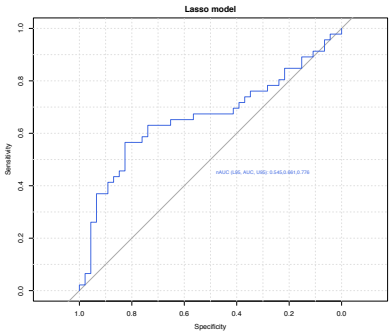


AUC:0
.84

PTGS2
IL6
DNAAF1
IL1A
IL1B
CCL4
MRPS24
EREG
FCER2
EPHB4
LINC00346
XCR1
FOLR3
TRAF1
AL627309.1
HBEGF
ADAMTS10
KAZALD1
FRMD6

INHBA
HS3ST3B1
PLK2
TNFRSF11A
MESDC1
TMEM88
WASF1
FAM65C
THBS1
GLTPD2
OMA1
PRUNE2
ZFP62
WDR89
PDIK1L
C10orf10
PRKCQ
PLA2G4C EPHB4

AUC:0
.85

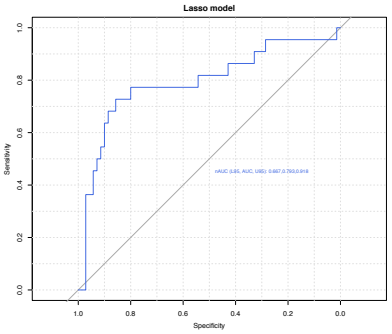


AUC:0
.661

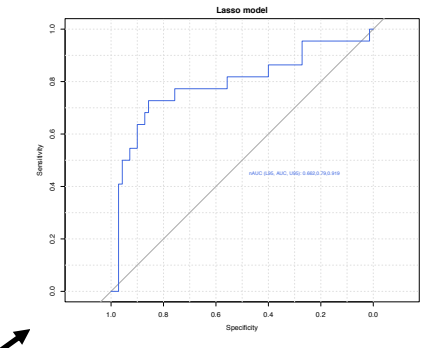
EPHB4

DM Common Biomarkers VoomDDA- WGCNA-Presto_Lasso

IL7R
TRAF1
PTGS2
IL1B
CD9
OLFM1
PDK4
VNN1
BEAN1



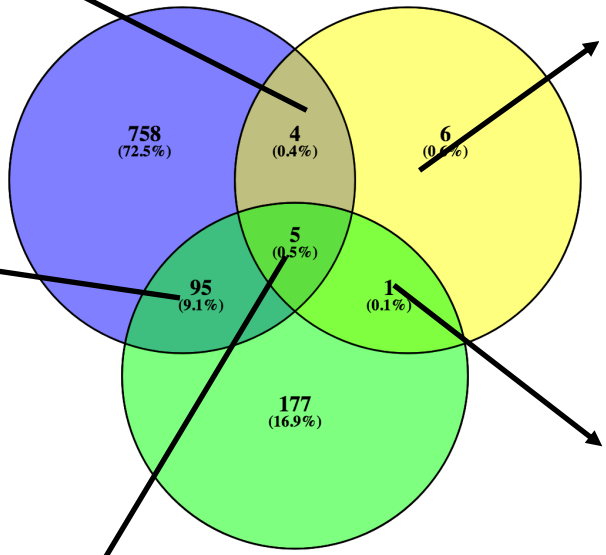
AUC:0
.793



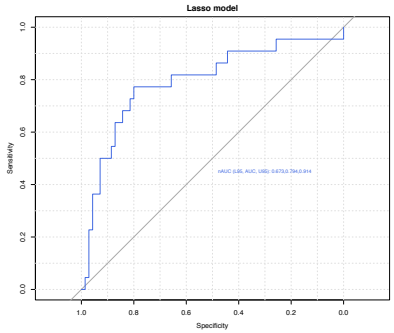
CD9
OLFM1
PDK4
VNN1
BEAN1

AUC:0
.79

IL6
CCL4
KCNJ2
AUTS2
SLA
TMCC3
IL7R
TRAF1
PTGS2
IL1B
FAM89A

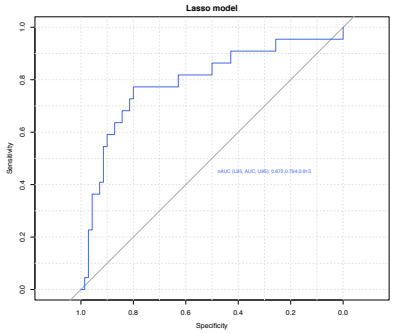


AUC:0
.85



AUC:0
.794

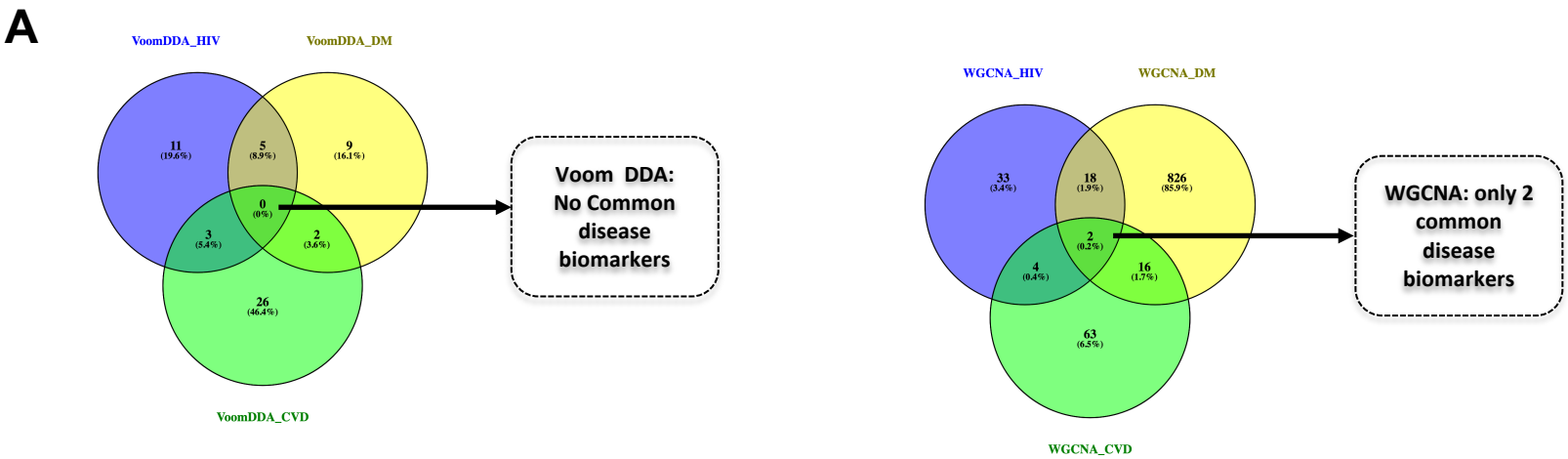
CD9
OLFM1
PDK4
VNN1
BEAN1

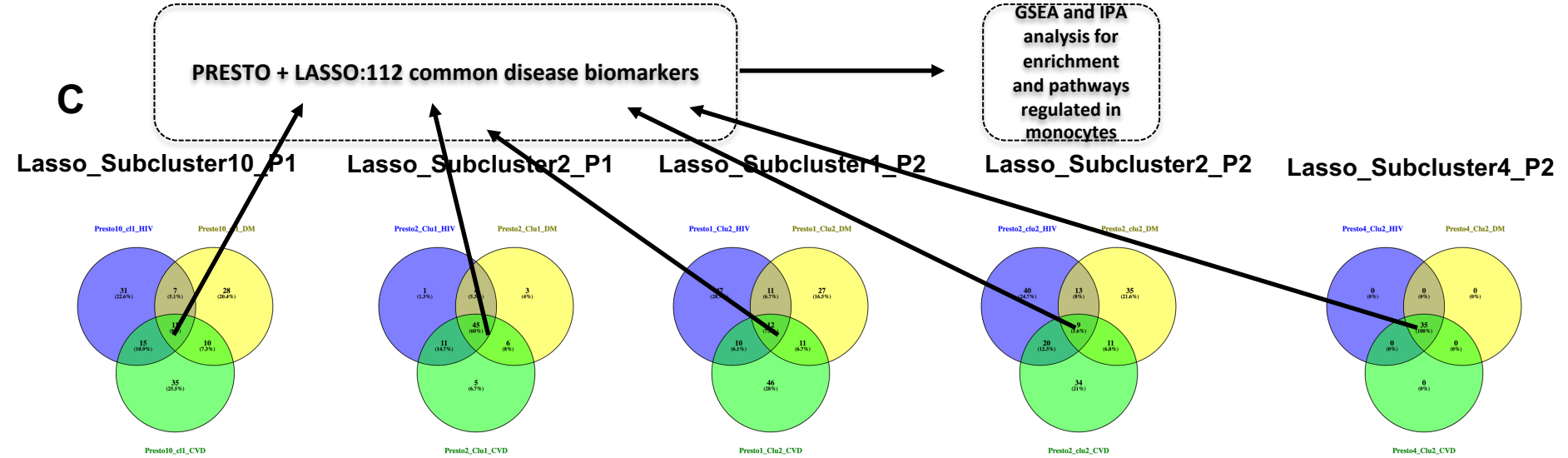


AUC:0
.794

CD9
OLFM1
PDK4
VNN1
BEAN1
FAM89A

Common Disease Biomarkers





GSEA

- 7 HIV GEO datasets; 6 human and 1 monkey, include transcriptomes from PBMCs and blood; GSE50834, GSE56837, GSE57730, GSE108294, GSE44228, GSE52900, and GSE98717

GEO Accession	Disease	Samples	Species	Tissue	Groups
GSE50834	HIV and HIV/TB	44	Homo sapiens	PBMCs	Two (HIV vs HIV/TB)
GSE56837	HIV	26	Homo sapiens	Blood	Two (Progressor vs NonProgressor)
GSE57730	HIV	12	Homo sapiens	Blood	Two (Progressor vs NonProgressor)
GSE108294	HIV	81	Homo sapiens	PBMCs	Two(HIVControllers vs cART)
GSE44228	HIV	72	Homo sapiens	PBMCs	Two(Before Treatment vs After Treatment)
GSE52900	HIV	14	Homo sapiens	Monocytes	Three (Control vs viral supression vs virologica

- 5 CVD GEO datasets: Human, GSE21545, GSE43292, GSE62646, GSE48060 and GSE41571

GSE21545	CVD (Atherosclerosis)	223	Homo sapiens	Carotid Plaques and PBMCs	Two (PBMC vs Plaque)
GSE43292	CVD (Atheroma)	64	Homo sapiens	Atheroma plaque (ATH MIT)	Two (Intact tissue vs Atheroma)
GSE62646	CVD (Myocardial Infarction)	98	Homo sapiens	PBMCs	Two (Control vs Patient)
GSE48060	CVD (Myocardial Infarction)	52	Homo sapiens	Blood	Two (Control vs Patient)
GSE41571	CVD (Atheroma)	11	Homo sapiens	Macrophages from Ruptured and Stable Plaque	Two (Ruptured vs Stable Plaque)

- 5 DM GEO datasets: Human, GSE38642, GSE41762, GSE25274, GSE33440 and GSE65561

GSE38642	Type 2 diabetes (T2D)	63	Homo sapiens	Pancreatic islets	Two(Non-diabetic vs Diabetic)
GSE41762	Type 2 diabetes (T2D)	77	Homo sapiens	Pancreatic islets	Two(Non-diabetic vs Diabetic)
GSE25724	Type 2 diabetes (T2D)	13	Homo sapiens	Pancreatic islets	Two(Non-diabetic vs Diabetic)
GSE33440	Type 1 diabetes (T1D)	22	Homo sapiens	CD14+ monocytes	Two(Non-diabetic vs Diabetic)
GSE65561	Type 2 diabetes (T2D)	18	Homo sapiens	Peripheral blood monocyte	Four(ControlNDUTvs ControlNDT vs DIABETICU

To be Done

- GSEA
- IPA analysis
- Kaplan Meier survival Analysis
- Network Analysis

To be Done

- ROC plots
- GSEA
- IPA analysis
- Kaplan Meier survival Analysis
- Network Analysis
- Conclusive Biomarkers if any from three methods