

NMF流程 以软件cNMF为例

①准备数据

```
library(Seurat)
library(tidyverse)

malignant.seu=readRDS("malignant.seu.1129.rds")

###导出样本的矩阵
for (si in as.character(unique(malignant.seu@meta.data$sample))) {
  small.meta.data=malignant.seu@meta.data %>% filter(sample == si)
  small.count=as.data.frame(malignant.seu[["RNA"]@counts[,small.meta.data$CB])
  small.count=small.count[rowSums(small.count) > 0,]
  small.count=small.count[!str_detect(rownames(small.count), "^MT-"),]
  small.count=as.data.frame(t(small.count)) ← 转置
  write.table(small.count,file = paste0(si,".count.txt"),quote = F,sep = "\t",row.names = T,col.names = T)
}
```

CB表示cellular barcode

注意：后面输入的是count矩阵，
cell x genes

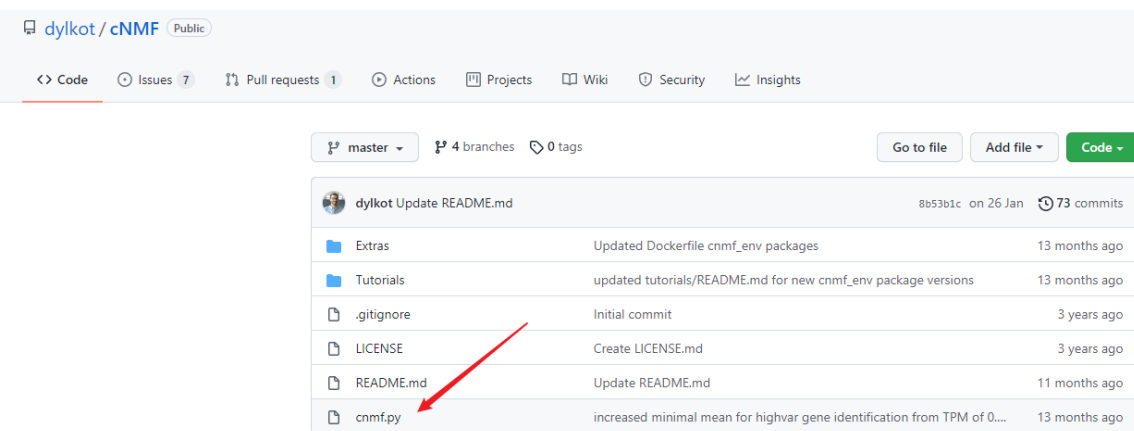
```
$ ll -lh
total 253M
-rw-r--r-- 1 root root 2.5M Dec 12 18:02 P15.count.txt
-rw-r--r-- 1 root root 7.2M Dec 12 18:02 P14.count.txt
-rw-r--r-- 1 root root 60M Dec 12 18:02 P13.count.txt
-rw-r--r-- 1 root root 42M Dec 12 18:01 P12.count.txt
-rw-r--r-- 1 root root 81M Dec 12 18:01 P11.count.txt
-rw-r--r-- 1 root root 8.0M Dec 12 17:59 P06.count.txt
-rw-r--r-- 1 root root 5.6M Dec 12 17:59 P05.count.txt
-rw-r--r-- 1 root root 3.1M Dec 12 17:59 P04.count.txt
-rw-r--r-- 1 root root 5.0M Dec 12 17:59 P03.count.txt
-rw-r--r-- 1 root root 1.9M Dec 12 17:59 P02.count.txt
-rw-r--r-- 1 root root 38M Dec 12 17:59 P01.count.txt
nmfuser1 16:05:01 ~/1212/count_data
```

将得到的表达数据上传至工作目录下面的某个文件夹下，
比如count_data：

[单细胞高级分析：非负矩阵分解\(NMF\)_哔哩哔哩_bilibili](#)

本教程由公众号【TOP生物信息】制作

② 下载程序



使用的软件为cNMF，2019年发表在eLife期刊上，链接：
<https://github.com/dylkot/cNMF>

下载的cnmf.py就是后面会用到的脚本，建议在单独的conda环境下运行（注意：子环境里面还需要安装一些依赖包，见github）

③ 运行程序

prepare：对数据标准化，以及指定一些参数

参数	解释
--output-dir	输出文件夹
--name	子文件夹，以及结果文件的名称前缀
-c	count矩阵的路径
-k	可以是多个空格隔开的值，表示每个表达矩阵在NMF之后被分解成几个表达程序
--n-iter	迭代的次数，为了增加准确性，会多次执行NMF
--total-workers	几个线程（一个就够用了，cNMF很快，且使用较少的计算资源）
--numgenes	高变基因的数量，简化分解（最终的program仍然会包含所有基因）

```
source /home/hsy/miniconda3/bin/activate cnmf_env
```

```
(cnmf_env) nmfuser1 10:04:25 ~/1212
$ ll -th
total 104K
-rw-r--r-- 1 nmfuser1 nmfuser1 41K Dec 14 10:02 cnmf.py
-rw-r--r-- 1 nmfuser1 nmfuser1 56K Dec 14 10:02 hallmark_cancersea.gmt
drwxr-xr-x 2 root      root      4.0K Dec 13 15:58 count_data/
(cnmf_env) nmfuser1 10:04:40 ~/1212
$ which python
/home/hsy/miniconda3/envs/cnmf_env/bin/python
```

```
python ./cnmf.py prepare --output-dir ./res1/ --name
P01_cNMF -c ./count_data/P01.count.txt -k 3 4 5 6 7 8 9 10
--n-iter 300 --total-workers 1 --numgenes 2000
```

factorize: 分解

--worker-index: 为0的时候表示在第一个核中，运行所有的计算任务（总任务数等于k可选值的个数×每个k值之下迭代的次数）

```
python ./cnmf.py factorize --output-dir ./res1/ --name P01_cNMF --worker-index 0
```

combine: 合并多次迭代（k一定的情况下），以获得更加可靠的结果

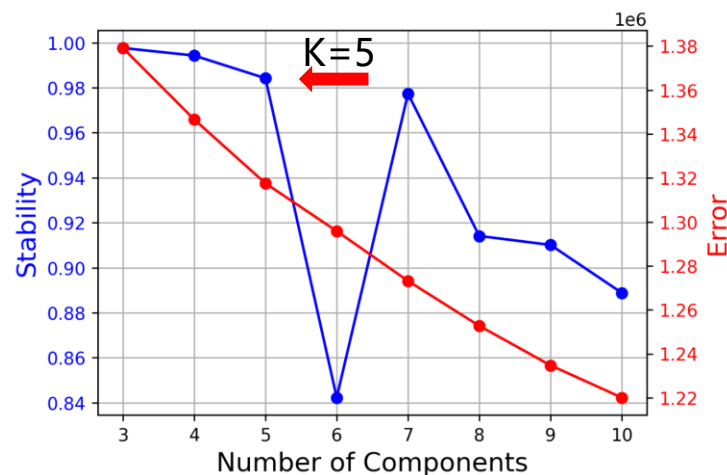
```
python ./cnmf.py combine --output-dir ./res1/ --name P01_cNMF  
rm -f ./res1/P01_cNMF/cnmf_tmp/P01_cNMF.spectra.k_*.iter_*.df.npz
```

k_selection_plot: 计算稳定性和出错可能性（k一定的情况下）

```
python ./cnmf.py k_selection_plot --output-dir ./res1/ --name P01_cNMF  
source /home/hsy/miniconda3/bin/deactivate cnmf_env
```

如何来选择最合适的k值

以病人P01为例

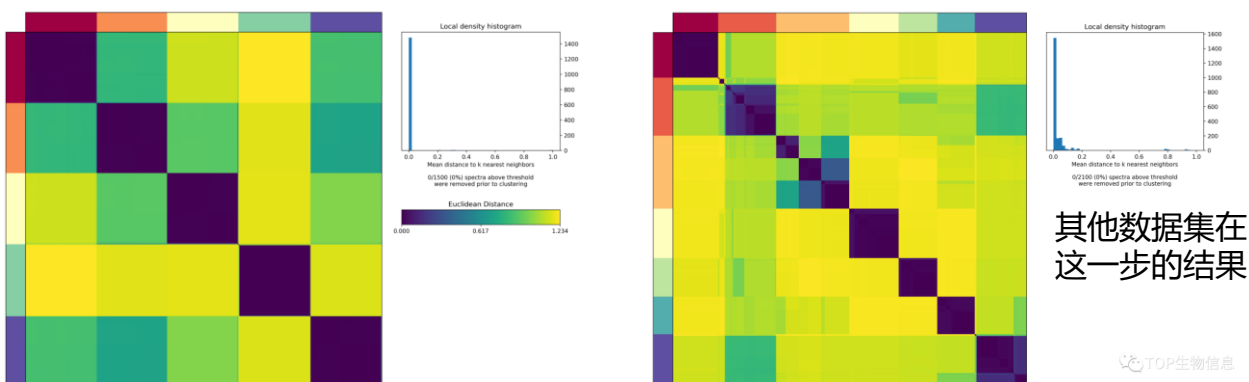


确定k值后，查看不同迭代次数之间的一致性

先看不做任何过滤的结果，dt阈值选为2

```
python ./cnmf.py consensus --output-dir ./res1/ --name P01_cNMF --components 5 --local-density-threshold 2 --show-clustering
```

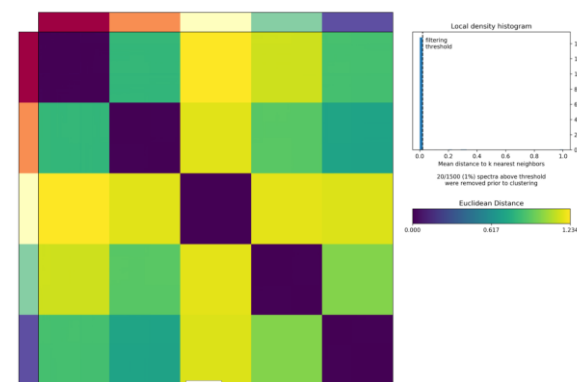
之后会得到一个名为 P01_cNMF.clustering.k_5.dt_2.png的热图



从热图和柱形图可以看出来，有几次分解迭代的结果与大多数结果不一致，可以去掉。从柱形图可知，过滤阈值选择0.02就可以。

```
python ./cnmf.py consensus --output-dir ./res1/ --name P01_cNMF --components 5 --local-density-threshold 0.02 --show-clustering
```

之后会得到一个名为 P01_cNMF.clustering.k_5.dt_0_02.png的热图



④查看输出结果

根据刚才的结果，k选择的是5，过滤阈值dt选择的是0.02，所以我们关注的文件是：

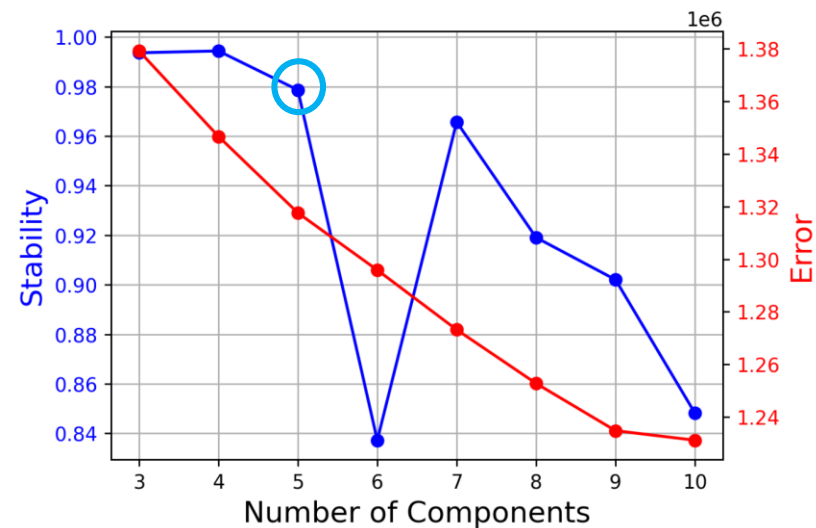
P01_cNMF.usages.k_5.dt_0_02.consensus.txt，P01_cNMF.gene_spectra_score.k_5.dt_0_02.txt，
细胞×表达程序的矩阵。存放的是每个细胞中，每个表达程序×基因的矩阵。存放的是每个program中，
每个基因的权重。往往根据权重最大的前20/30/50个基因来确定该program的功能。

P01_AAACCTGAGGACACCA_1	3.760131265381992	653.6581193953476	30.113706037268667	66.10880383652221	430.04274070799386
P01_AAACCTGGTAGGACAC_1	39.838888973649	1594.597700144216	28.076304517290332	165.30496325412938	403.12521649987957
P01_AAACGGGAGACCTTGG_1	140.15781598222298	0.0	51.614725138434686	170.398071033986	492.8520542323195
P01_AAACGGGAGATCCTGT_1	0.0	934.9690974061734	25.40190132504505	381.60924190753116	0.0
P01_AAACGGGTGGTTAAC_1	36.332465157108	174.23942285694625	5.175263444914462	109.04251830636404	212.0440482368194
P01_AAAGATGTCCACATACC_1	26.36282564845834	529.0198635206373	7.3262341553807895	309.61874296418796	0.0
P01_AAAGATGTCCACGTGG_1	240.10041235939977	0.0	19.834818436697205	113.92725708021097	659.6786116054902
P01_AAAGCAAGAGACTAT_1	164.31525492699475	409.38156080166093	0.0	198.2954654183657	837.5425253037064
P01_AAAGCAAGATCACGG_1	132.7217964202966	119.14985795947918	5.786401144892716	400.40227326117815	534.4268580306068
P01_AAAGCAATCATCACCC_1	0.0	640.5639086608088	26.914679967618284	113.11991324717845	14.03010753166626
P01_AAAGTAGGTACCCAAT_1	0.0	1132.022024954182	323.12776376013636	390.9490911830261	15.40966385167293
P01_AAATGCCCAACACGCC_1	27.638188121778754	210.88864855759104	15.054507130105058	506.1448394517766	250.93965774278095
P01_AACCATGCAAGTACCT_1	168.31827179808514	645.5989147412655	64.94768928647704	302.3788804953865	521.3903019471145
P01_AACCATGGTGCACGAA_1	0.07442268520514918	252.88699003628278	24.300056048471006	131.1946536305838	0.0
P01_AACCGCAGGAGTTGC_1	372.5960346360971	563.4332853483386	20.131453997601884	0.0	299.3086728842354
P01_AACCGCGCACAGACTT_1	44.074003274954286	692.7269953989178	0.0	73.88451722264927	442.945401723854
P01_AACCGCGCAGATGGCA_1	535.0016647283875	1035.637584400387	398.98668941326247	258.317030229975	746.0717671559873
P01_AACCGCGGTAAAGGCT_1	234.77646192869153	210.4301439059867	748.0470970875134	263.2465217558393	677.1408570298331
P01_AACCGCGGTGATGTT_1	746.5276202929501	783.9879645023442	8.108021993300715	308.9902255021265	216.57365339376514
P01_AACGTTGAGACTAGAT_1	43.05492310713551	332.8872032714493	28.275024692394368	72.73543686885823	0.0
P01_AACGTTGAGAGTGAGA_1	362.0283253774249	142.96978071343722	49.02197279230273	79.9737244175198	686.0187095103814
P01_AACGTTGAGGGATACC_1	446.87840275712756	0.0	518.3636662865596	0.0	803.5302199473145
P01_AACTCAGCACCTATCC_1	223.96023705156435	0.0	26.418563270707065	0.0	1034.2090160735358
P01_AACTCAGTCCGTCAAA_1	6.6672302870307405	308.57121772211974	42.707266513823335	149.58390334046996	0.0
P01_AACTCCCGTCTAGGTT_1	473.2350295480852	589.4050885138216	45.32556280221995	63.5125929807197	1019.4445180274072
P01_AACTCCCTCATTATC_1	15.647021514226479	179.9718668318199	351.16493053459703	369.78661715737746	0.0
P01_AACTCTTGTCAAATCT_1	629.6561968001578	337.8972699926963	0.0	206.0678002560102	249.61050589161164
P01_AACTGGTCACTCACTA_1	772.352718792216	593.682065443483	0.0	0.0	223.3651723194563
P01_AACTGGTCTACCAAGA_1	99.12116480205157	760.0874083515557	0.0	0.0	686.2912676781494
P01_AACTTTCCACCATGTA_1	6.590764029687398	576.616148266252	0.0	417.075077101468	0.0
P01_AACTTTCTCACTTACT_1	0.0	612.1347848654929	0.0	405.7925385851727	67.11652469226371
P01_AAGACCTTCATGGTCA_1	0.0	870.5240812706242	0.0	505.0616967039144	892.9602858896525
P01_AAGCGCAGGGTTCCT_1	251.79362251895333	1137.897579720903	12.062370973427594	0.0	189.7152714695885
P01_AAGCGCAGTGTACCT_1	0.0	1269.9662119739282	143.33860821161585	252.1449020353454	766.7442734960904
P01_AAGCGCGCTAAGAGGA_1	0.0	227.08466402670135	12.953912946674308	139.31922439465276	0.0
P01_AAGCGCGCTCAAGTGT_1	292.98469684599314	761.8939115463138	371.18574016871224	865.6849055398212	1246.5232543760785

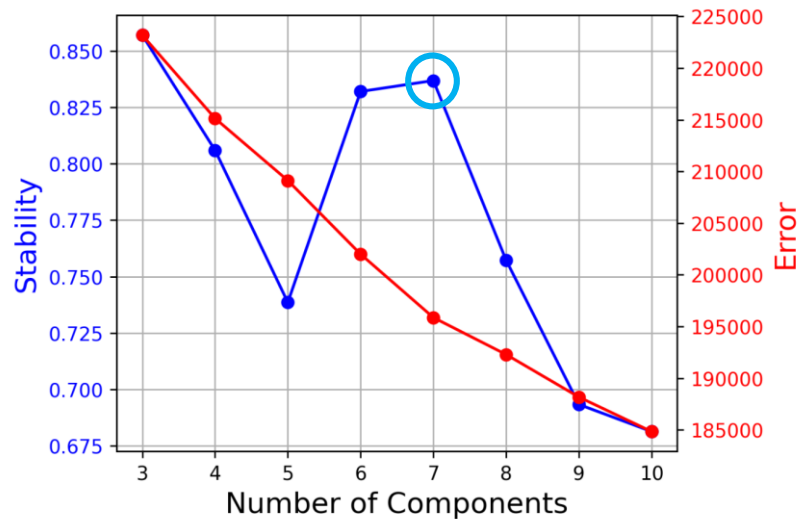
	AL627309.1	AL669831.5	LINC00115	FAM41C	AL645608.7
1	4.229067287149837e-05	0.00011365677108898182	-6.489053205210509e-05		
2	0.00010104555406918348	-8.18175998205438e-05	8.493749075511788e-06		
3	-0.00012163971006952491	0.0002946535937365169	0.0001168019276111907		
4	-0.0001239722687926997	3.800678988208237e-05	-2.9602230861572994e-05		
5	-4.868944839861656e-05	-4.8245028911907234e-05	-5.981249206925331e-05		

以上只是一个样本，我们需要对剩余样本进行同样的处理（依次运行③④步）

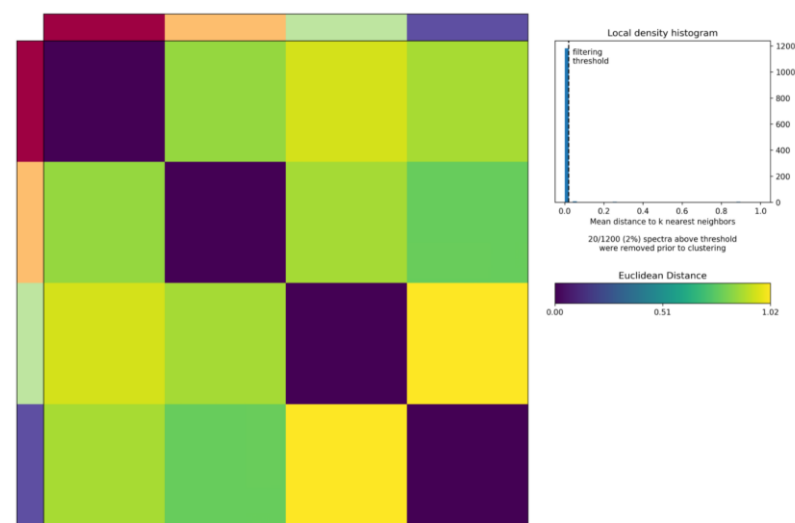
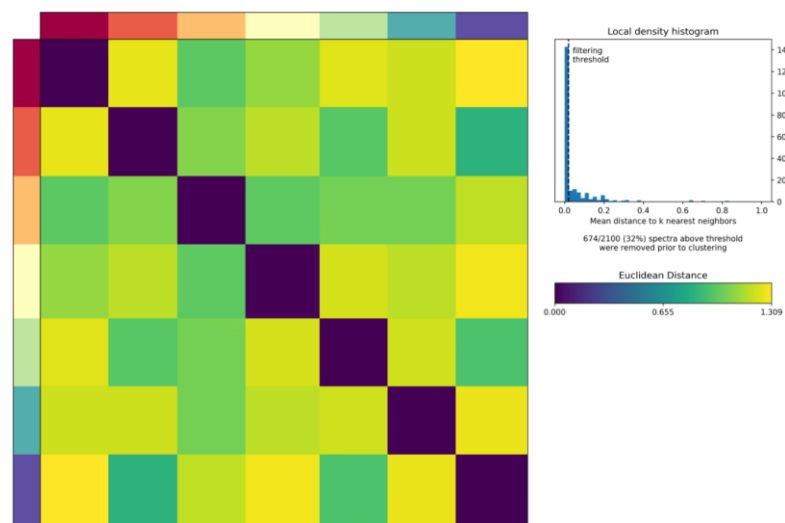
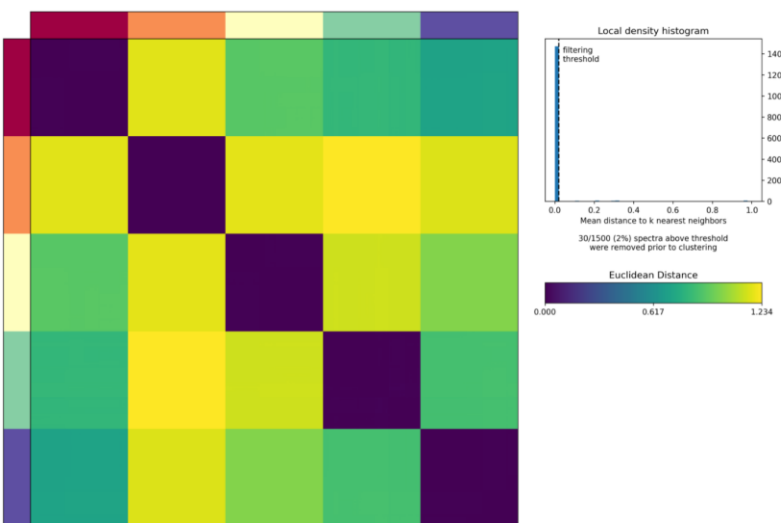
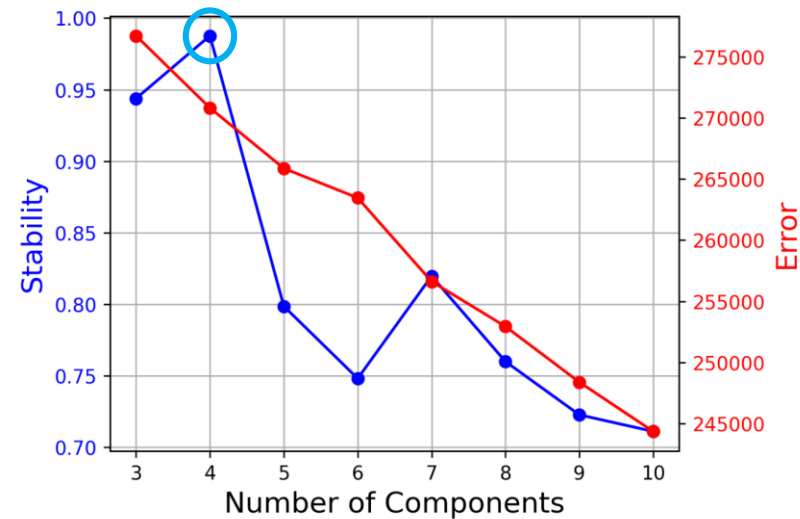
P01.k_selection.png



P03.k_selection.png



P05.k_selection.png

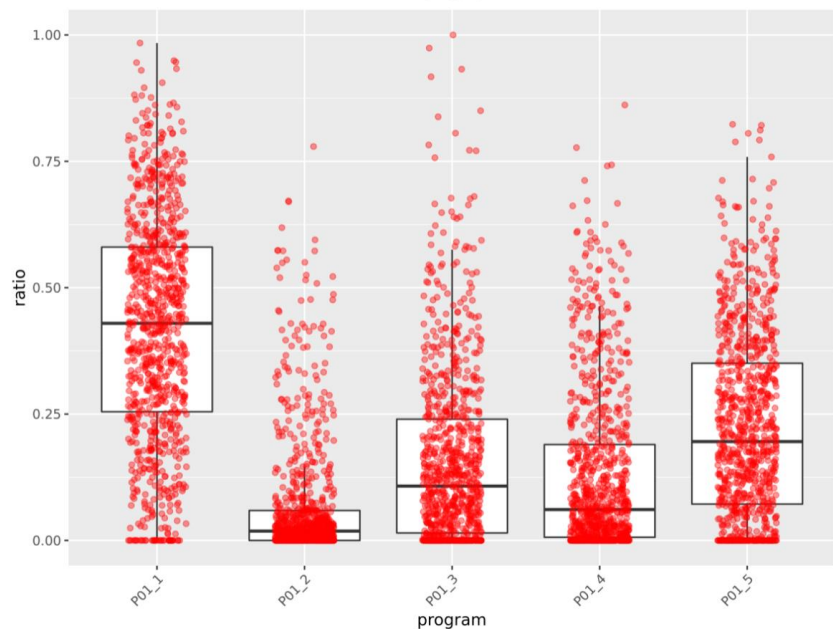


P01.clustering.k_5.dt_0_02.png

P03.clustering.k_7.dt_0_02.png

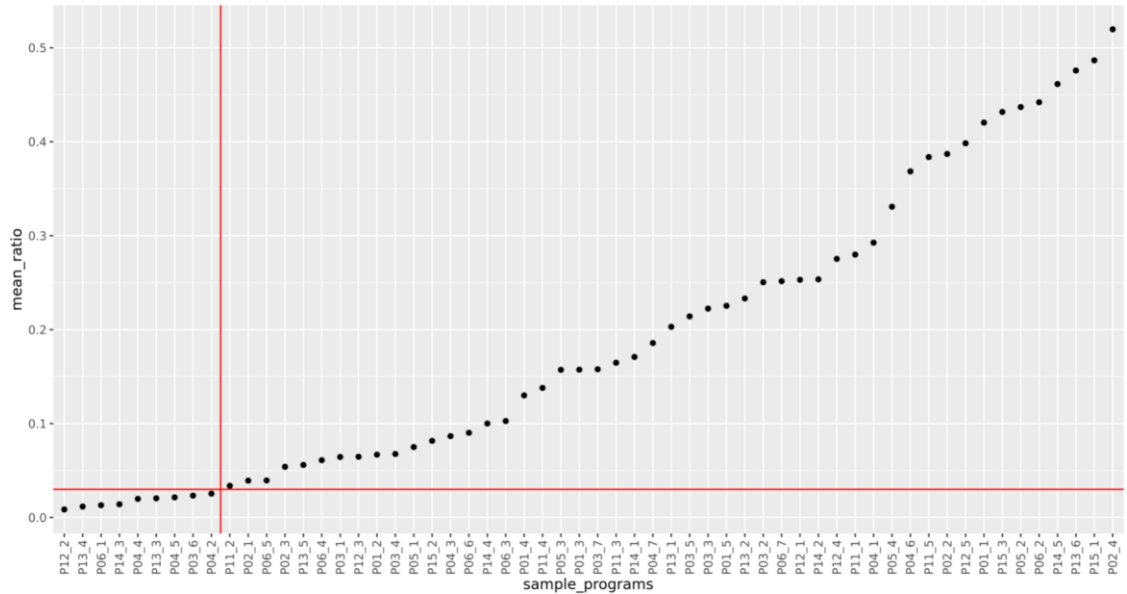
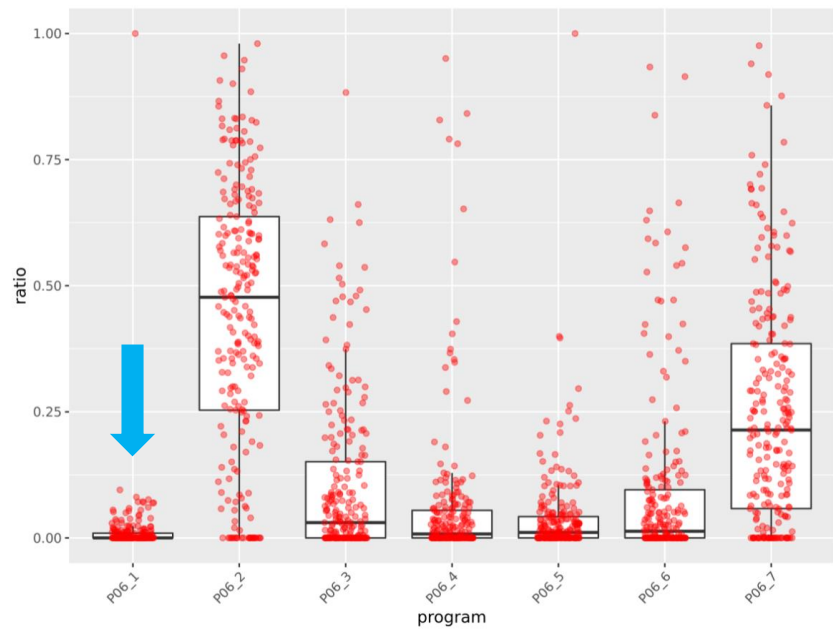
P05.clustering.k_4.dt_0_02.png

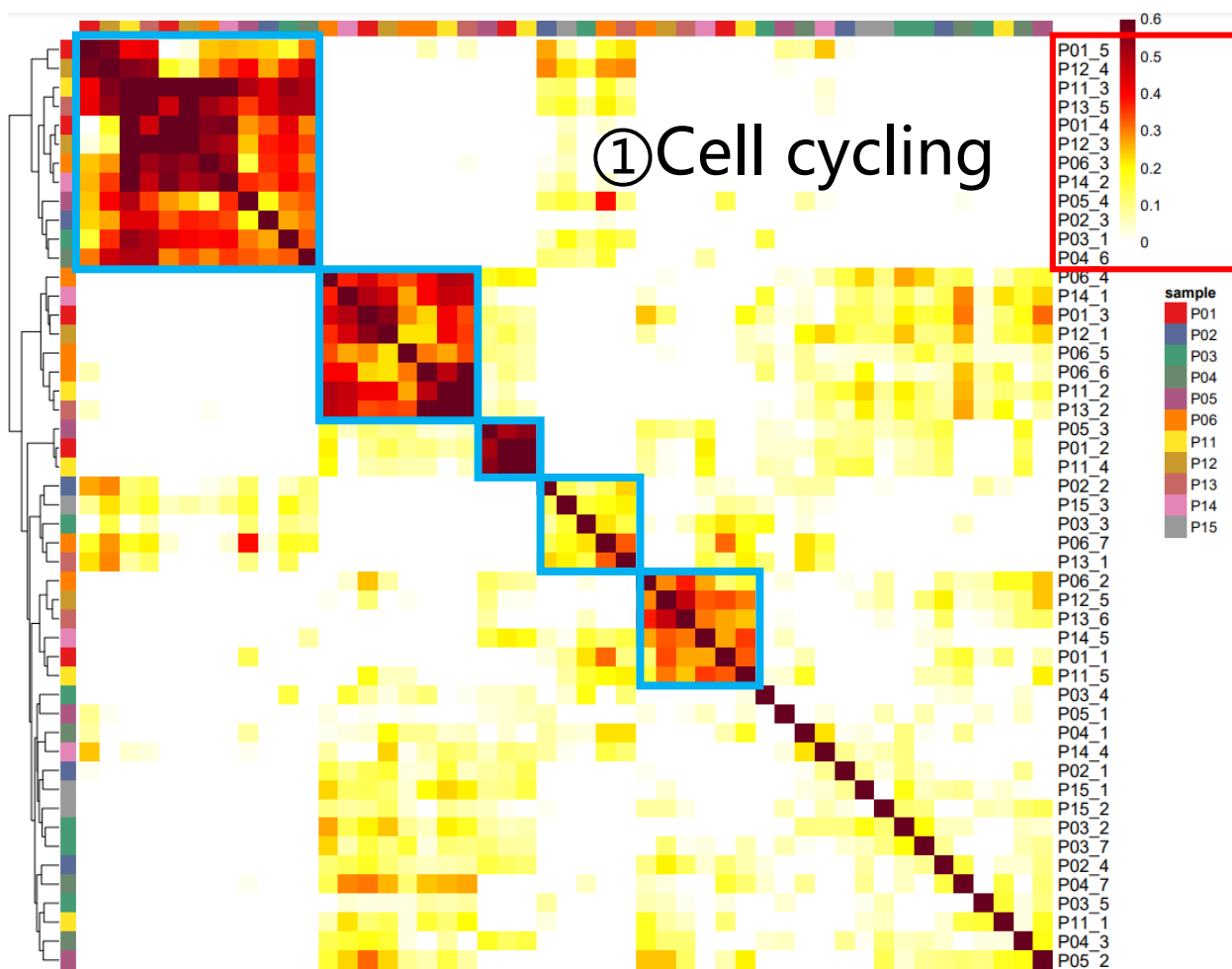
P01



P01_1	P01_2	P01_3	P01_4	P01_5	
P01_AAACCTGAGGACACCA_1	0.552231729533517	0.0254461768919148	0.0557573833382566	0.0031278171038677	0.363436893132443
P01_AAACCTGGTAGGACAC_1	0.714737911746967	0.0125802716537343	0.0739677766672148	0.0178598879064276	0.180854152625656
P01_AAACGGGAGACCTTTG_1	0	0.0603758852907579	0.19928398642568	0.163739653971292	0.57660127431227
P01_AAACGGGAGATCCTGT_1	0.696909494358213	0.018932556070729	0.284157949571058	0	0
P01_AAACGGGTGCGTTAAC_1	0.324789327723271	0.0096496886867886	0.282993315327106	0.0674654909250185	0.395102185156718
P01_AAAGATGTGCATACG_1	0.606686570644366	0.00839460776506399	0.354721229015087	0.0301975925754833	0
P01_AAAGATGTGCACGTGG_1	0	0.0191806298230162	0.110230014370605	0.232042035821153	0.638530319076432
P01_AAAGCAAAAGACATAT_1	0.254401441419626	0	0.123147318845728	0.101910417823722	0.520540821910924
P01_AAAGCAAAAGATCACGG_1	0.100183968767177	0.00485949331913561	0.335670242619522	0.111009708126517	0.448276587167649
P01_AAAGCAATCATCACCC_1	0.806212340654723	0.0338696874176007	0.142167214507434	0	0.017750757420242
P01_AAAGTAGGTATCCCAAT_1	0.608232872027736	0.173588382316198	0.209835269564719	0	0.00834347609134698
P01_AATATGCCAACACGCC_1	0.208754000394812	0.0148928317088979	0.500739974512834	0.0272758238559727	0.248337369527484
P01_AACCATGCAAGTACCT_1	0.379284554123685	0.0381492238114929	0.177497553067903	0.0986831440718786	0.306385524925041
P01_AACCATGGTCACGAA_1	0.619406917990329	0.0594916079577499	0.320952640943122	0.000148833108799072	0
P01_AACCGGAGGAGTTGC_1	0.448708174624365	0.0160164946987164	0.296524536191572	0.238750794485347	0
P01_AACCGGCACAGACTT_1	0.552523930326932	0	0.0588599903158472	0.0350863386530129	0.353529740704208
P01_AACCGGCAGATGGCA_1	0.348324196237357	0.134159604211263	0.0867492075924305	0.179660932675186	0.251106059283763
P01_AACCGGGTAAGGGCT_1	0.0987351516348518	0.350617383871684	0.123339836248117	0.109833362144814	0.317474266100533
P01_AACCGGGTCGATTGT_1	0.380006241112591	0.00391140841340751	0.149505470750949	0.361236348975465	0.105340530747588
P01_AACGTTGACACTAGT_1	0.698167931720834	0.0592891812002899	0.152318660832025	0.0902242278468507	0
P01_AACGTTGACACTGCA_1	0.108435858139506	0.0371309686020659	0.0605340790759597	0.273850400373353	0.520038813710115
P01_AACGTTGAGGATACC_1	0	0.293007908975294	0	0.2523080616595273	0.454603394429433
P01_AACTCAGCACCTATCC_1	0	0.0205759610676705	0	0.174091557386128	0.805332481546201
P01_AACTCAGTCCGTCAAA_1	0.60832832578007	0.0041581295553713	0.294477663317064	0.0130358813474951	0
P01_AACTCCGCTAGGTTT_1	0.269112314150026	0.0206895879158835	0.0289238587855277	0.215649271236068	0.465624967912495
P01_AACTCCCTCACTTATC_1	0.196475567192026	0.383133203031486	0.403314727704418	0.0170765020720608	0
P01_AACTCTTGCTCAAATC_1	0.23775862058497	0	0.144585781852855	0.4417933780083803	0.175862219478372
P01_AACTGGCTCAGTCACTA_1	0.373592129008909	0	0	0.485383375187258	0.141024495803833
P01_AACTGGTTCTACCAGA_1	0.491647847721133	0	0	0.0641261709623347	0.444225981316532
P01_AACTTTCCACCATGTA_1	0.576621628675964	0	0.416768372548884	0.00660999877515143	0
P01_AACTTTCTCACTTACT_1	0.56419770616448	0	0.373821789324154	0	0.0619805045113659
P01_AAGACCTTCATGGTCA_1	0.383602481736933	0	0.222577215943396	0	0.393820302319671
P01_AAGCCGACAGGTTCC_1	0.714812661872262	0.00755307714615673	0	0.15813905525443	0.119495205727151
P01_AAGCCGACAGTGTACT_1	0.522116056087109	0.0080354102091337	0.1035466639037652	0	0.315401059906105
P01_AAGCCGCTAACAGGA_1	0.598880487095474	0.0341446380131709	0.366066874801355	0	0
P01_AAGCCGCTGTCAGATT_1	0.215384783180651	0.104908732203371	0.244603150276192	0.082672601210032	0.352430733208054
P01_AAGCCGCTGCTAGGAG_1	0.353483982595959	0.0159689902730099	0.0512854713733761	0.162892727147753	0.416368828609093
P01_AAGCCGCTCTACTCAT_1	0.505743337278812	0.00220458413884705	0	0.303640609721358	0.188411468060983
P01_AAGGCAGGTGTCAGTAG_1	0	0	0.757220923423007	0	0.242779076576993
P01_AAGGCAGGTTCCGGCT_1	0.628675797849862	0.155508847268488	0.0761513036935701	0	0.13966405118808
P01_program.usage.norm.txt					

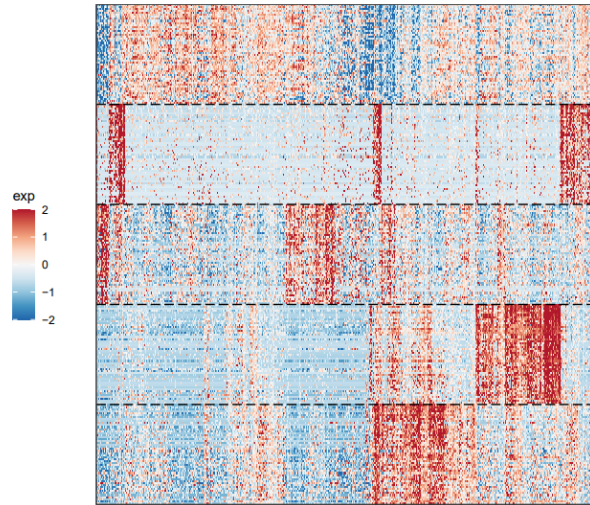
P06





A	B	C	D
program	pathway		
P01_5	Cell Cycle	E2F_TARGETS	DNA_REPAIR
P12_4	Cell Cycle	E2F_TARGETS	MYC_TARGETS_V1
P11_3	G2M_CHECKPOINT	Cell Cycle	E2F_TARGETS
P13_5	Cell Cycle	E2F_TARGETS	G2M_CHECKPOINT
P01_4	G2M_CHECKPOINT	Cell Cycle	E2F_TARGETS
P12_3	Cell Cycle	G2M_CHECKPOINT	E2F_TARGETS
P06_3	Cell Cycle	G2M_CHECKPOINT	E2F_TARGETS
P14_2	G2M_CHECKPOINT	Proliferation	E2F_TARGETS
P05_4	MYC_TARGETS_V1	OXIDATIVE_PHOSPHORYLATION	Invasion
P02_3	Cell Cycle	G2M_CHECKPOINT	MITOTIC_SPINDLE
P03_1	G2M_CHECKPOINT	Cell Cycle	E2F_TARGETS
P04_6	Cell Cycle	E2F_TARGETS	G2M_CHECKPOINT
P06_4	KRAS_SIGNALING_UP	Hypoxia	TNFA_SIGNALING_VIA_NFKB
P14_1	COMPLEMENT	Metastasis	Quiescence
P01_3	INTERFERON_ALPHA_RESPONSE	INTERFERON_GAMMA_RESPONSE	TNFA_SIGNALING_VIA_NFKB
P12_1	TNFA_SIGNALING_VIA_NFKB	INTERFERON_GAMMA_RESPONSE	UV_RESPONSE_UP
P06_5	Hypoxia	GLYCOLYSIS	HYPOXIA
P06_6	APOPTOSIS	COMPLEMENT	PI3K_AKT_MTOR_SIGNALING
P11_2	TNFA_SIGNALING_VIA_NFKB	APOPTOSIS	Metastasis
P13_2	Hypoxia	Metastasis	HYPOXIA
P05_3	ALLOGRAFT_REJECTION	Quiescence	KRAS_SIGNALING_UP
P01_2	ALLOGRAFT_REJECTION	APICAL_JUNCTION	COMPLEMENT
P11_4	ALLOGRAFT_REJECTION	IL2_STAT5_SIGNALING	INTERFERON_GAMMA_RESPONSE
P02_2	EPITHELIAL_MESENCHYMAL_TRANSITION	OXIDATIVE_PHOSPHORYLATION	Invasion
P15_3	EPITHELIAL_MESENCHYMAL_TRANSITION	KRAS_SIGNALING_DN	MYOGENESIS
P03_3	CHOLESTEROL_HOMEOSTASIS	ALLOGRAFT_REJECTION	ESTROGEN_RESPONSE_LATE
P06_7	MYC_TARGETS_V1	ALLOGRAFT_REJECTION	DNA damage
P13_1	ALLOGRAFT_REJECTION	IL6_JAK_STAT3_SIGNALING	Inflammation
P06_2	KRAS_SIGNALING_DN	APICAL_SURFACE	Apoptosis
P12_5	TNFA_SIGNALING_VIA_NFKB	Hypoxia	HYPOXIA
P13_6	Stemness	INTERFERON_ALPHA_RESPONSE	ESTROGEN_RESPONSE_EARLY
P14_5	MYC_TARGETS_V1	ALLOGRAFT_REJECTION	P53_PATHWAY
P01_1	MYC_TARGETS_V1	OXIDATIVE_PHOSPHORYLATION	DNA damage
P11_5	MYC_TARGETS_V1	P53_PATHWAY	ALLOGRAFT_REJECTION
P03_4	Metastasis	EPITHELIAL_MESENCHYMAL_TRANSITION	COMPLEMENT
P05_1	XENOBIOTIC_METABOLISM	Proliferation	COAGULATION
P04_1	TNFA_SIGNALING_VIA_NFKB	MYOGENESIS	COAGULATION
P14_4	INFLAMMATORY_RESPONSE	ALLOGRAFT_REJECTION	TNFA_SIGNALING_VIA_NFKB
P02_1	INTERFERON_ALPHA_RESPONSE	INTERFERON_GAMMA_RESPONSE	INFLAMMATORY_RESPONSE
P15_1	INTERFERON_GAMMA_RESPONSE	INTERFERON_ALPHA_RESPONSE	APOPTOSIS
P15_2	ANDROGEN_RESPONSE		

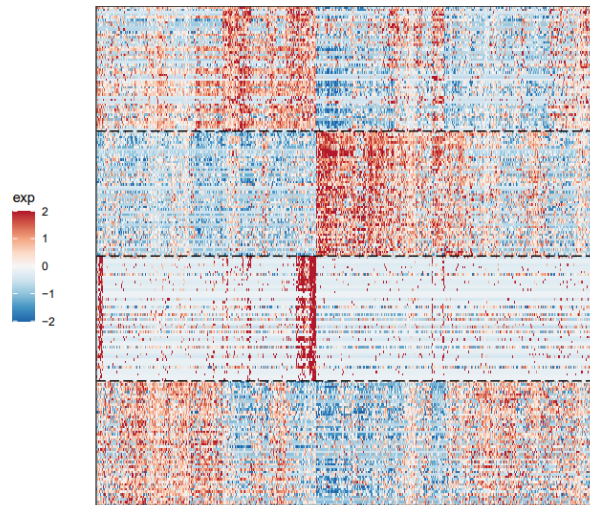
P01: 919 cells; 5 programs



P01_1; P01_2; P01_3; P01_4; P01_5

RPL4	BTF3	RPL5	NACA	YBX3	RPL19	RPL3	EEF1A1	TSPO	RPL13A
RACK1	LDHB	RPS3	PEBP1	RPL7	EEF2	RPL7A	RPL23	MDH2	EEF1B2
RPS3A	RPS6	RPS8	RPL6	AHCY	RPS7	NPM1	ATP5F1B	RPS14	NPM3
RPL35A	RPL10	RPS4X	RPL17	SLC25A3	BAG1	RPS25	EBPL	SLC25A6	RPS27A
RPL9	EIF3K	RPL10A	RPLP0	VDAC1	EIF3E	HSP90AB1	ATP5MC3	RPL14	EIF4B
TMSB4X	CD3D	CORO1A	COTL1	CD2	ARHGD1B	IFITM2	CD53	TRAC	CD48
PTPRC	HCST	LAG3	CD3E	CD7	NKG7	LCK	GMFG	GPSM3	TIGIT
CD3G	EVL	ACAP1	CS77	SLA	S100A4	CLEC2D	SAMSN1	CD8A	SIT1
SH2D1A	HCLS1	GZMB	CD37	GZMK	ITGB2	LAT	CYTOR	CD247	FYB1
ITM2A	FMNL1	CRIP1	CD27	FERMT3	HAVCR2	TRBC2	FABP5	GIMAP7	DNAJC15
TNFSF10	OAS1	SFN	CLEC2B	S100A11	IFI44L	IFI44	DDIT4	CSTB	IFITM1
UBALD2	DDIT3	SAT1	BTG1	IFI6	NUPR1	HCAR3	S100A6	IRF7	GNPMB
EFNA1	CDKN1A	C4orf3	ISG15	SEMA4B	VEGFA	NEURL3	SQSTM1	ITGA3	IFI35
PPM1K	PLP2	IFI16	SP110	PLSCR1	IRF6	RNF19B	HCAR2	ITM2B	ID1
ETS2	GBP5	CRABP2	IL1RN	TGIF1	TACSTD2	PLAU	RTP4	CD9	CMPK2
CCNB1	UBE2C	PLK1	KPNA2	TOP2A	CDC20	CCNB2	CENPF	AURKA	TPX2
ARL6IP1	PTTG1	NUSAP1	BIRC5	CKS1B	CKS2	CENPA	UBE2S	HMMR	CDC43
CENPE	FAM83D	TUBA1C	DEPDC1	CCNA2	KNSTRN	NUF2	ASPM	HMGB2	PRR11
CDKN3	AURKB	CALM2	TUBB4B	PSRC1	CDK1	PRC1	KIF23	MKI67	KIF2C
NEK2	TACC3	GTSE1	JPT1	KIF20B	TROAP	SGO2	NUCKS1	RACGAP1	CEP55
TK1	DUT	TYMS	ASF1B	RPA3	PCLAF	PSMC3IP	RRM2	CLSPN	TMEM163
FAM111B	RFC2	ZWINT	PCNA	CDC6	DHFR	CDC45	GINS2	MYBL2	FEN1
DTL	CENPU	HSPB11	CCDC28B	RFC4	TMEM106C	UBE2T	PXMP2	CHCHD2	MCM5
SIVA1	RRM1	HELLS	CENPX	TAGLN2	ACAT2	TUBG1	CENPM	EBP	USP18
CSRP2	MCM7	SLBP	FDPS	WDR34	ORC6	SVIP	DSN1	H2AFZ	RAD51AP1

P13: 1501 cells; 4 programs



P13_1; P13_2; P13_5; P13_6

RPS27A	SOCS1	ITIH5	SRGN	RPL18A	CSTA	PRRX1	TNFAIP2	IL2RA	NCF4
LITAF	RPS16	SH3PXD2A	VCAM1	LGALS1	WDR91	RPL28	RPLP1	TYK2	OGFRL1
RPS17	ICAM1	IL27RA	UBD	CD70	RPL37A	RPS5	PAPPA	GYPC	LINC01127
RPS23	RPS18	PARP1	RPL8	RPL23A	RPS10	EBI3	STAR	IL1R2	RPL27
ANGPT2	RPS24	RPS2	CHCHD6	TMEM14A	CCND1	RPS19	RPS15	RPS28	NRP2
TMSB10	SAT1	SFN	CSTB	SOD2	S100A11	S100A10	KRT6A	S100A9	S100A8
IGFBP3	PKM	PDZK1IP1	CTSB	IL1RN	S100A6	LAMC2	CARHSP1	TMPPRSS11D	CD9
NEAT1	TPH1	SDCBP2	PLAT	CD47	ADAM28	LYN	SERINC2	ANXA2	SERPINB1
LDHA	TACSTD2	MALL	LAMB3	ERO1A	EIF1	DNAJB6	MMP7	WFDC2	FBTH1
CFL1	TPM4	MYL12A	CDKN1A	FAM3C	CLIC1	IL36G	SOCS3	MMP10	CXADR
HMGB2	PBK	MKI67	PCLAF	CDK1	TOP2A	NUSAP1	TUBA1B	ESCO2	TK1
RAD51AP1	TYMS	UBE2C	ZWINT	KIF4A	DEPDC1	GTSE1	UBE2T	PRC1	CEP55
H2AFZ	BIRC5	NUF2	HIST1H4C	CCNA2	CKS1B	AURKB	CENPF	CKS2	KIF2C
HMG2N	CENPA	MAD2L1	CENPU	UXM1	KIF23	DUT	RRM2	TPX2	ATAD2
CENPW	CENPE	CDC43	TACC3	RANBP1	ASPM	CCNB1	KIFC1	HMMR	ANLN
SOX4	TXNIP	KRT15	BCL11A	ZKSCAN1	IFI44L	IFI27	ARHGAP24	ARL4C	CARS
BCL2	DAPL1	NTRK2	PABPC1	H3F3A	TNRC6B	RASSF6	PCDH7	MSI2	WASF2
SEMA6D	STMN1	IGFBP2	FUS	GCLC	WINK2	TNFSF10	NAV2	RHOBTB3	PBX1
MAPK10	RPL10A	BAZ2B	EEF1A1	ZBTB20	RPL10	NSD3	MEX3A	APEX1	RPS4X
FTL	TRIM38	MARCKSL1	SYNPO2	RPL13A	FEZ2	SMAD5	SLC38A1	GABPB1.AS1	GTF2I

manually (Fig. 2c; Supplementary information, Table S5). For each signature, we then combined the top 100 genes of each metagene and calculated the average loadings for each gene. We summarized the total loadings for repetitive genes, retained the original loadings for exclusive genes, and divided the loadings of each gene by the number of metagenes within the signature. Finally, the top 30 genes with the highest loading were defined as the marker genes for the signature.