**Identifying methylation biomarks of cervical cancer**

Data Collection

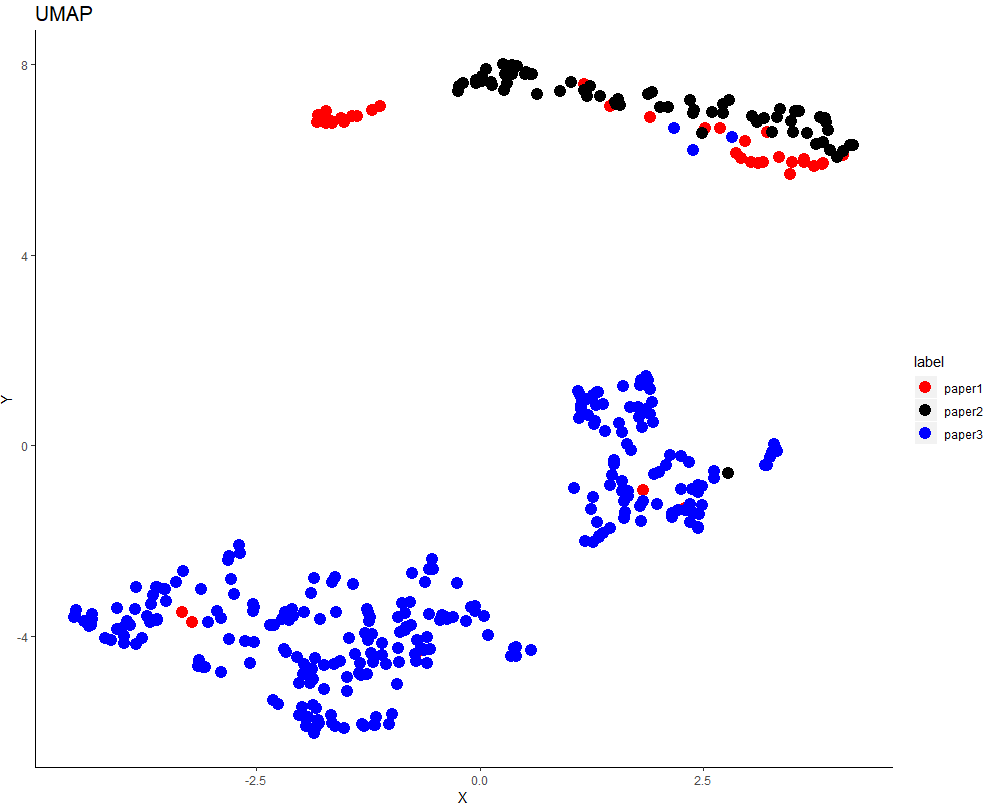
All the data are extracted from 3 papers via searching GEO of NCBI by Xihui Yao:

1. Farkas S A, Milutin-Gašperov N, Grce M, et al. Genome-wide DNA methylation assay reveals novel candidate biomarker genes in cervical cancer[J]. Epigenetics, 2013, 8(11): 1213-1225.
2. Verlaat W, Snoek B C, Heideman D A M, et al. Identification and validation of a 3-gene methylation classifier for HPV-based cervical screening on self-samples[J]. Clinical Cancer Research, 2018, 24(14): 3456-3464.
3. Lando M, Fjeldbo C S, Wilting S M, et al. Interplay between promoter methylation and chromosomal loss in gene silencing at 3p11-p14 in cervical cancer[J]. Epigenetics, 2015, 10(10): 970-980.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Reference | platform |  |  |  |  |  |  |  |  |
| normal | Pre-Cancer | cancer | Total |  |  |  |  |
| Ref01 | 450K | 20 | 18(CIN3) | 6 | 48 |  |  |  |  |
| Ref02 | 450K | 28 | 36(CIN3) | 4(SCC) | 68 |  |  |  |  |
| Ref03 | 450K+? | 270 cancer: 9-2A; 5-3A; 13-4A;  8-1B1; 15-1B2; 164-2B; 56-3B | | | 270 |  |  |  |  |
| TCGA | 450 | 299 | 3 | | 302 |  |  |  |  |
|  |  | Overlap probe number：99964 | | | | | | | |

Checking the system variance

左图所示为3篇文献数据的聚类结果。Paper3对应的蓝色样本点明显分为2个cluster，因为paper3的数据出自两个不同技术平台。Paper3和其它两篇文献的差异也很显著。差异之源包含生物本身差异，但更多的是实验方法导致的差异。所以这3篇paper的数据如果简单地pool到一起分析，分析的结果通常 with high FDR



由于paper3采用了不同于450k之外的技术平台，故将其数据排除在我们集成分析之外。我们尝试将TCGA的数据加入到我们的分析。