Assign: a software for single-cell genomic data integration

Weiqiang Zhou

Assistant Scientist

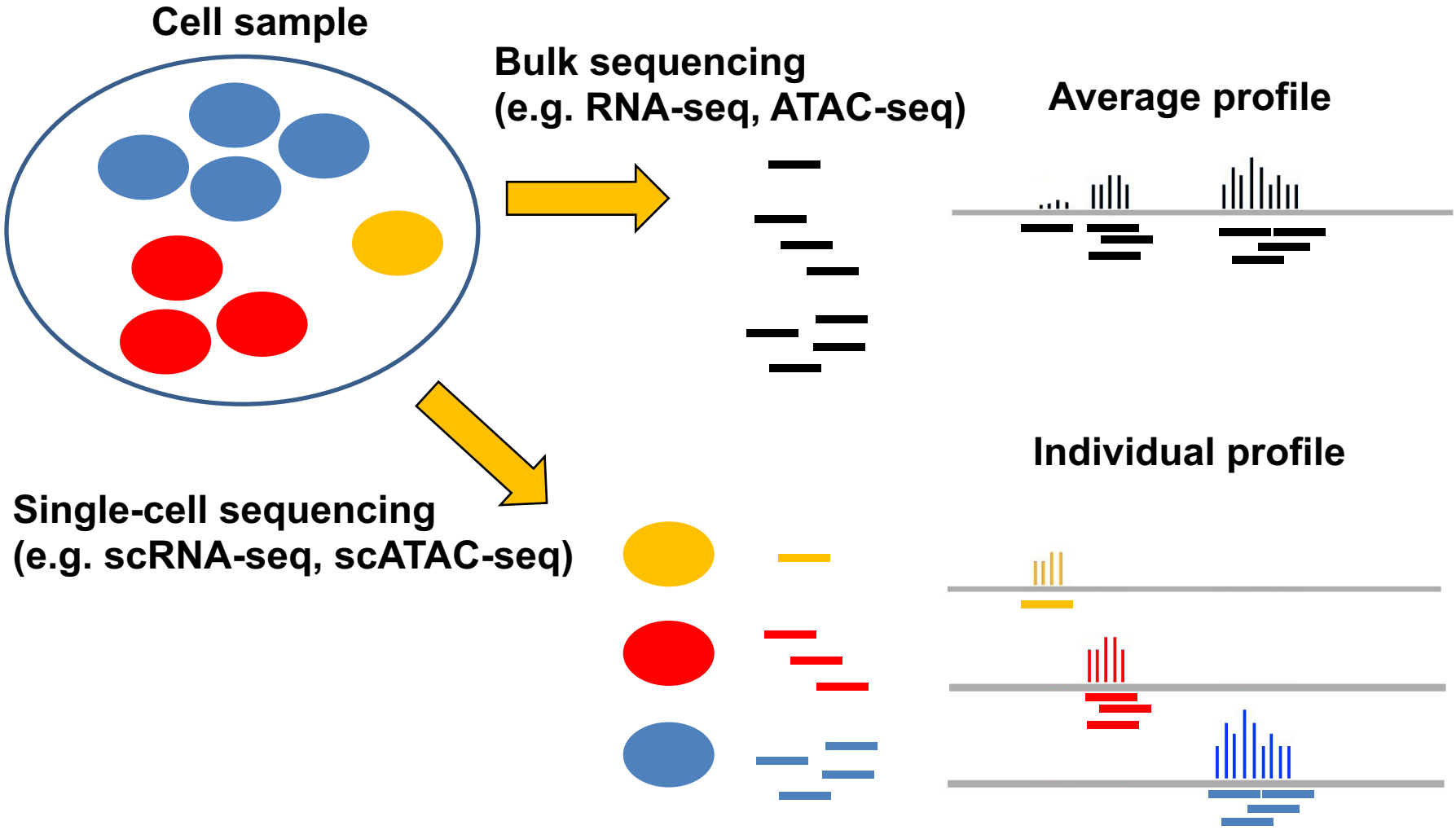
Department of Biostatistics

JOHNS HOPKINS

BLOOMBERG SCHOOL

of PUBLIC HEALTH

From bulk to single cell



Available single cell assays

10x genomics

- Single-cell RNA-seq
- Single-cell ATAC-seq
- Single cell immune profiling
- Single cell CNV

Fluidigm

- Single-cell RNA-seq
- Single-cell ATAC-seq
- Single-cell DNA-seq

Single cell multiomics

Macaulay *et al*. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nature Methods* (2015). (G&T-seq: genomic DNA and gene expression)

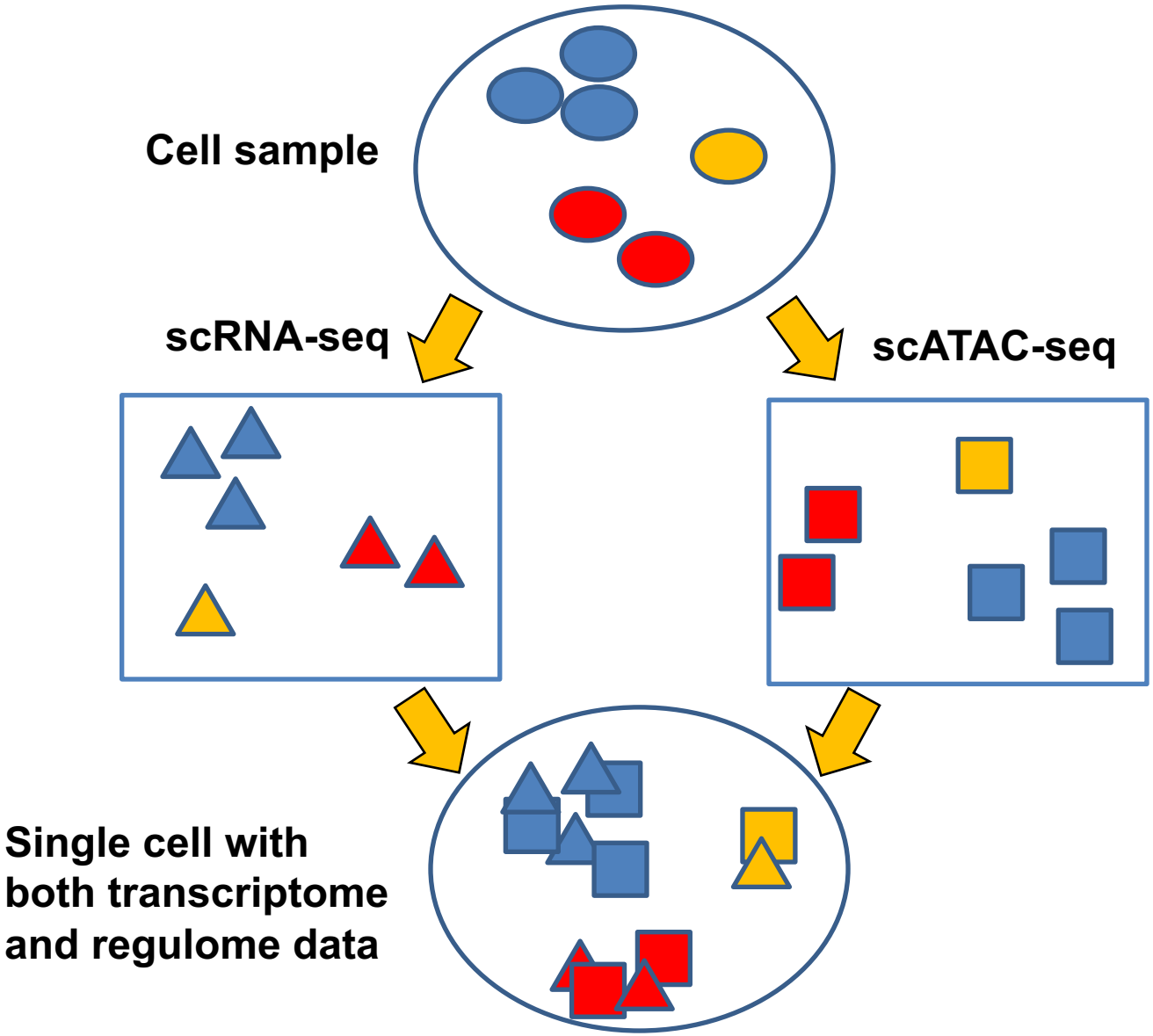
Hu *et al*. Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biology* (2016). (scMT-seq: DNA methylation and gene expression)

Cheow *et al*. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nature Methods* (2016). (sc-GEM: genotype, DNA methylation, and gene expression; use Fluidigm system)

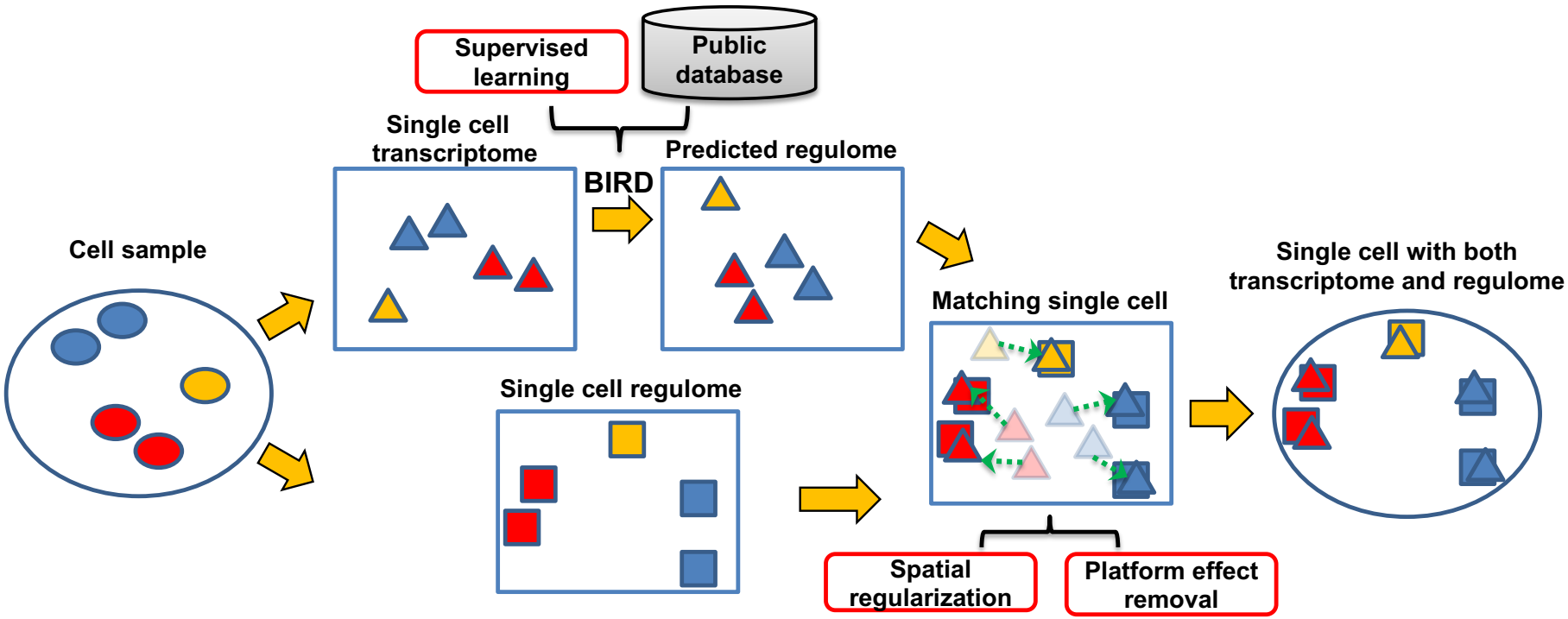
Clark *et al*. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nature Communications (2018). (scNMT-seq: chromatin accessibility, DNA methylation, and gene expression)

Cao, J. *et al*. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* (2018). (sci-CAR: chromatin accessibility, and gene expression)

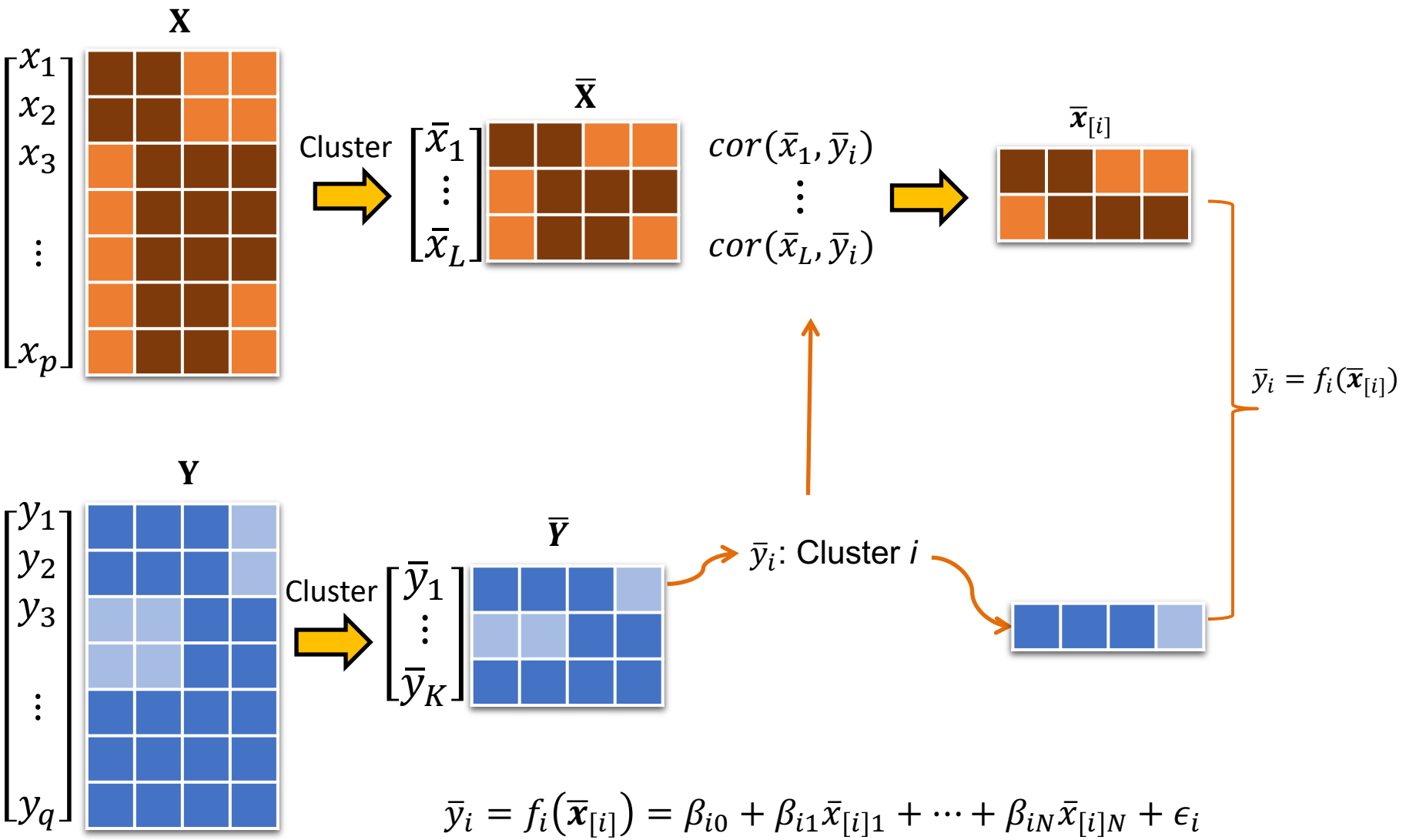
Integrate multiple data types



Overview



BIRD



$$\bar{y}_i = f_i(\bar{\boldsymbol{x}}_{[i]}) = \beta_{i0} + \beta_{i1}\bar{x}_{[i]1} + \cdots + \beta_{iN}\bar{x}_{[i]N} + \epsilon_i$$

Platform effect removal

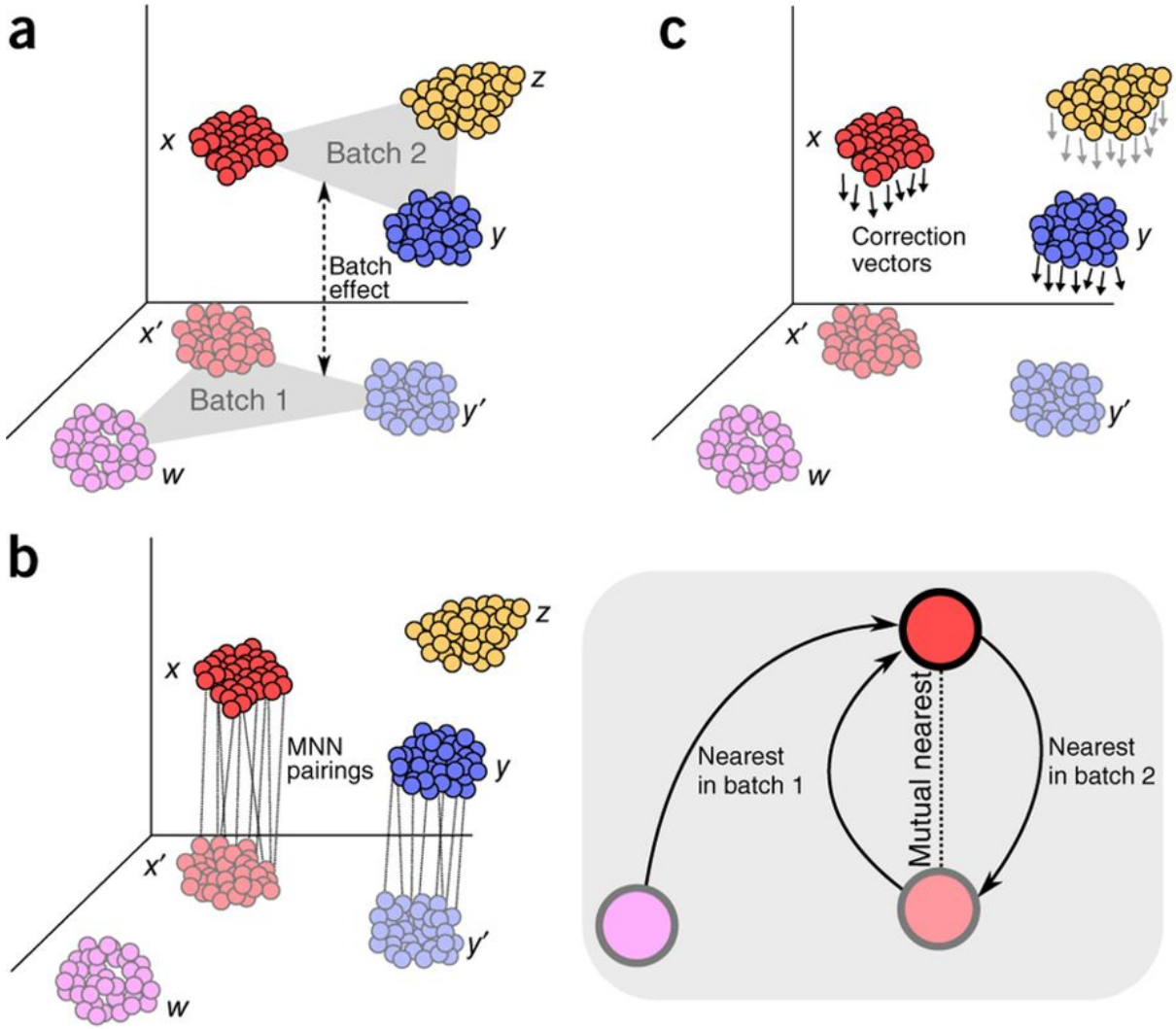


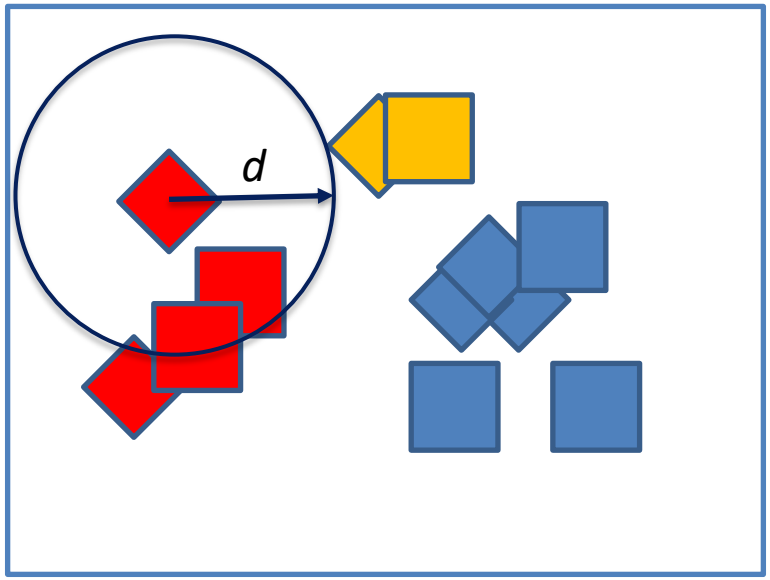
Figure obtained from Haghverdi *et al*. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology* **36**, 421–427 (2018)

Spatial regulation

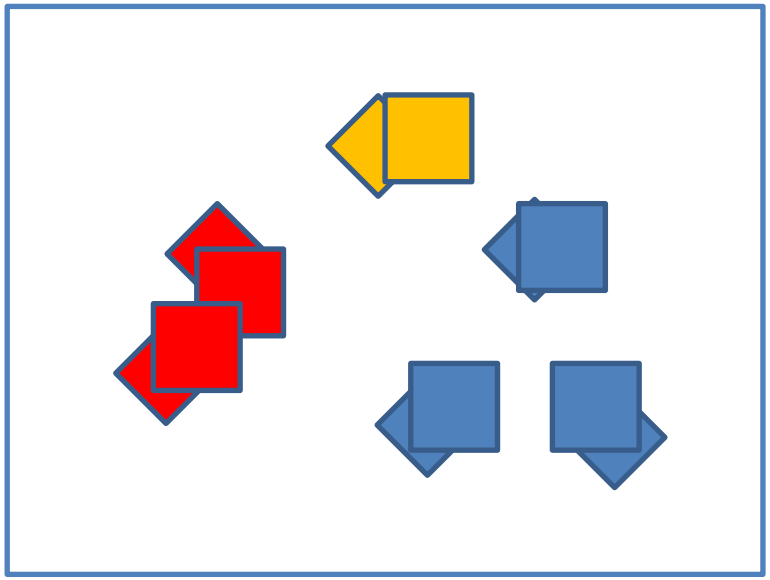
Predicted scATAC

True scATAC

Before optimization



After optimization

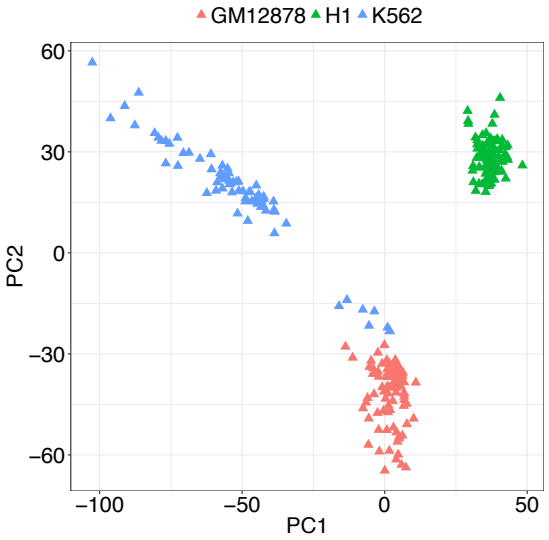


Fisher's extract test:

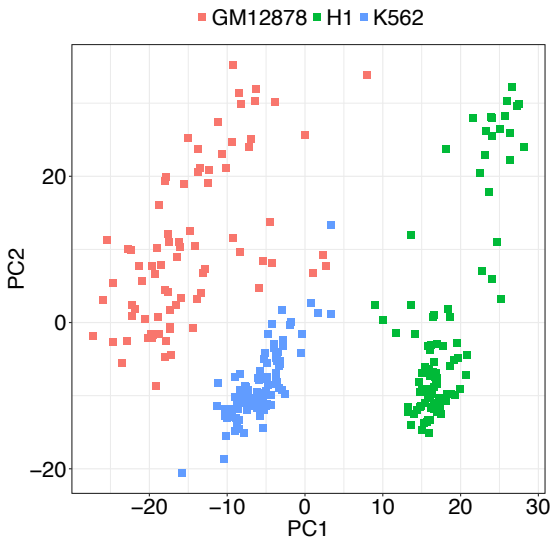
$N_{P\_obs}/ N_{T\_obs} = N_P/ N_T$

Application to a test dataset

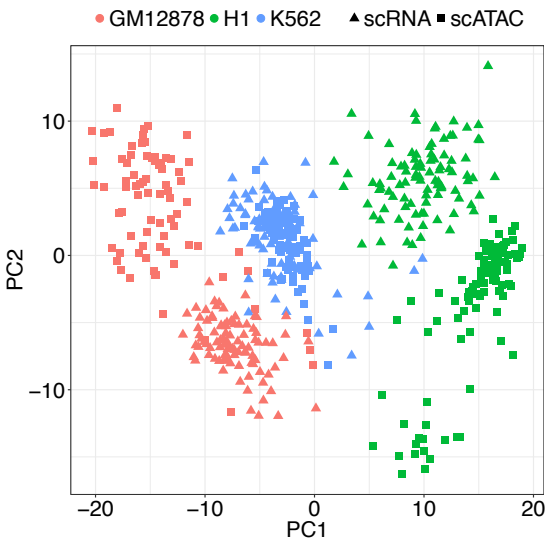
scRNA-seq



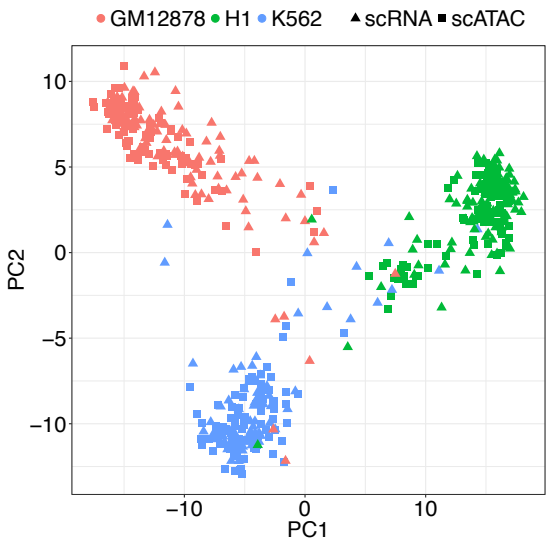
scATAC-seq



Predicted + True scATAC-seq

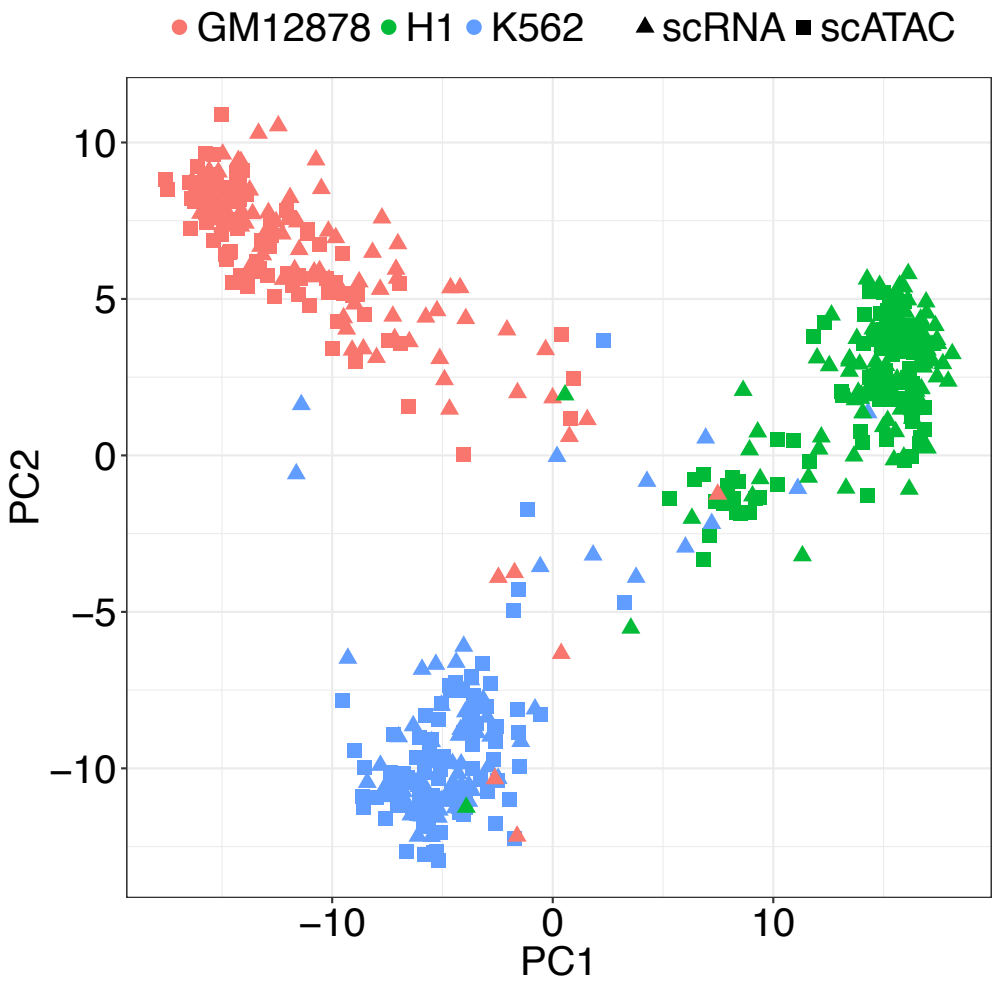


Assign

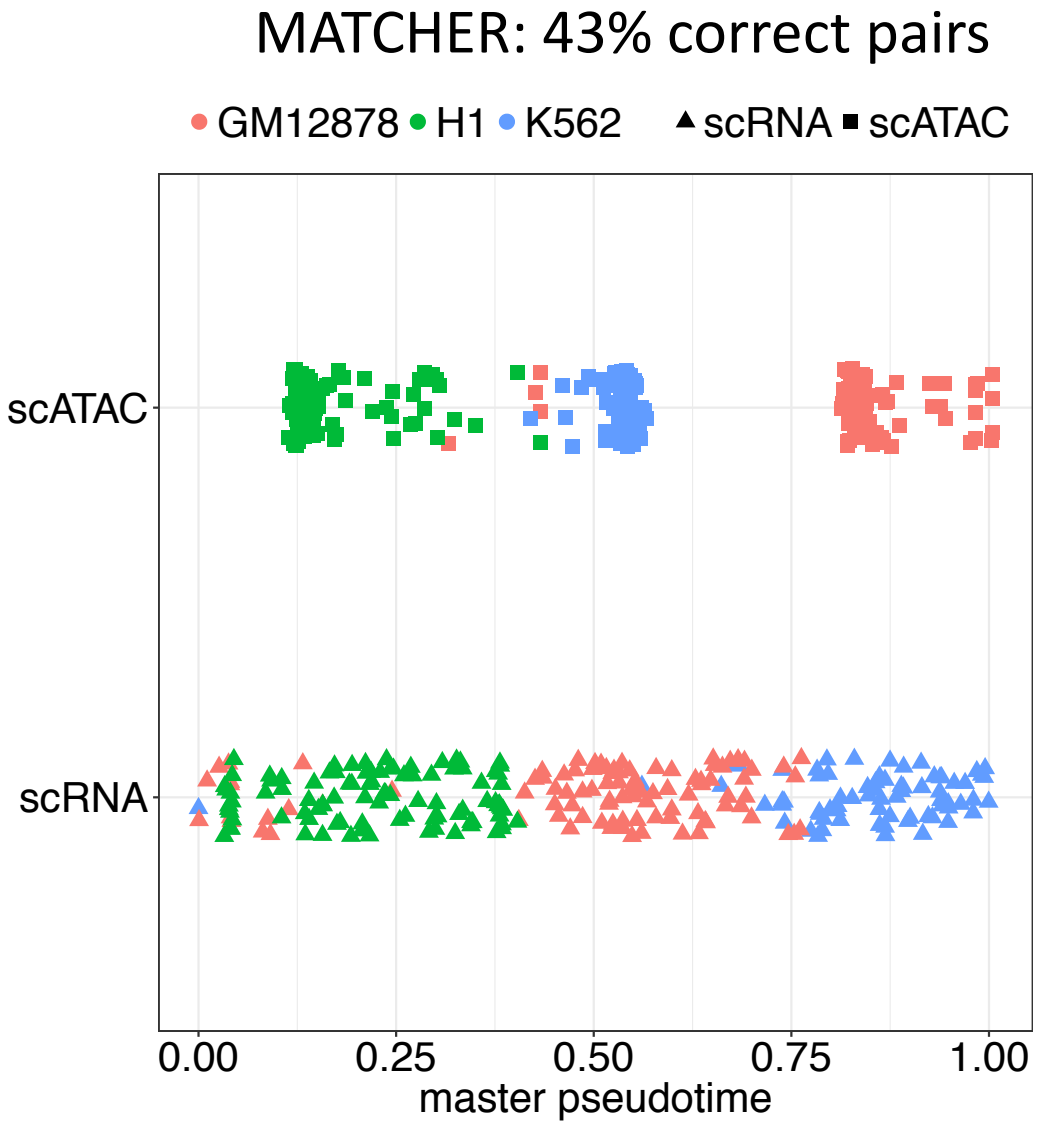


Comparison with other methods

ASSIGN: 94% correct pairs

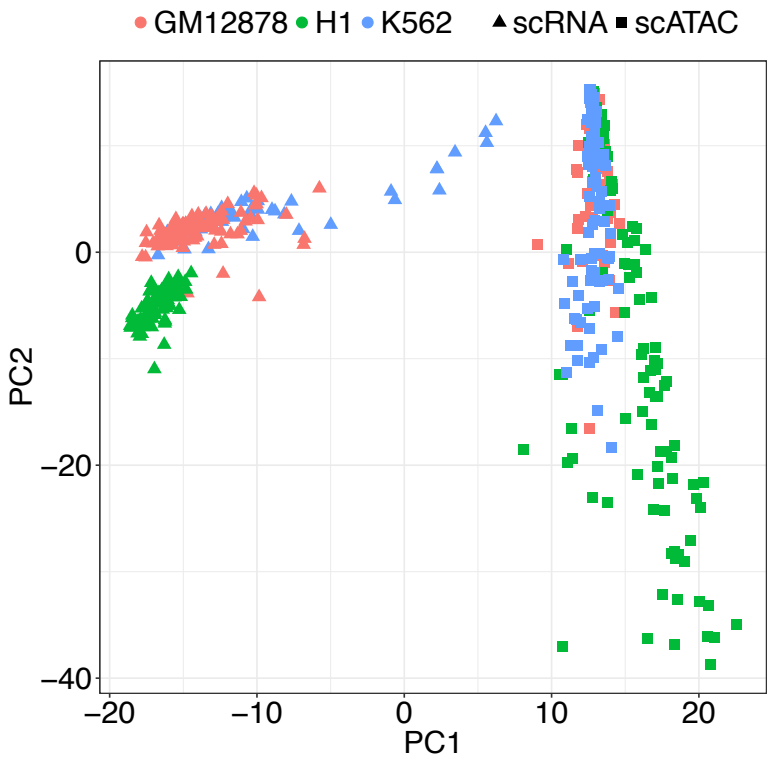


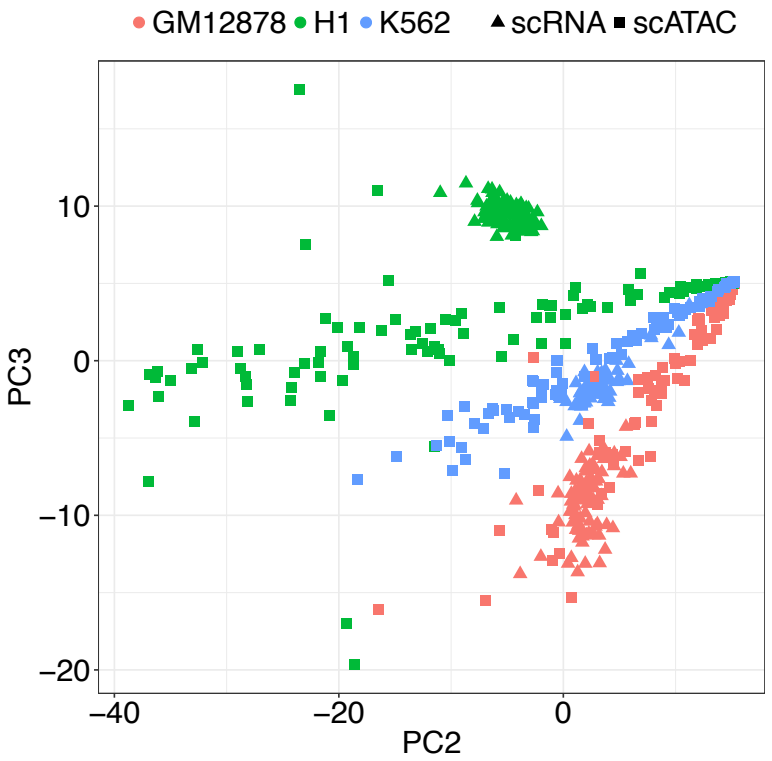
Comparison with other methods



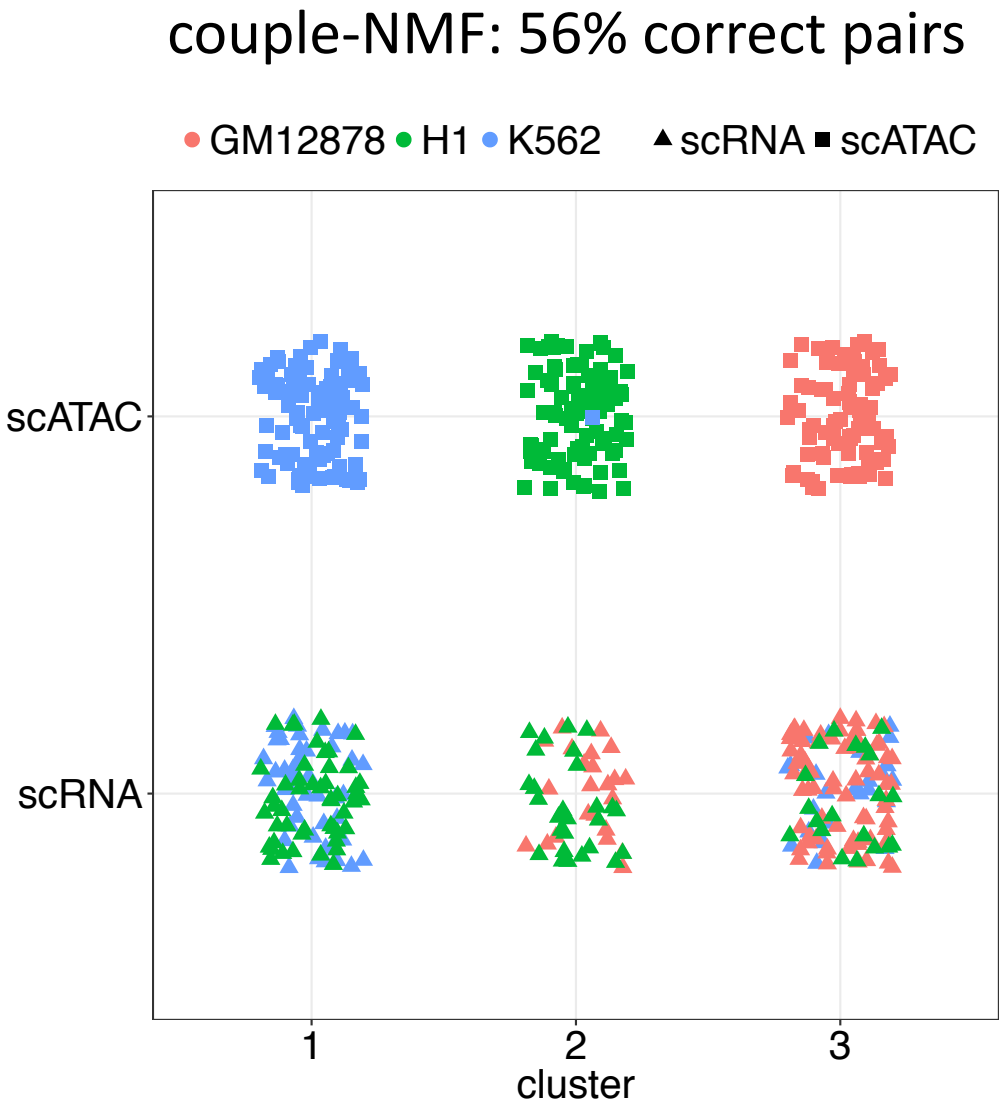
Comparison with other methods

Cicero: 79% correct pairs

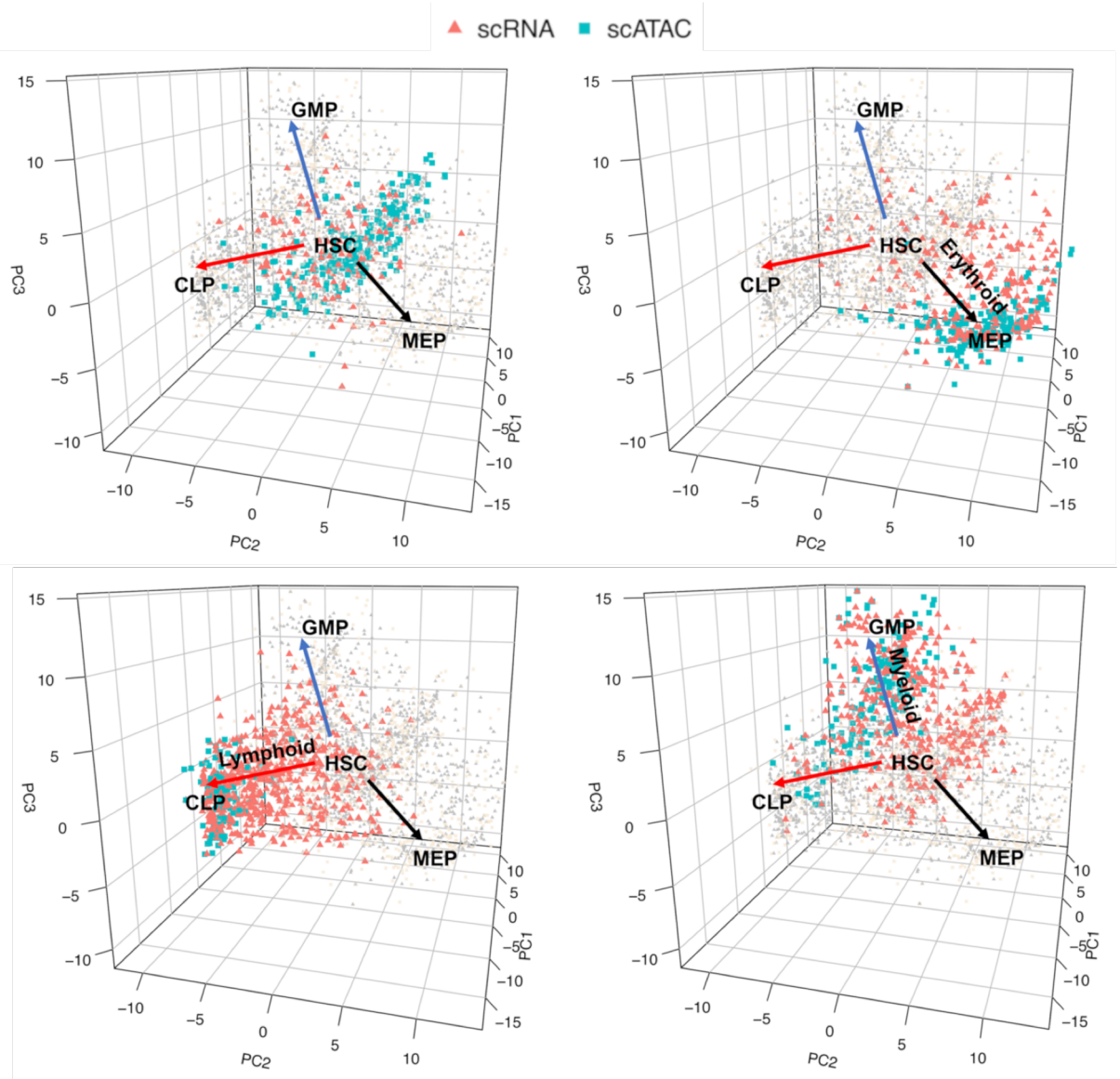




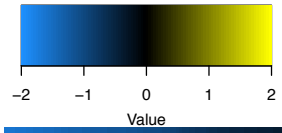
Comparison with other methods

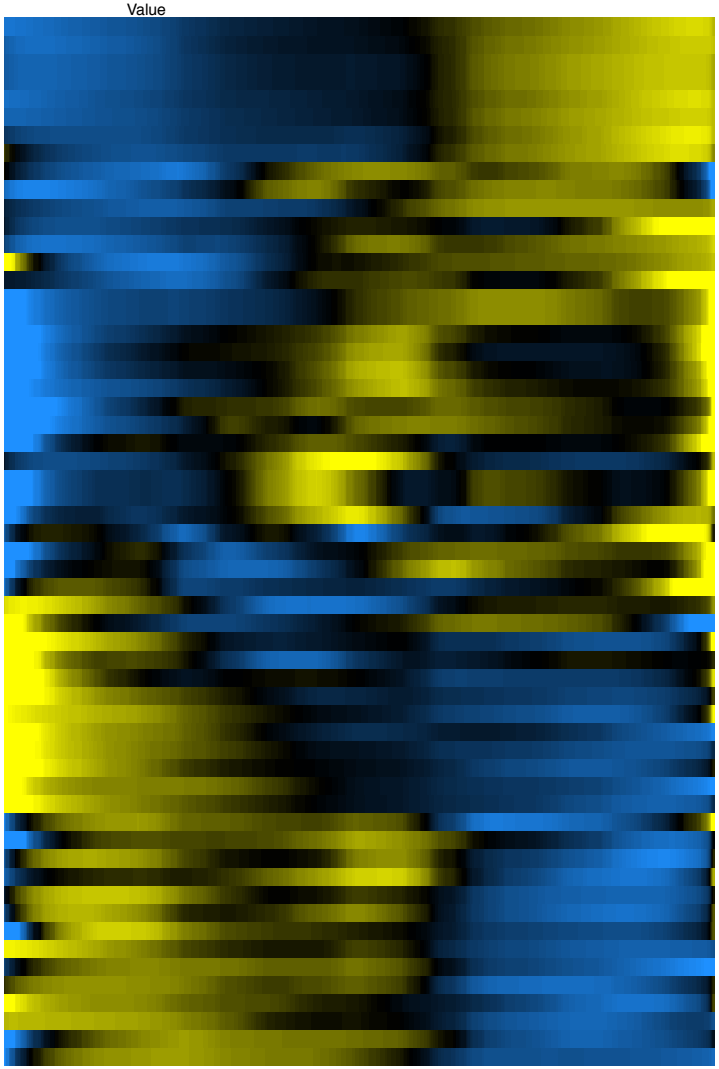


Application to HCA data

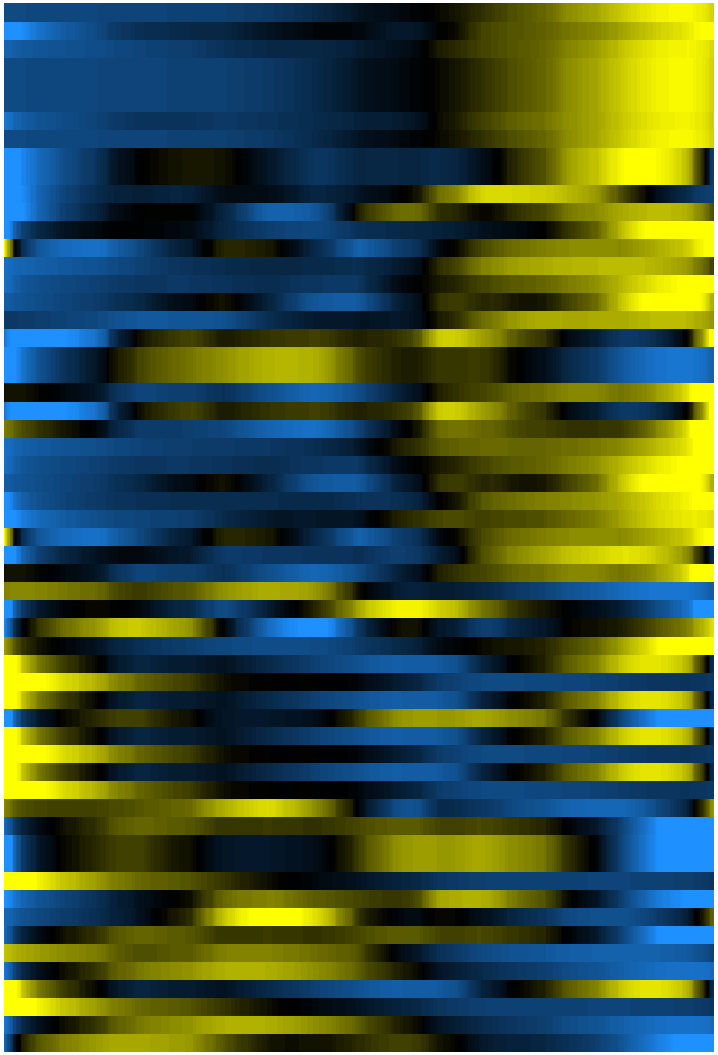


Application to HCA data





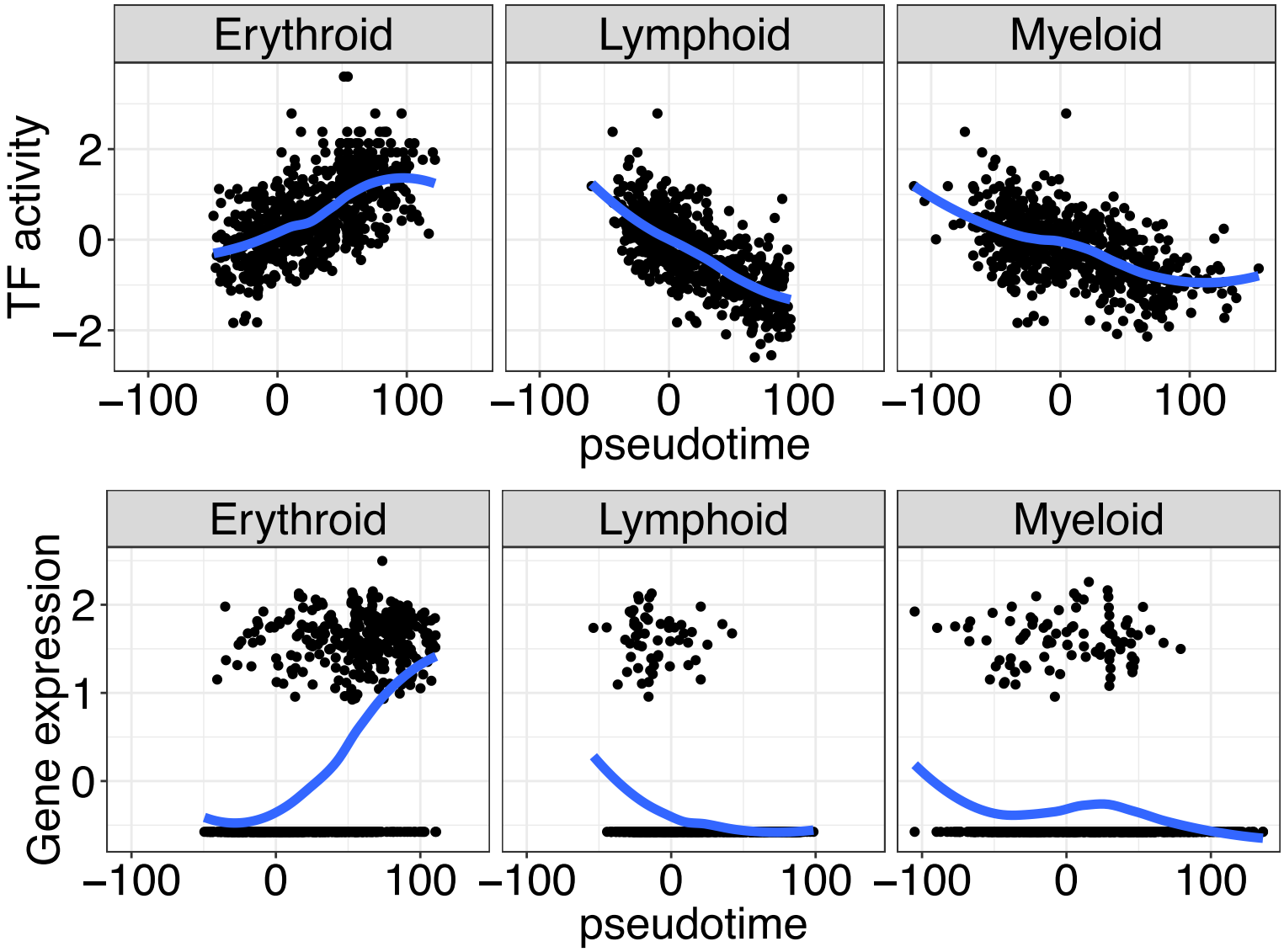
MA0035.3:Gata1
MA0036.2:GATA2
MA0140.2:GATA1::TAL1
MA0140.2:GATA1::TAL1
M00346_GATA1_05
M00128_GATA1_04
M00278_LMO2COM_02
M00126_GATA1_02
M00152_SRF_01
MA0083.3:SRF
MA0151.1:Arid3a
MA0852.1:Foxk1
M00791_HNF3_Q6
M00980_TBP_Q6
M00477_FOXO3_01
MA0089.1:MAFG::NFE2L1
MA0089.1:MAFG::NFE2L1
MA0462.1:BATF::JUN
MA0491.1:JUND
MA0476.1:FOS
MA0099.2:FOS::JUN
MA0795.1:SMAD3
MA0492.1:JUND_var.2
M00474_FOXO1_02
MA0501.1:MAF::NFE2
M00284_TCF11MAFG_01
M00284_TCF11MAFG_01
MA0770.1:HSF2
MA0797.1:TGIF2
MA0108.2:TBP
MA0152.1:NFATC2
M00701_SMAD3_Q6
M00454_MRF2_01
MA0605.1:Atf3
M00794_TTF1_Q6
MA0865.1:E2F8
M00222_HAND1E47_01
MA0014.2:PAX5
M00804_E2A_Q2
M00159_CEBP_01
M00002_E47_01
M00144_PAX5_02
MA0522.2:TCF3
M00143_PAX5_01
M00747_IRF1_Q6
MA0836.1:CEBPD
M00912_CEBP_Q2_01
MA0102.3:CEBPA
M00977_EBF_Q6
MA0518.1:Stat4
MA0772.1:IRF7
M00621_CEBPDELTA_Q6
MA0081.1:SPIB
M00699_ICSBP_Q6
M00071_E47_02
MA0154.3:EBF1
MA0652.1:IRF8
MA0080.4:SPI1



GATA1
GATA2
TAL1
GATA1
GATA1
GATA1
LMO2
GATA1
SRF
SRF
ARID3A
FOXK1
FOXM1
TBP
FOXO3
NFE2L1
MAFG
BATF
JUND
FOS
FOS
SMAD3
JUND
FOXO1
NFE2
NFE2L1
MAFG
HSF2
TGIF2
TBP
NFATC2
SMAD3
ARID5B
ATF3
TTF1
E2F8
TCF3
PAX5
TCF3
CEBPA
TCF3
PAX5
TCF3
PAX5
IRF1
CEBPD
CEBPA
CEBPA
EBF1
STAT4
IRF7
CEBPD
SPIB
IRF8
TCF3
EBF1
IRF8
SPI1

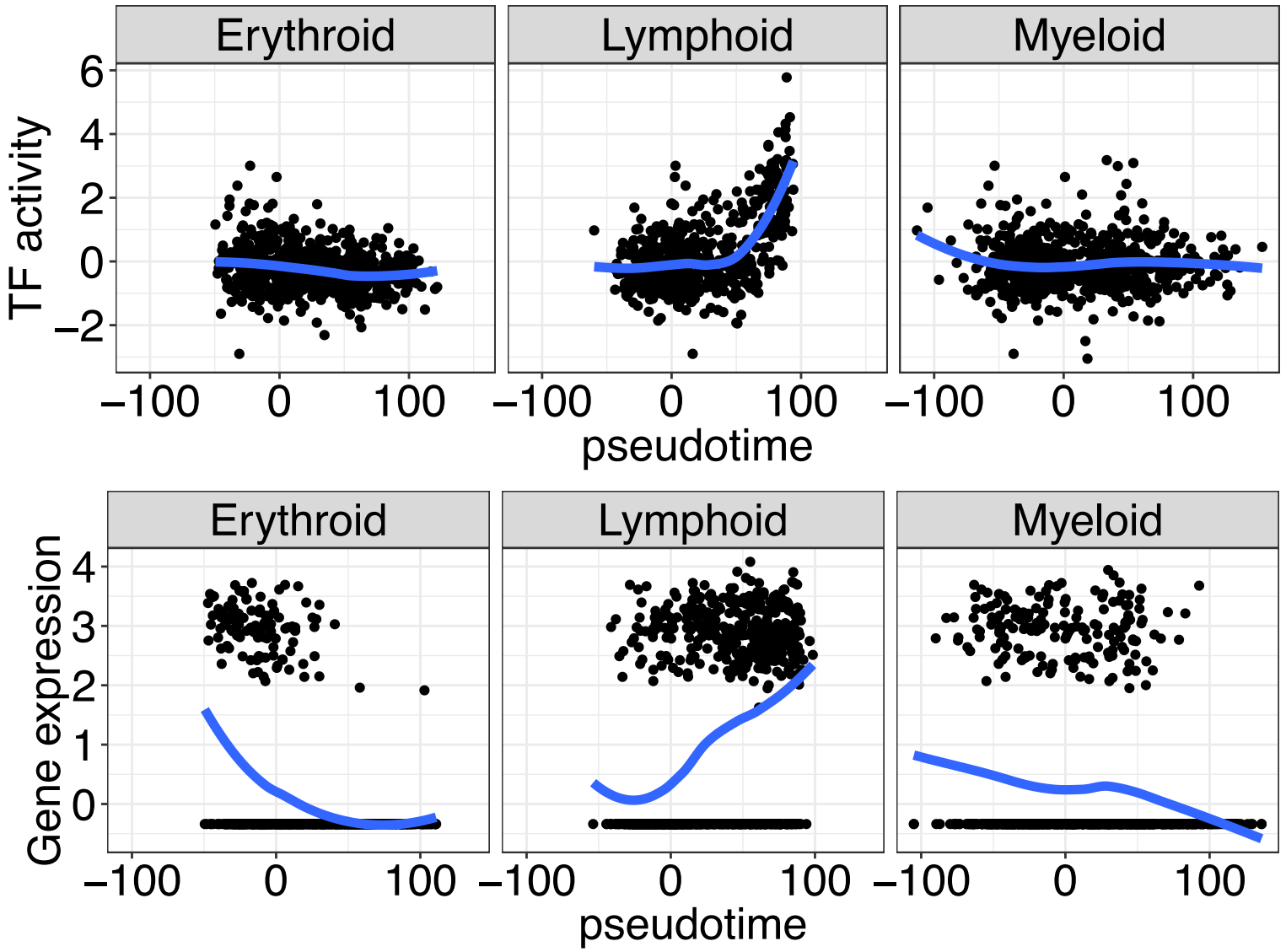
Application to HCA data

GATA1



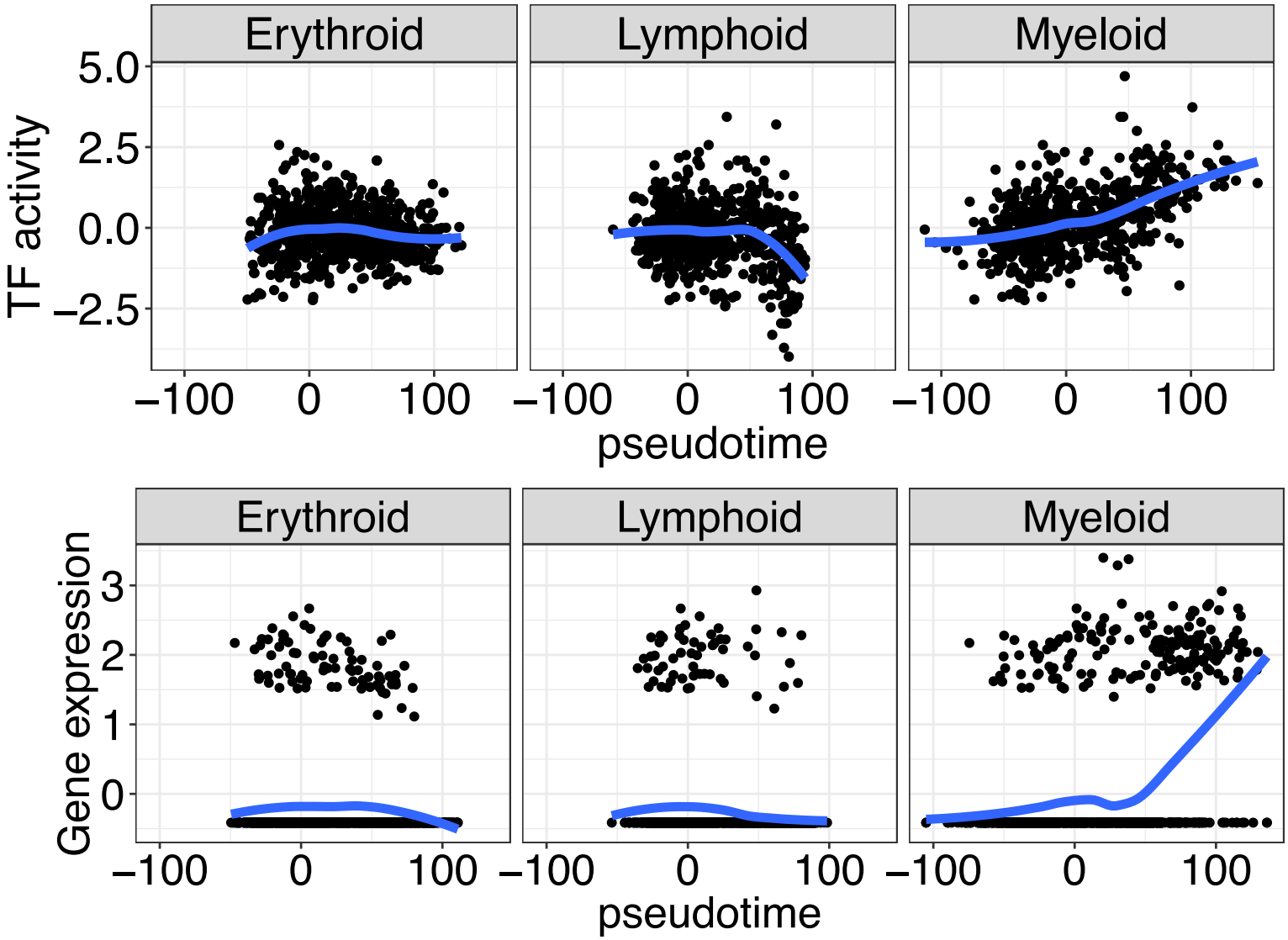
Application to HCA data

EBF1

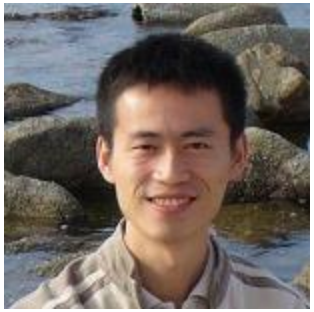


Application to HCA data

CEBPD

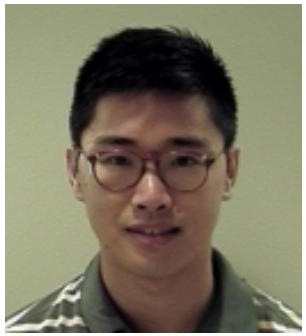


Acknowledgement



Hongkai Ji
Professor,
Department of Biostatistics,
Johns Hopkins Bloomberg
School of Public Health

Zhicheng Ji
PhD Candidate,
Department of Biostatistics,
Johns Hopkins Bloomberg
School of Public Health



Weixiang Fang
PhD Candidate,
Department of Biostatistics,
Johns Hopkins Bloomberg
School of Public Health