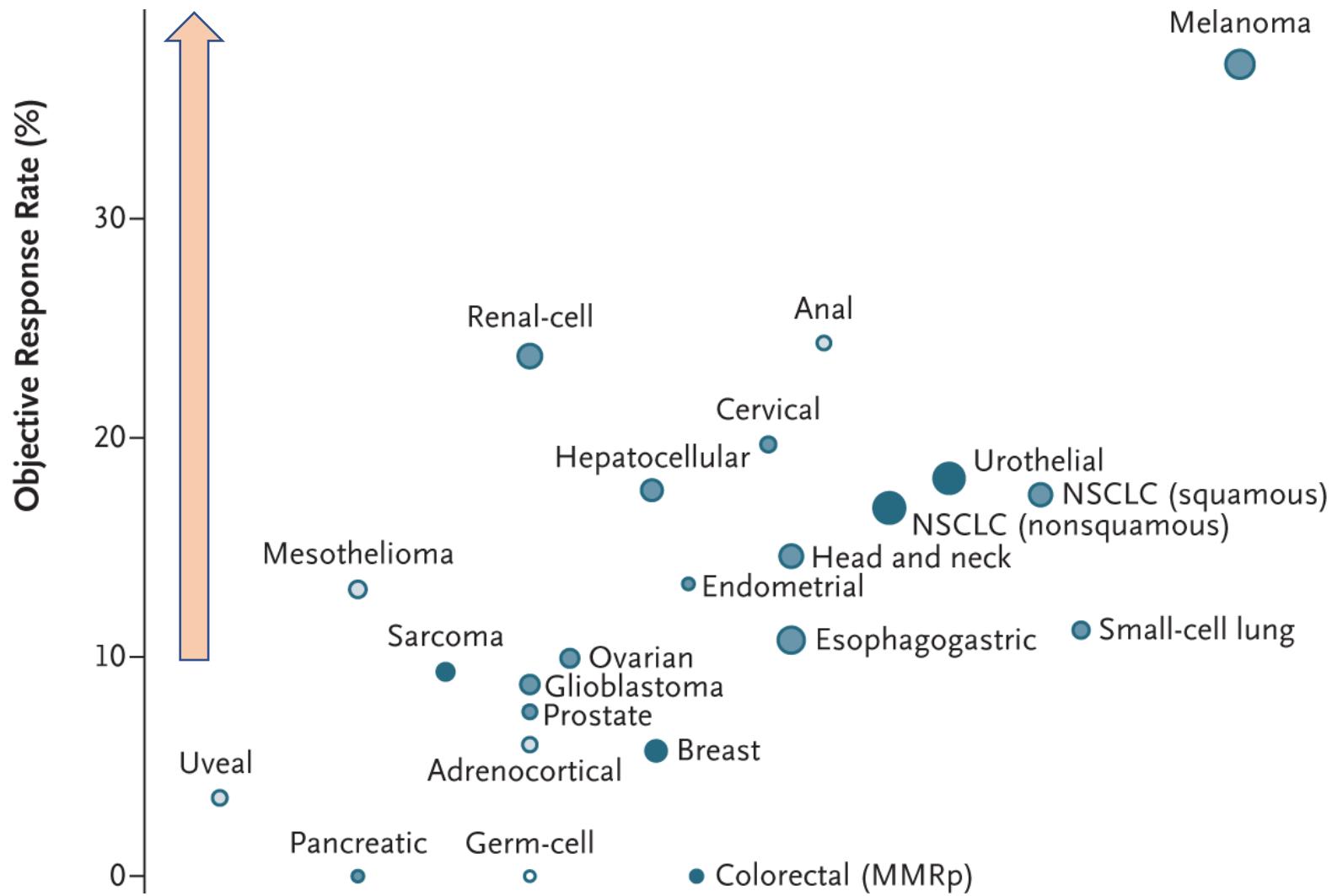


Identifying differentially expressed genes in single-cell data in response to immunotherapy

Avinash Das Sahu

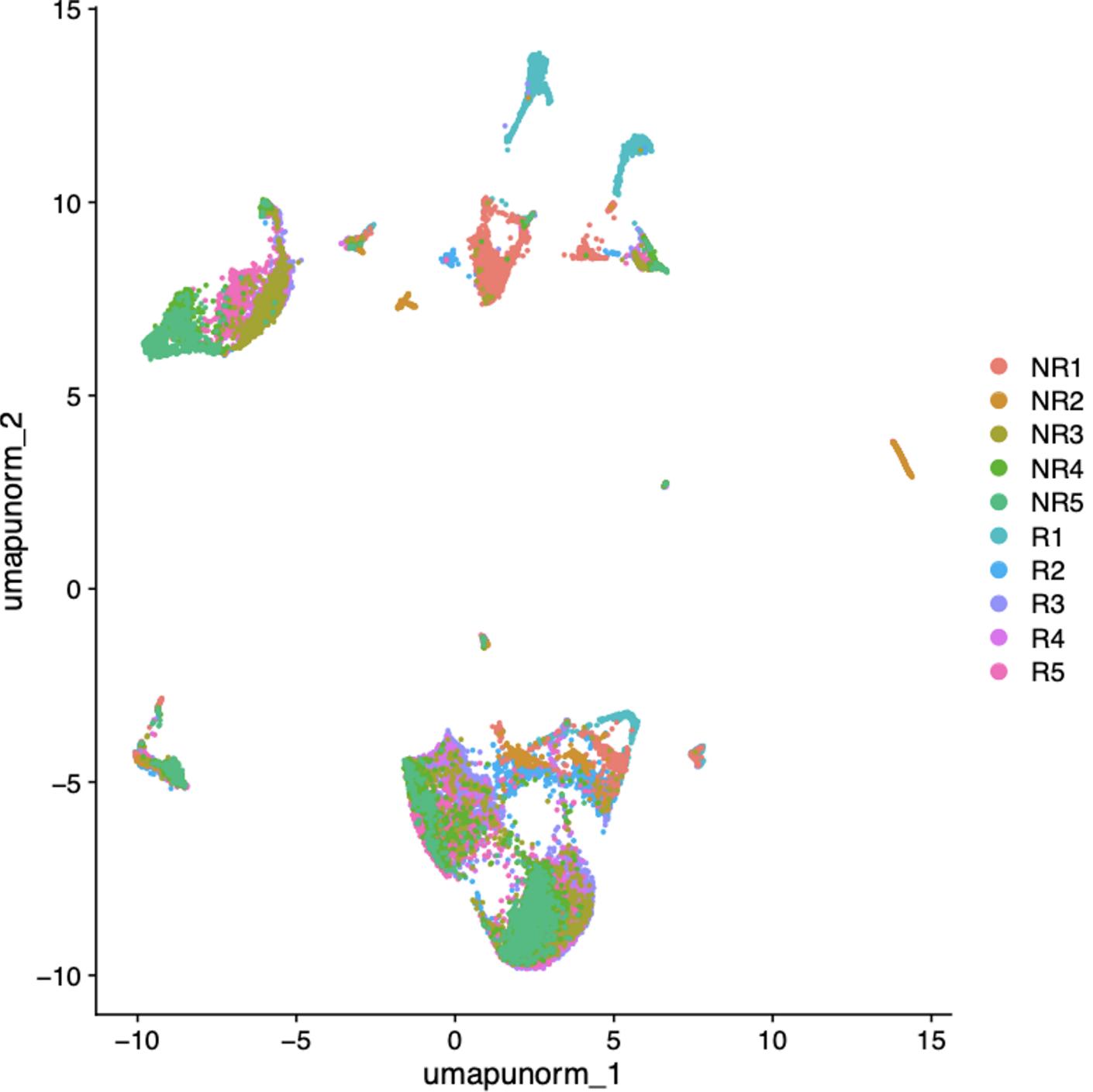
Immunotherapy response rate is low



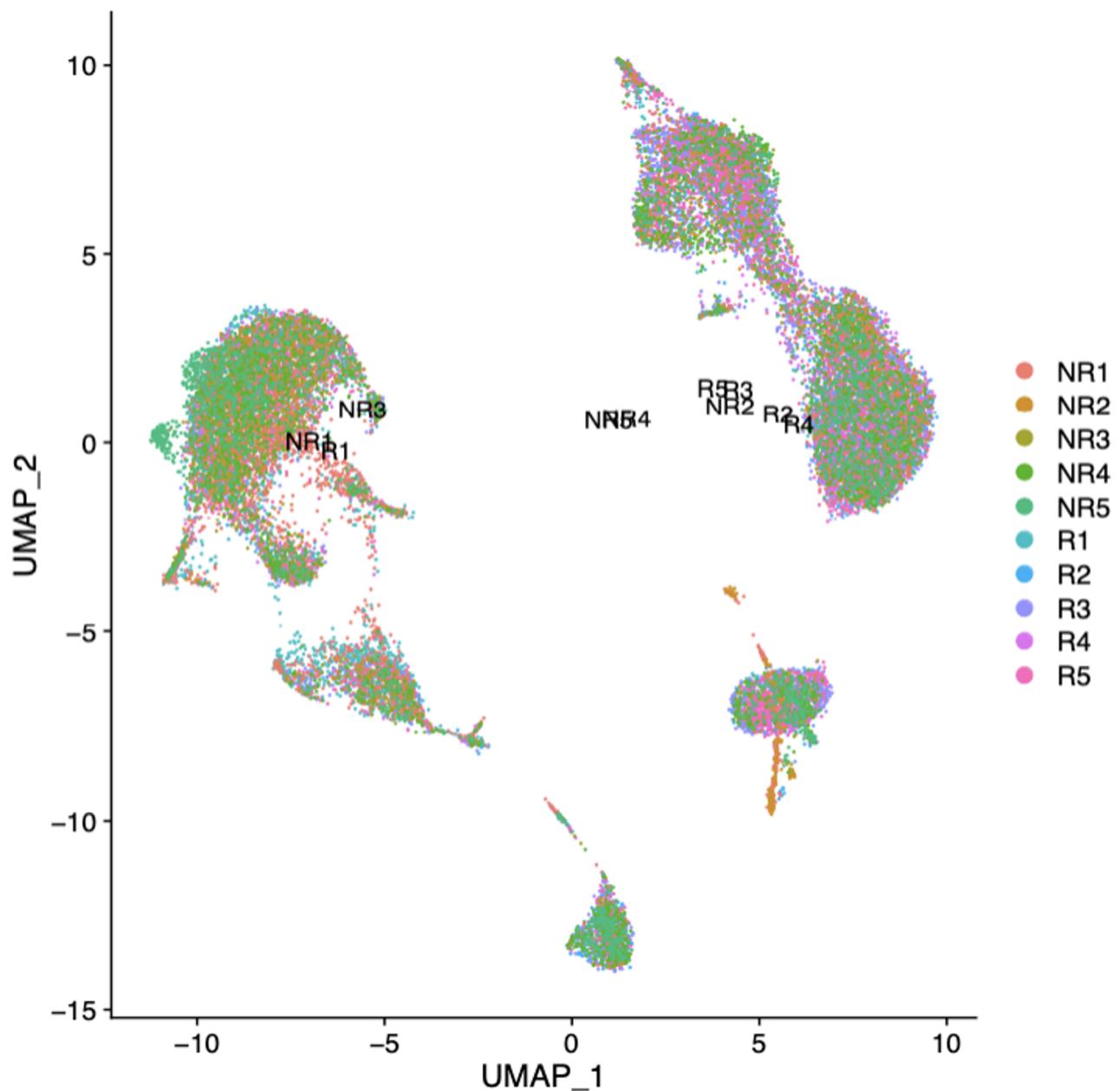
Single cell dataset description

- Peripheral blood mononuclear cells (PBMC) samples from 10 cancer patients with 5 responders and 5 non-responders in metastatic urothelial carcinoma
- 26K single cells

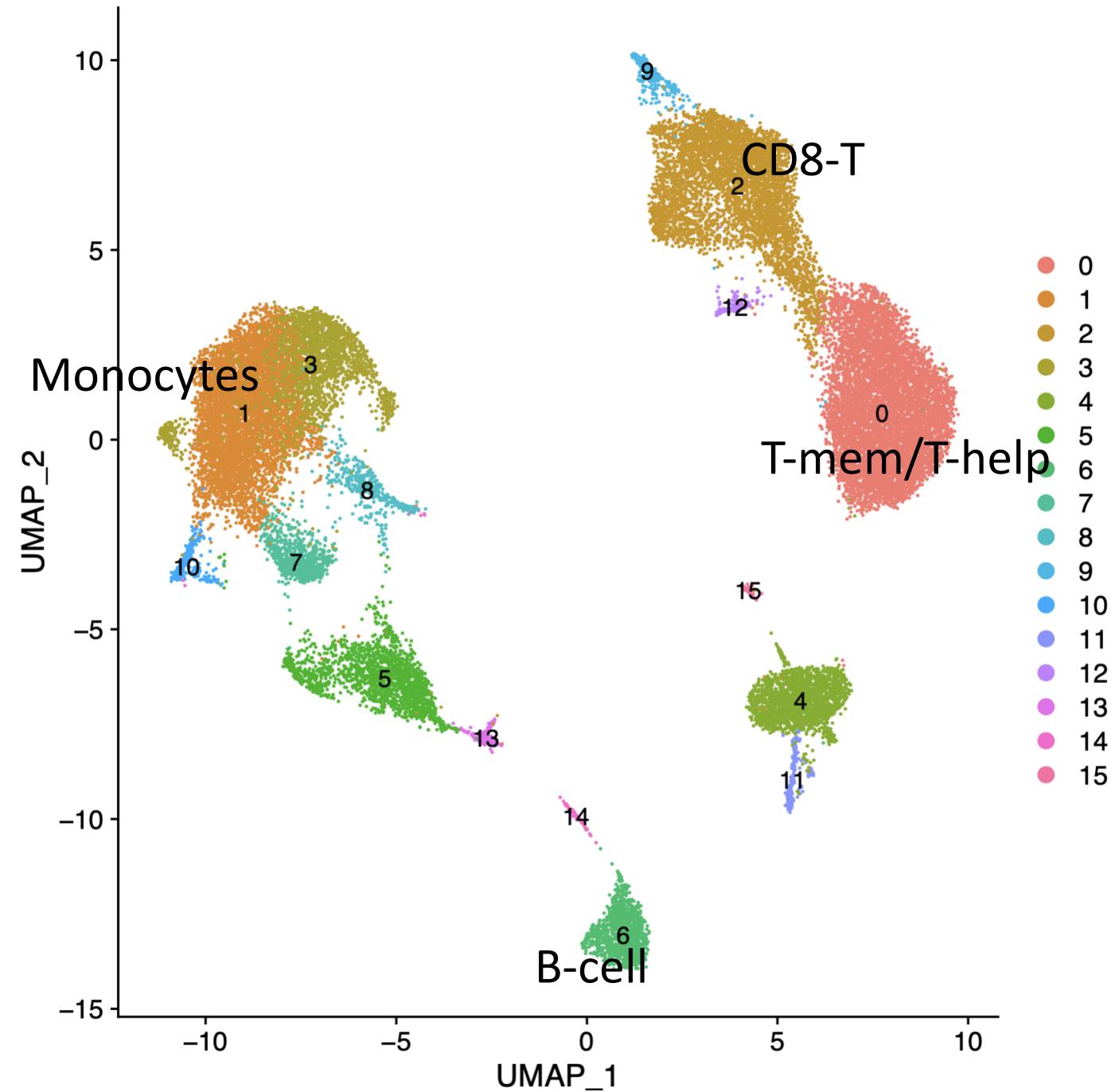
scRNA shows
patient-specific
batch effect



Batch
correction
methods can
remove the
batch effect



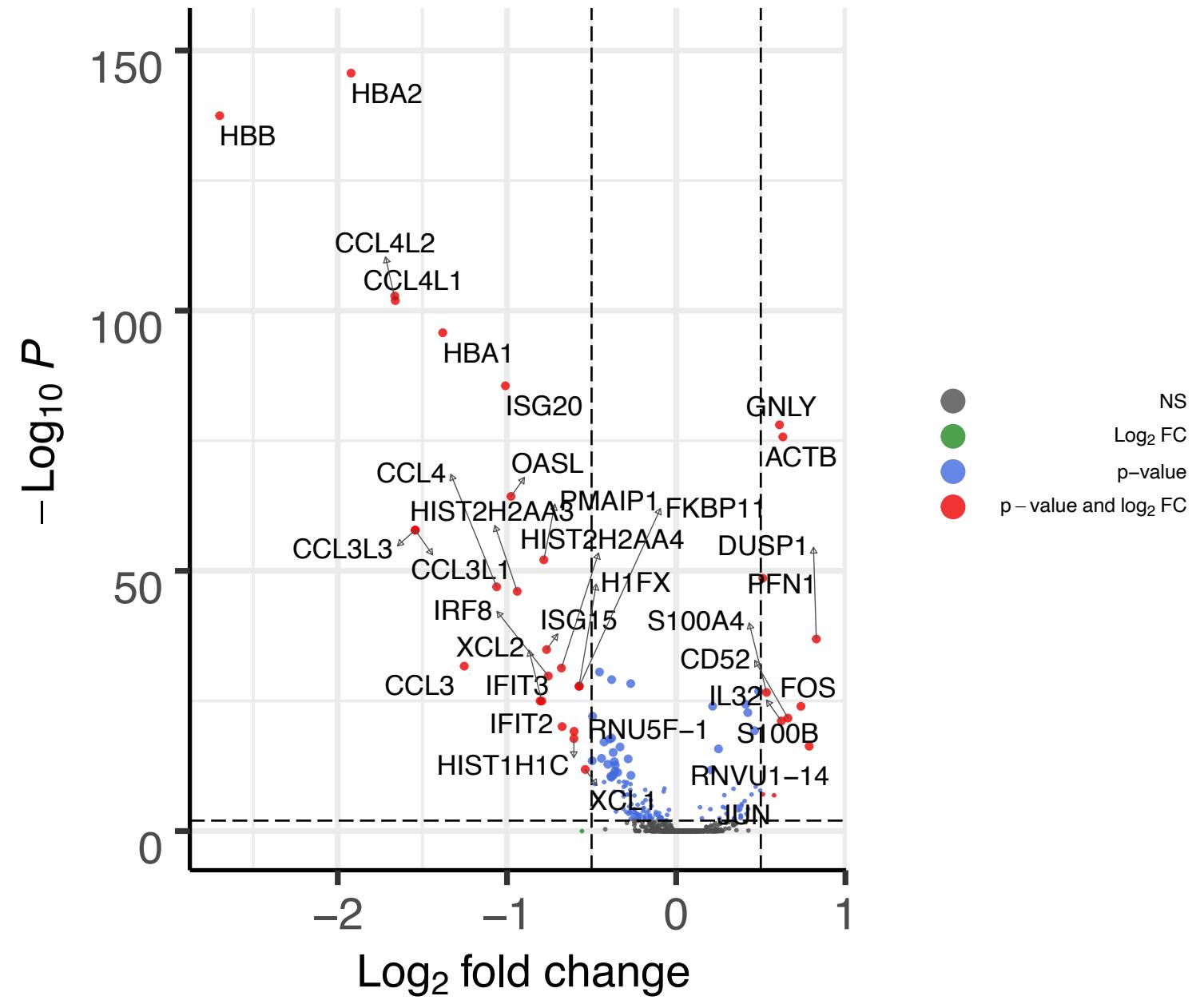
Manual annotation of cells



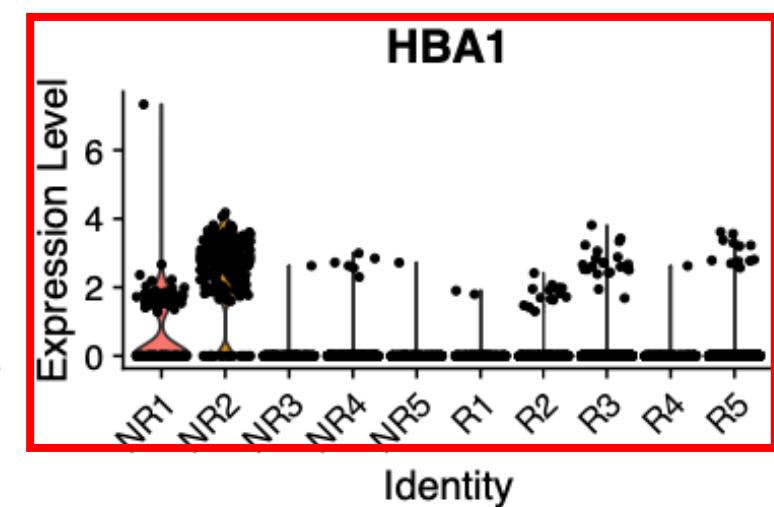
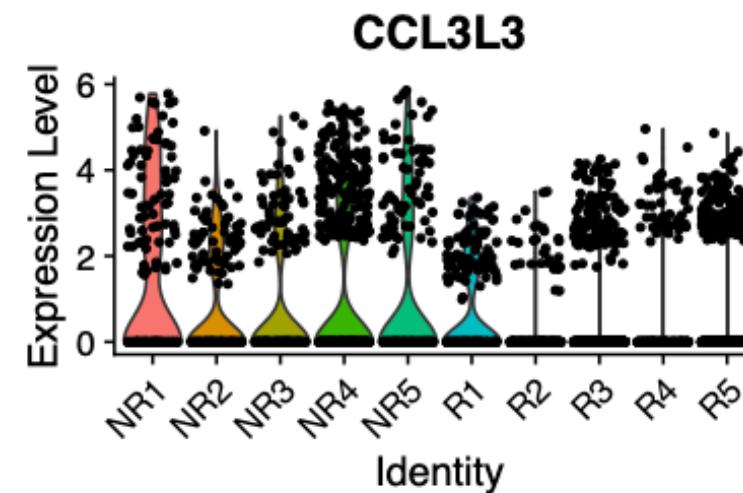
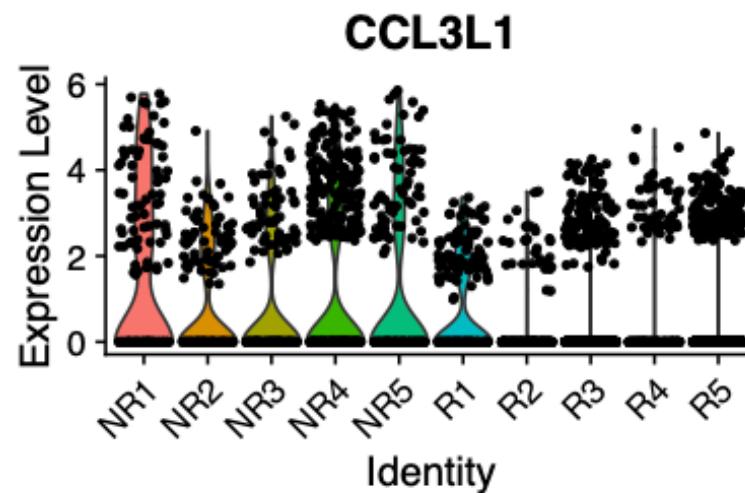
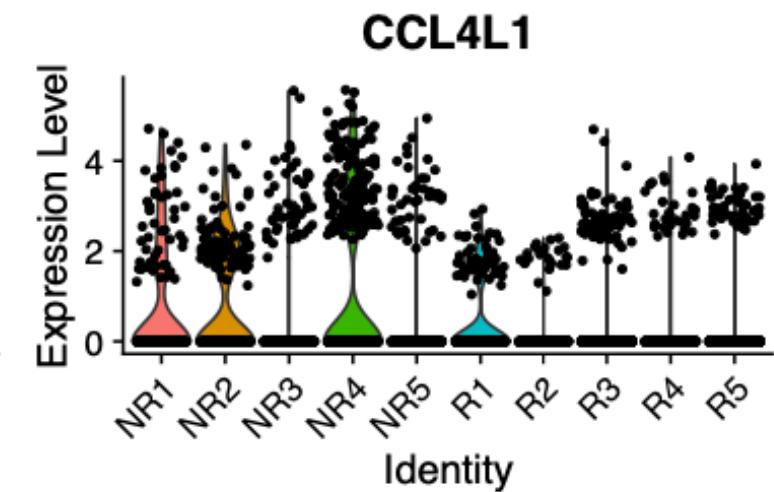
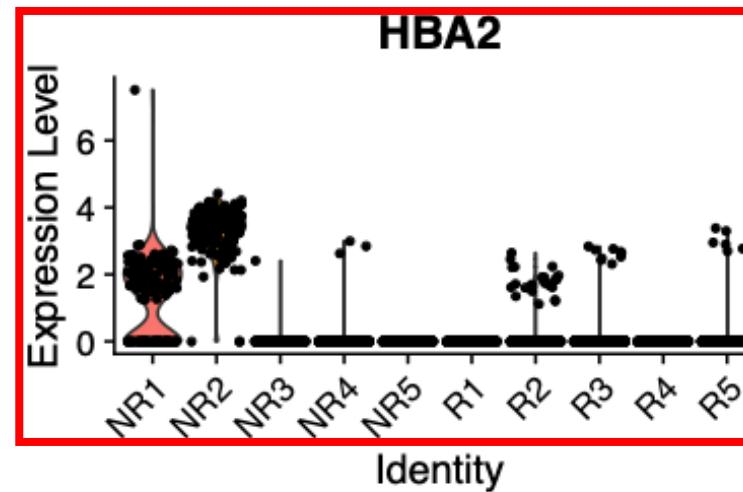
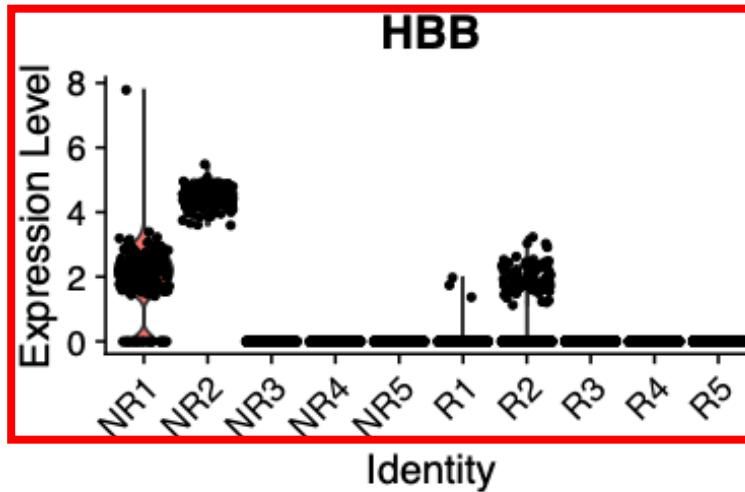
Within each cluster identify differential genes between responders and non-responders

- Wilcoxon test
- T-test
- Negative binomial GLM (Deseq)
- Poisson GLM (Deseq)
- Logistic regression
- Hurdle model (MAST)
- ROC

Several genes
are significantly
differentially
expressed in
responders vs.
non-responders



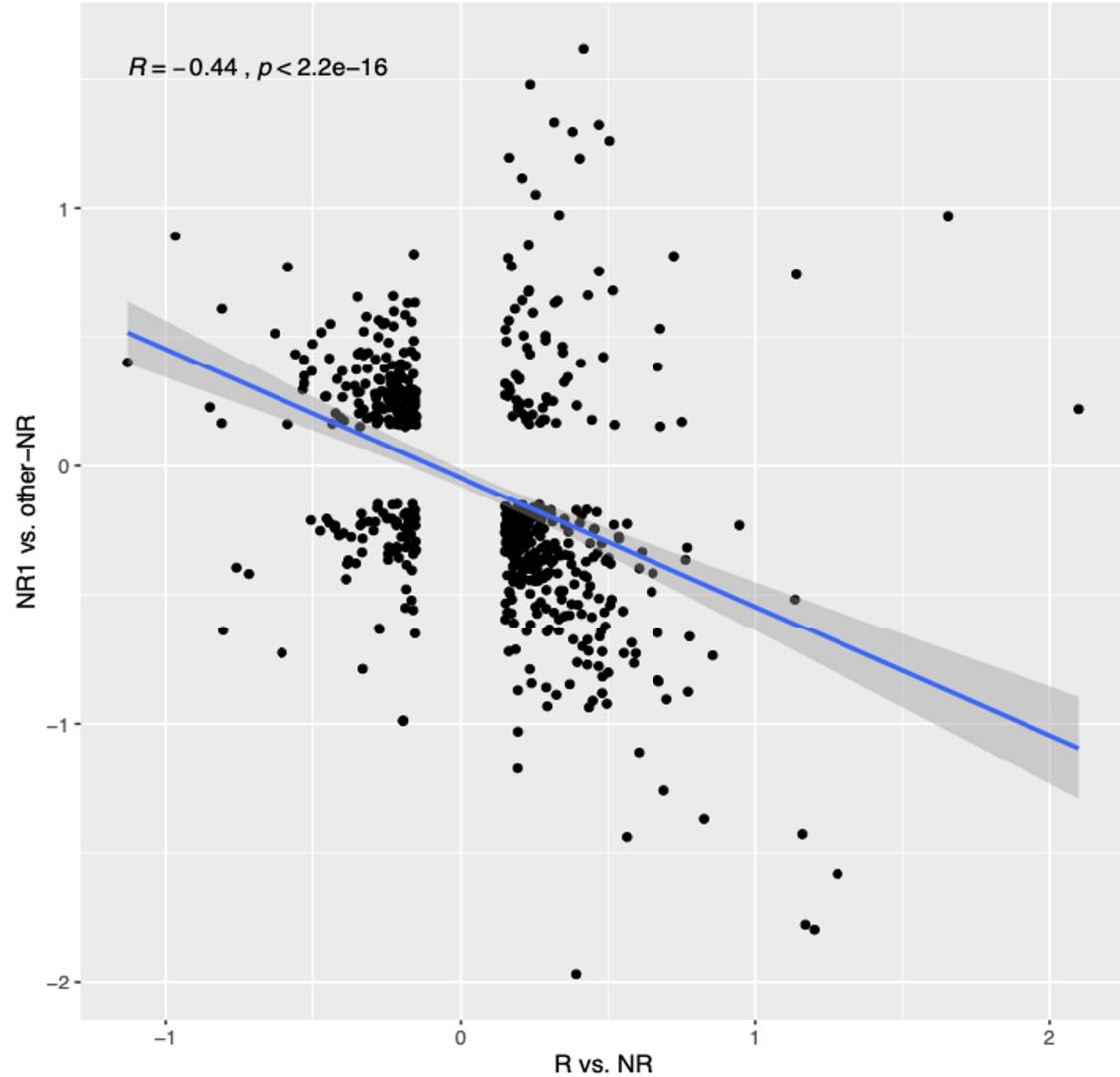
Patient-specific effects cause high false positives



Systematic evaluation of patient-specific effects

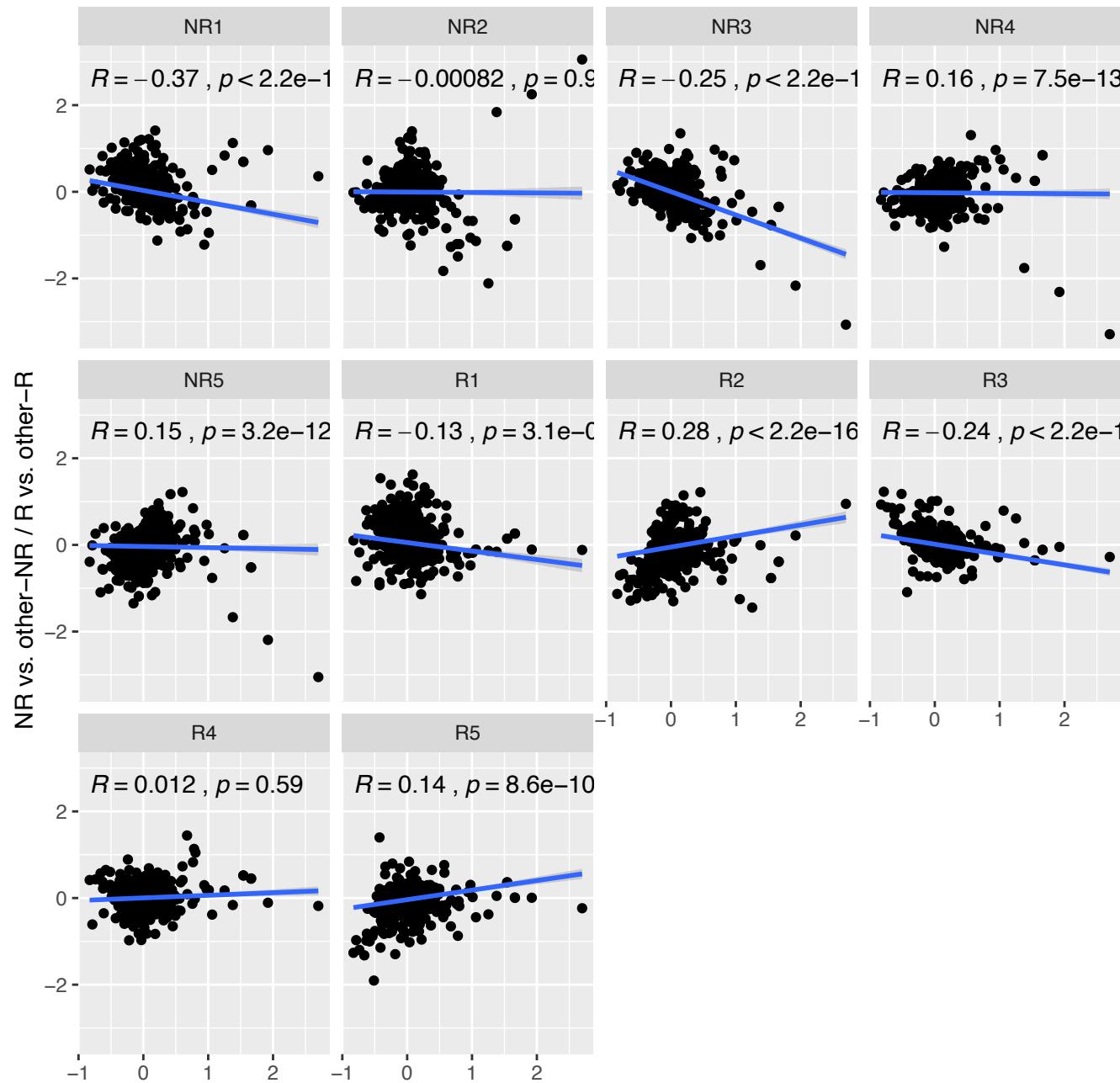
- Genes differentially expressed between responders vs. non-responders
 - must NOT be differentially expressed within responder
 - must NOT be differentially expressed within non-responder

Patient-specific effects confound differential expression analysis



All significant differentially expressed genes are displayed

Patient-specific effects confound differential expression analysis

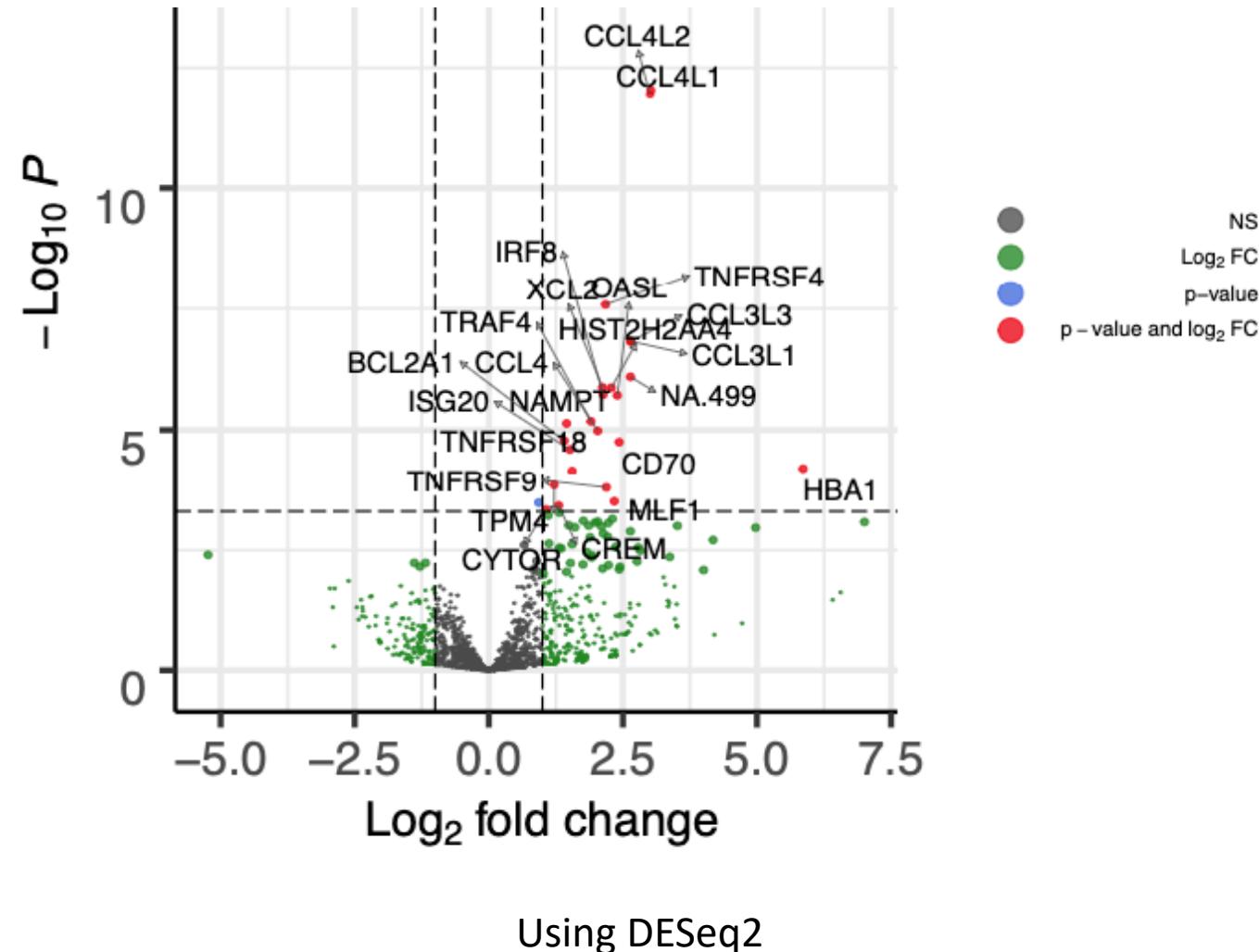


Design and power analysis for multi-sample single cell genomics experiments

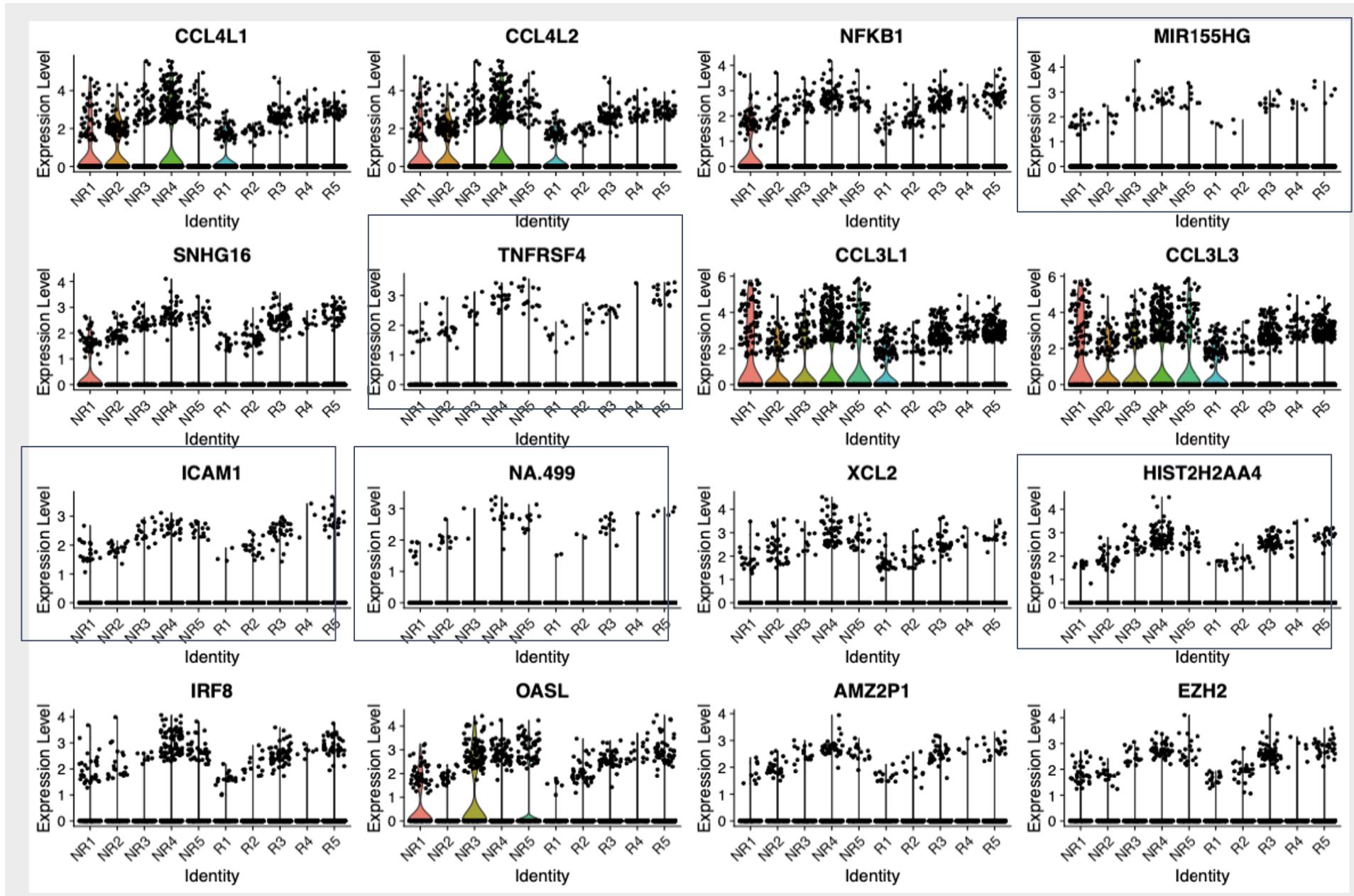
Katharina T. Schmid, Cristiana Cruceanu, Anika Böttcher, Heiko Lickert, Elisabeth B. Binder,
Fabian J. Theis, Matthias Heinig*

- Gene expression levels for each individual/sample is approximated as the sum of UMI counts over all cells of the cell type (pseudo-bulk)
- Differentially analysis of pseudo-bulk

Pseudo-bulk do
not suffer from
problem of
patient-specific
effects



Pseudo-bulk DEG is vulnerable to gene expressed in few cells



Any bulk RNA-seq might be also vulnerable to gene expressed in few cells!

≡ MENU



Search



UNDER CONSIDERATION

Pseudoreplication bias in single-cell studies; a practical solution

Kip D Zimmerman, Mark A Espeland, Carl D Langefeld



Nature Communications Jan 16, 2020

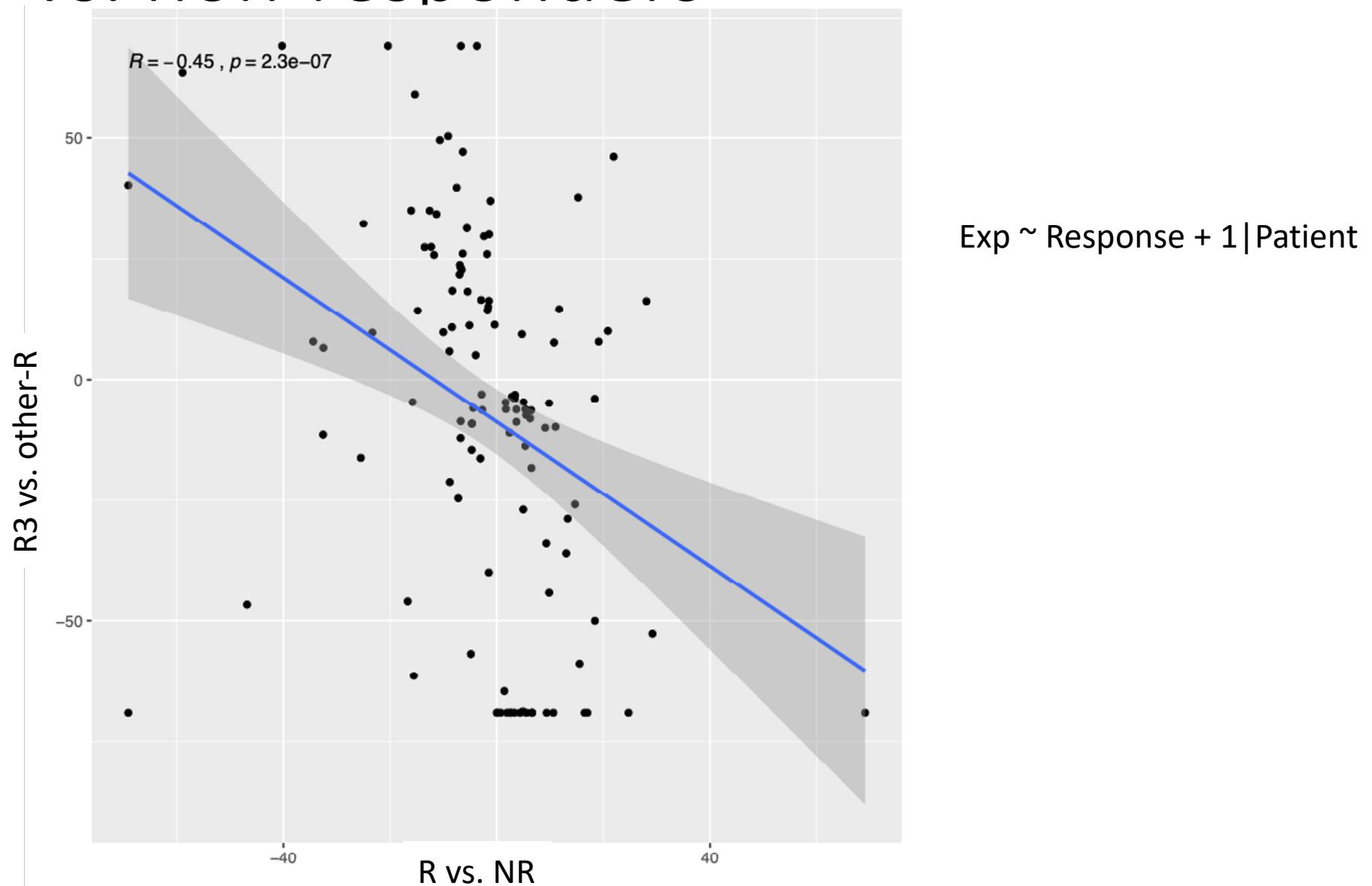


Pseudoreplication bias

- The use of inferential statistics where replicates are not statistically independent.
- Cells from the same individual share a common genetic and environmental background and are not independent, therefore they are pseudoreplicates.
- They propose to include random effects for differences among persons.

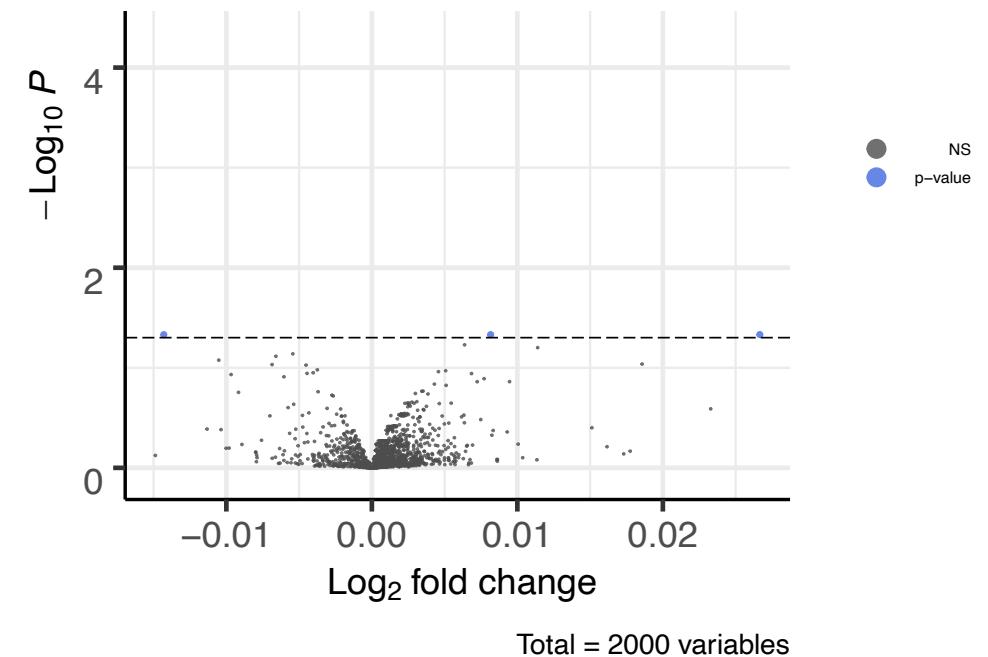
$\text{Exp} \sim \text{Response} + 1 \mid \text{Patient}$

But the method will not work when comparing responders vs. non-responders



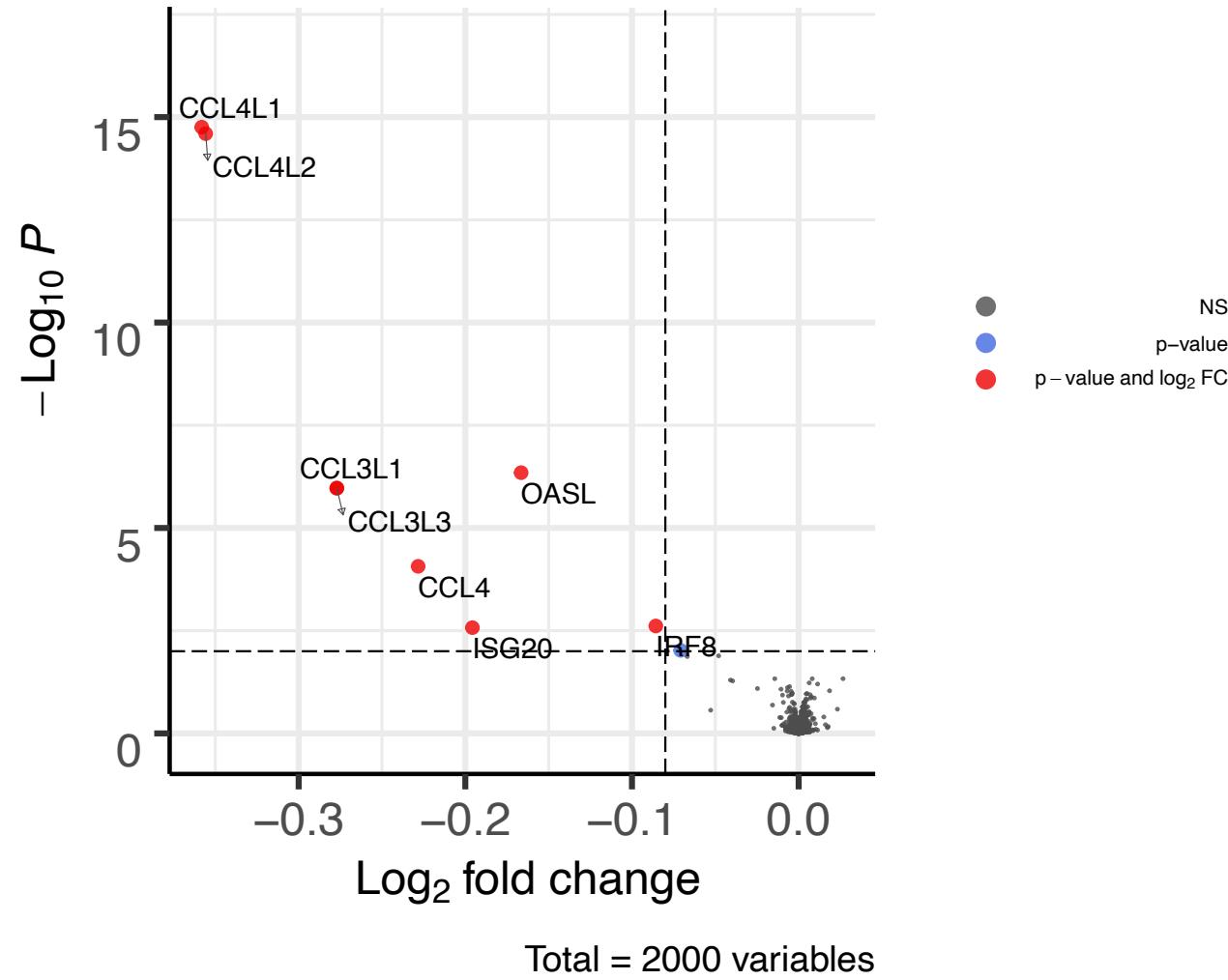
Controlling for patient-specific effect explicitly

- Step 1:
 - Model 1 : $\text{Exp} \sim 1 | \text{Patient}$
 - Calculate $\widehat{\text{Exp}} = \text{Predict}(\text{Model 1})$
- Step 2 : $(\text{Exp} - \widehat{\text{Exp}}) \sim \text{Response}$

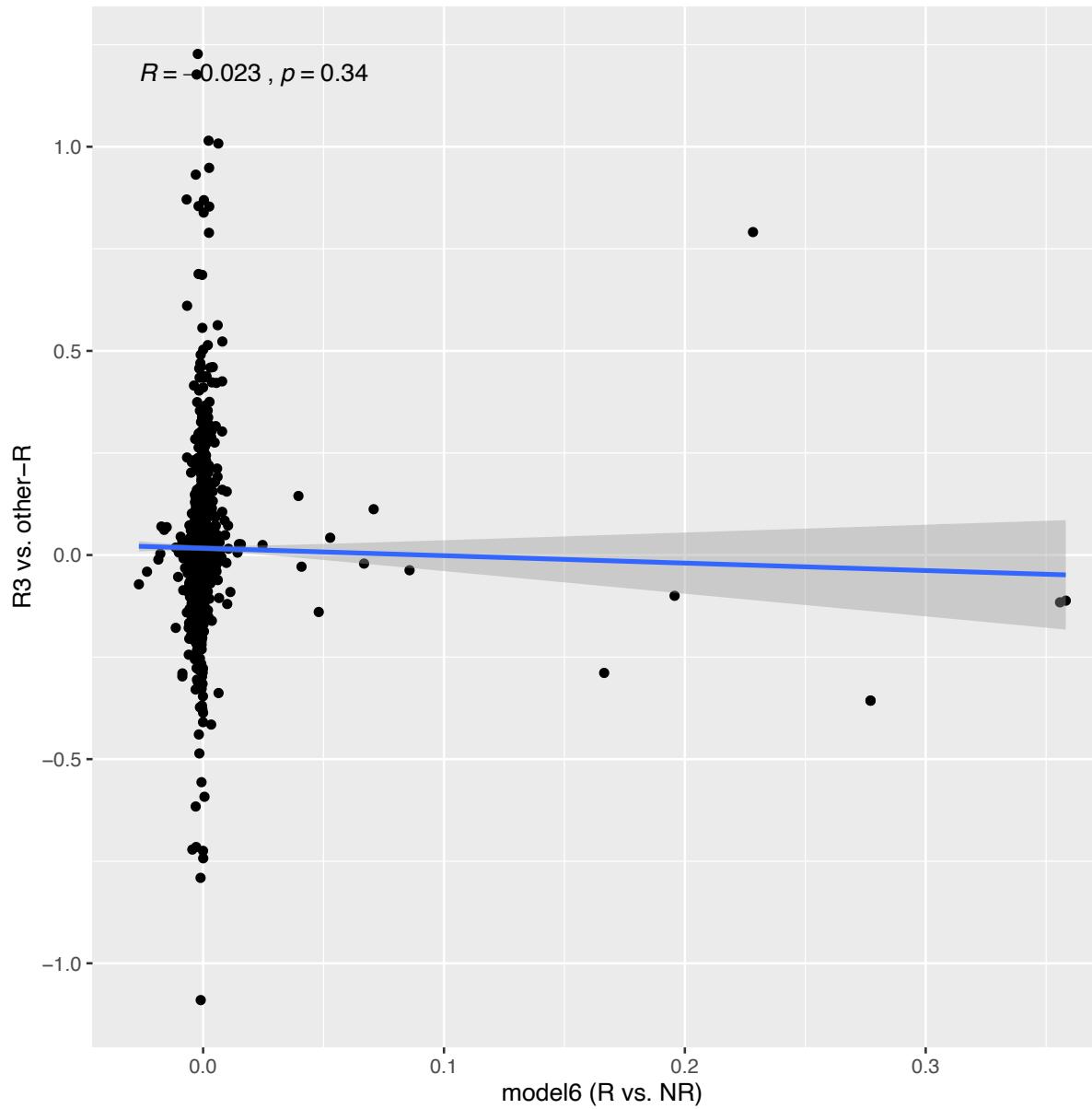


Using estimates from pseudo-bulk analysis as priors (heuristic)

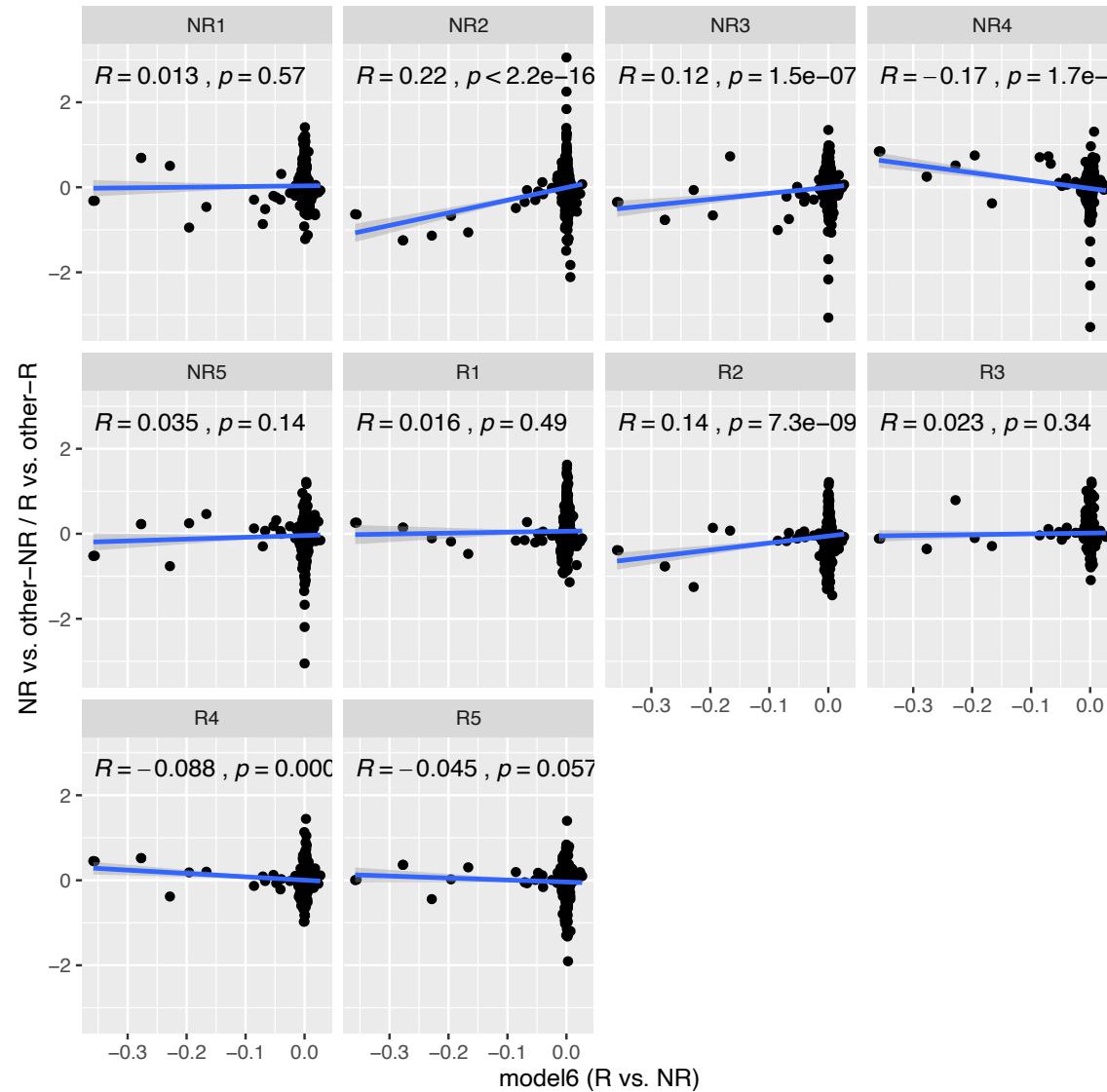
- Pseudo-bulk model:
 - $\text{Exp} \sim \text{Response}$
 - $\widehat{\text{Exp}} = \text{Predict}(\text{Pseudo-bulk model})$
- Step 2:
 - Model 1 : $(\text{Exp} - \widehat{\text{Exp}}) \sim 1 | \text{Patient}$
 - Calculate $\widehat{\text{Exp}} = \text{Predict}(\text{Model 1})$
- Step 3 : $(\text{Exp} - \widehat{\text{Exp}}) \sim \text{Response}$



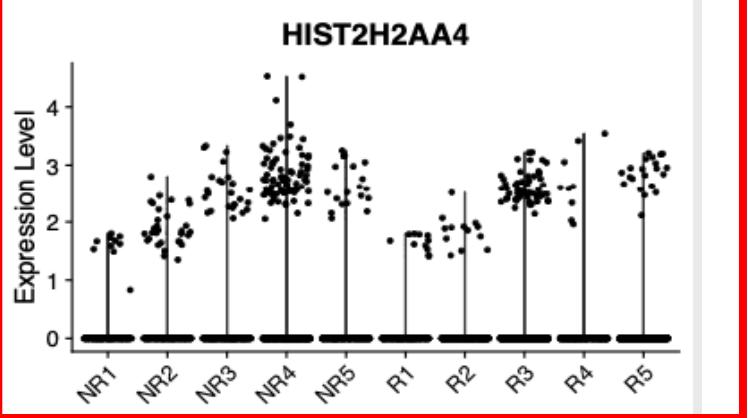
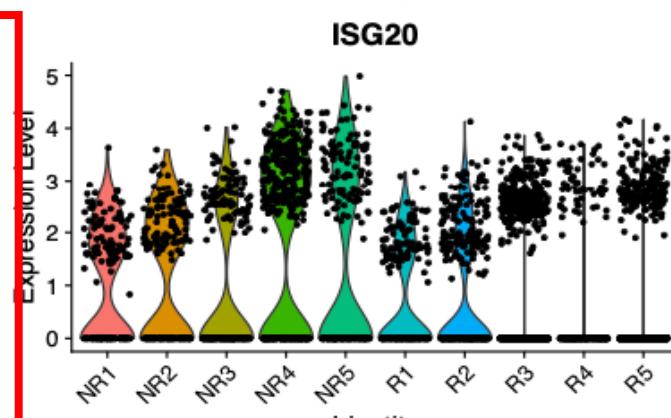
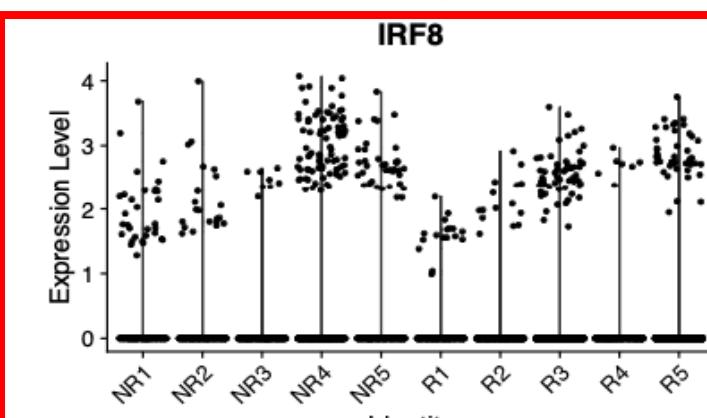
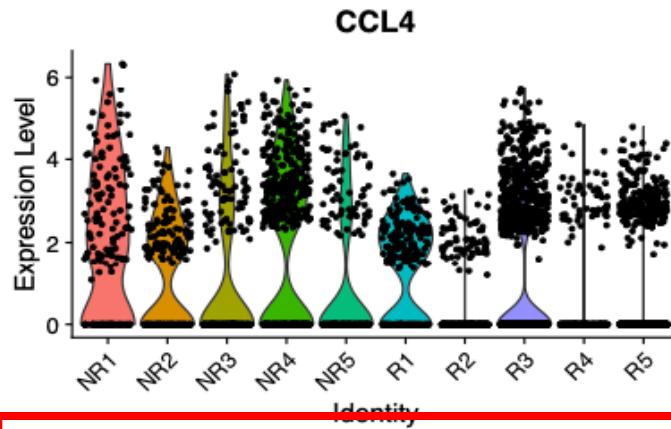
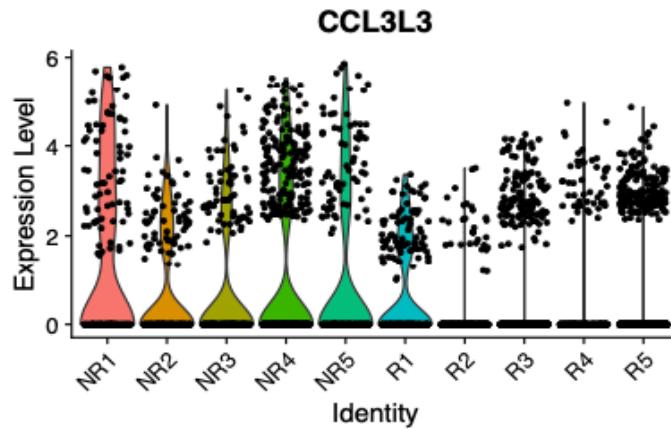
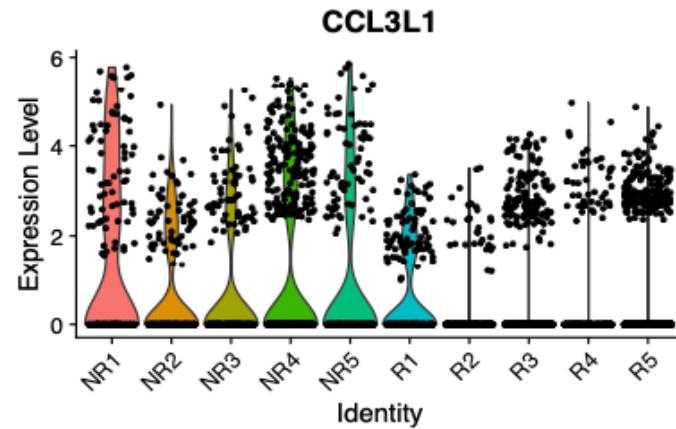
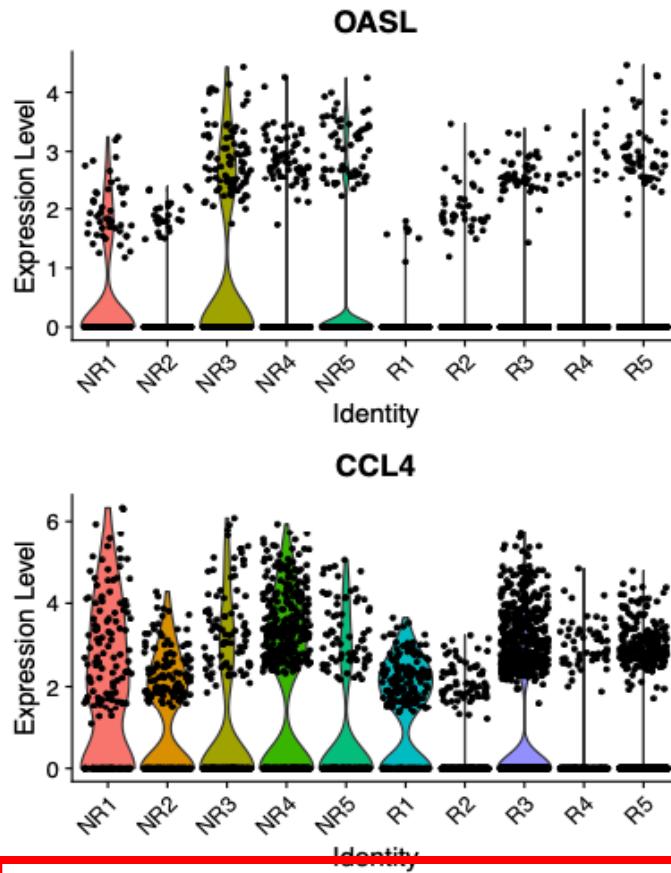
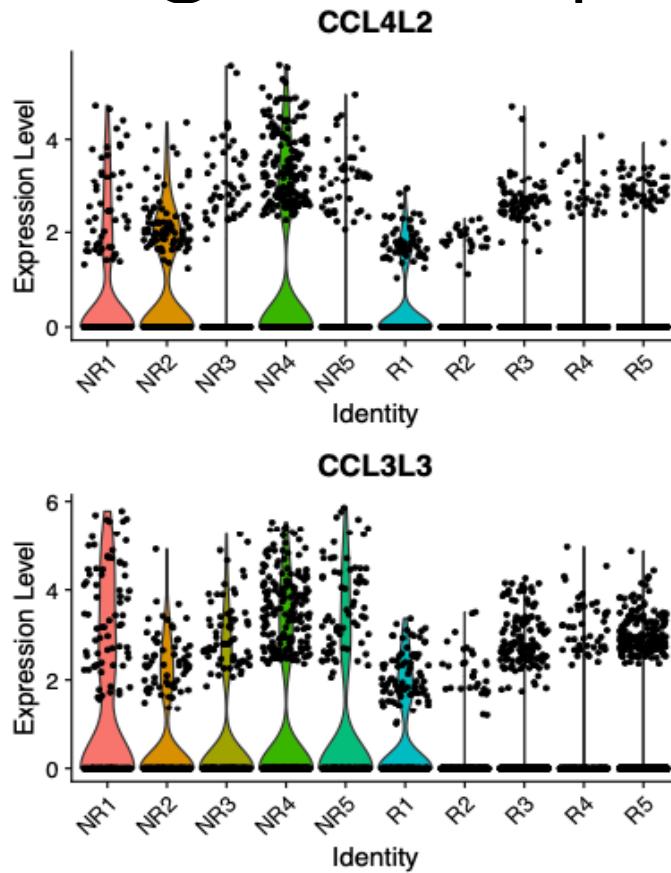
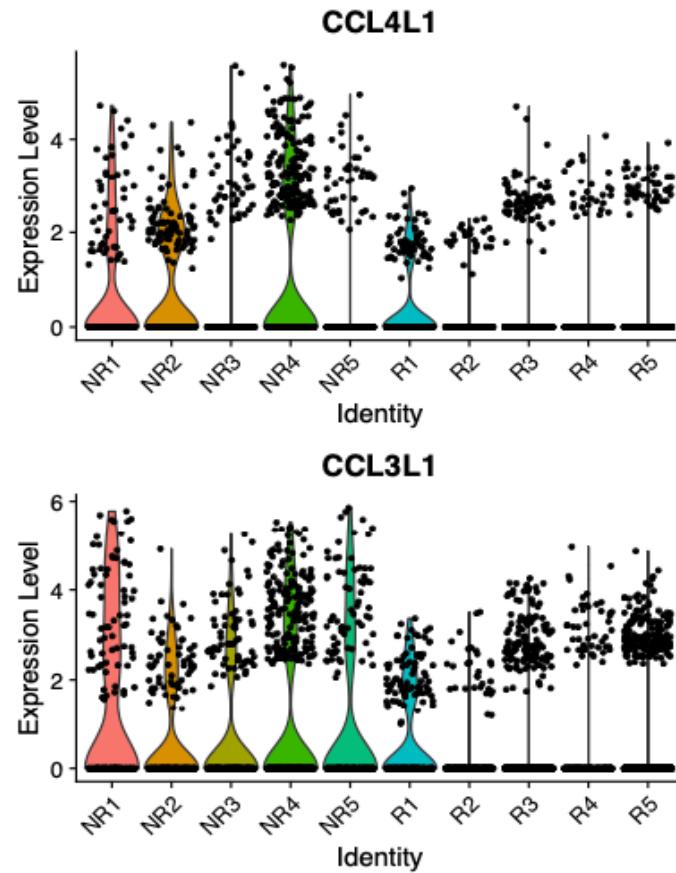
Mitigates patient-specific effects



Mitigates patient-specific effects



But still vulnerable to genes expressed in few cells



Summary

1. Single cell differential expression methods
 - a. take into account cell specific differences
 - b. but are poor in handling patient-specific batch effects
2. Pseudo-bulk method
 - a. is better in handling patient-specific batch effects
 - b. but loses advantage of single cell data
 - c. and are vulnerable genes expressed in cells.
3. Can we combine those two methods?
4. Can we use something from batch correction methods for single cell clustering? CCA, MNN etc.