Slalom: scalable and versatile factor analysis for single-cell RNA-seq

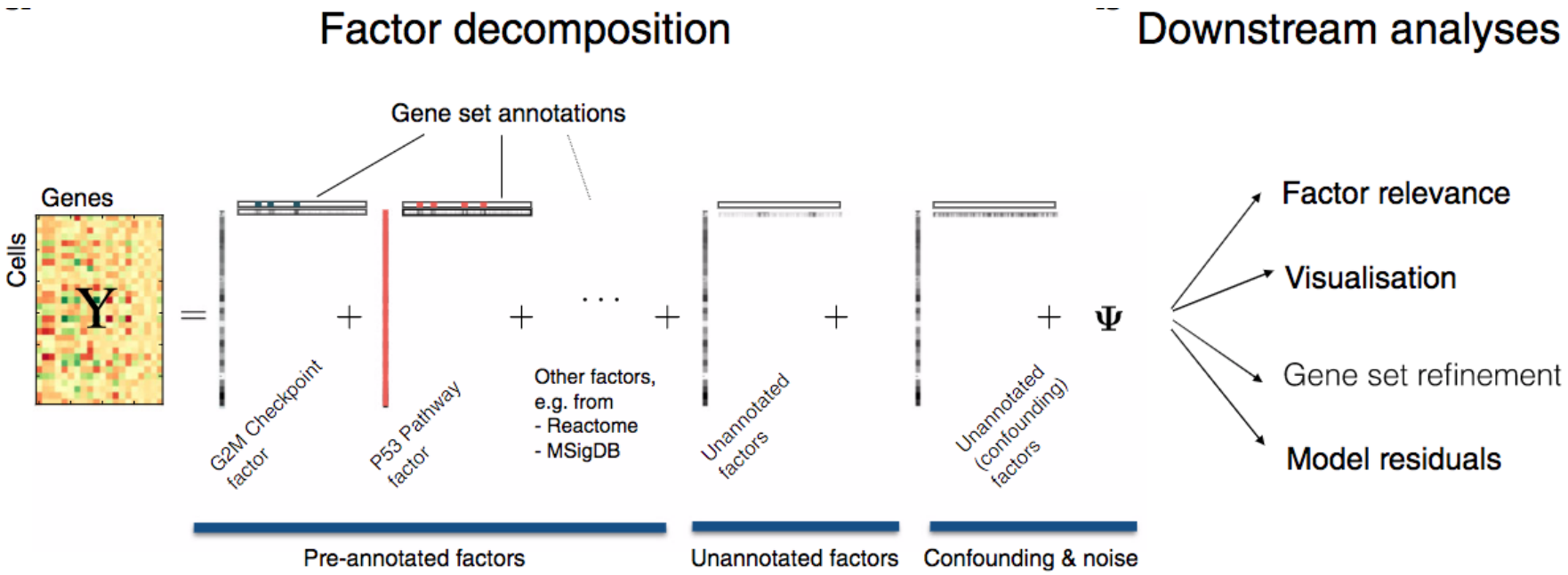
Davis McCarthy
NHMRC Early Career Fellow

Stegle Group, EMBL-EBI

www.ebi.ac.uk

EMBL-EBI

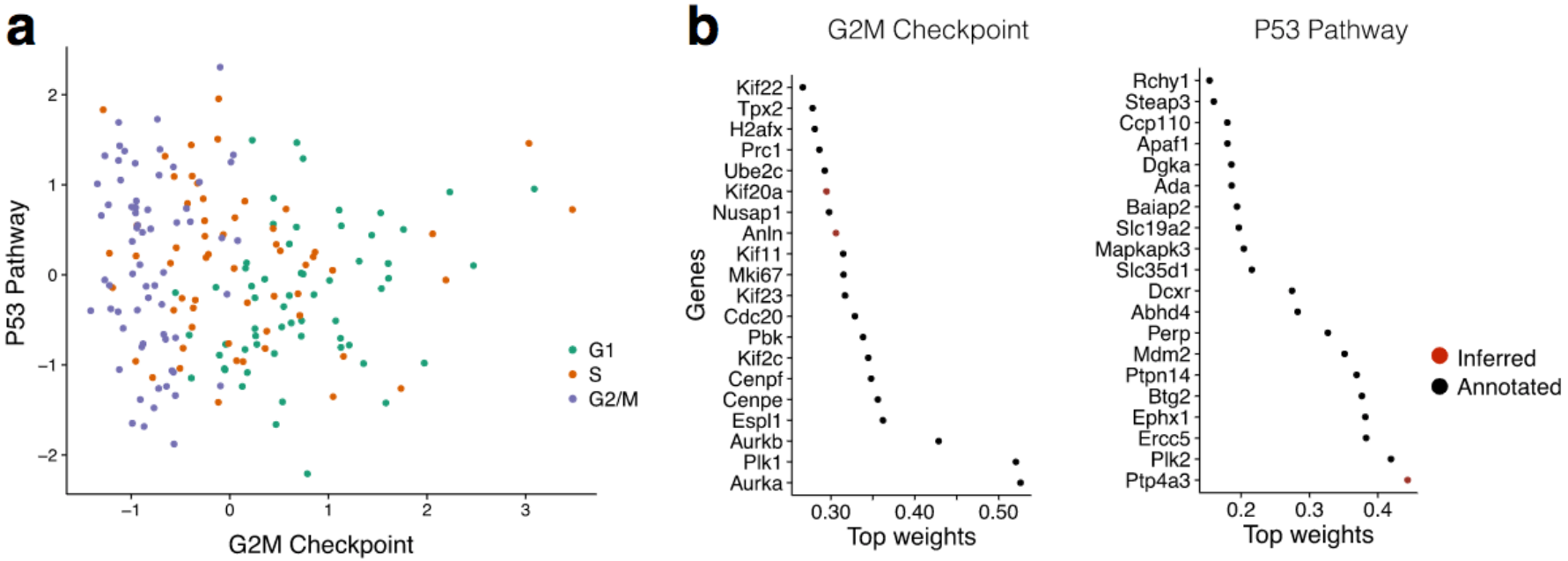
Factorial single-cell latent variable model



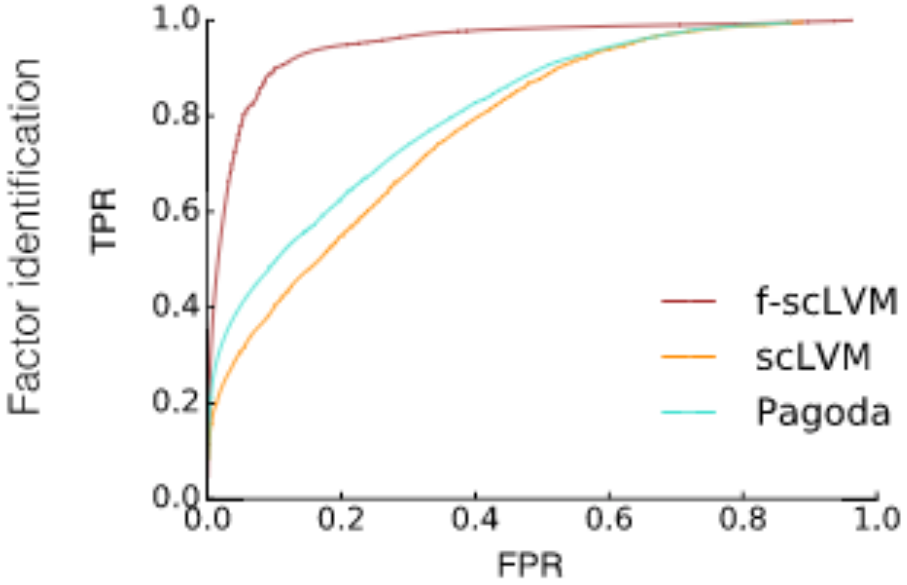
Buettner et al. *Genome Biology*, 2017

EMBL-EBI

Slalom finds interpretable factors explaining variation

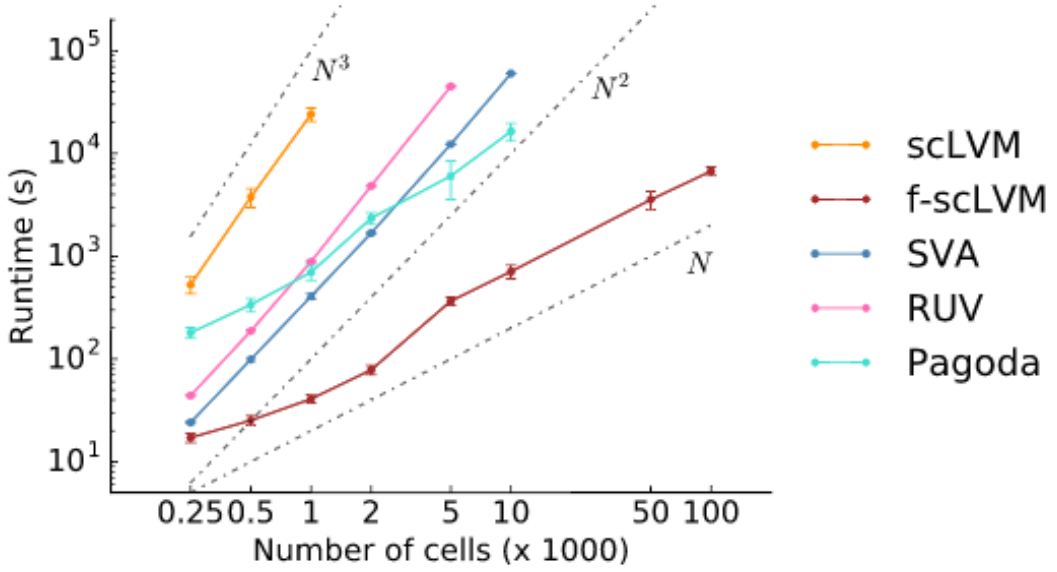


Mouse ESC data



Slalom outperforms scLVM and Pagoda

Slalom scales linearly in number of cells, so is feasible to run up to ~100,000 cells



Easy to create, initialise and train a slalom model

```
model <- newSlalomModel(mesc, genesets, n_hidden = 5, min_genes = 10)
```

```
## 14 annotated factors retained;  16 annotated factors dropped.
## 196  genes retained for analysis.
```

Next we need to *initialise* the model with the `init` function.

```
model <- initSlalom(model)
```

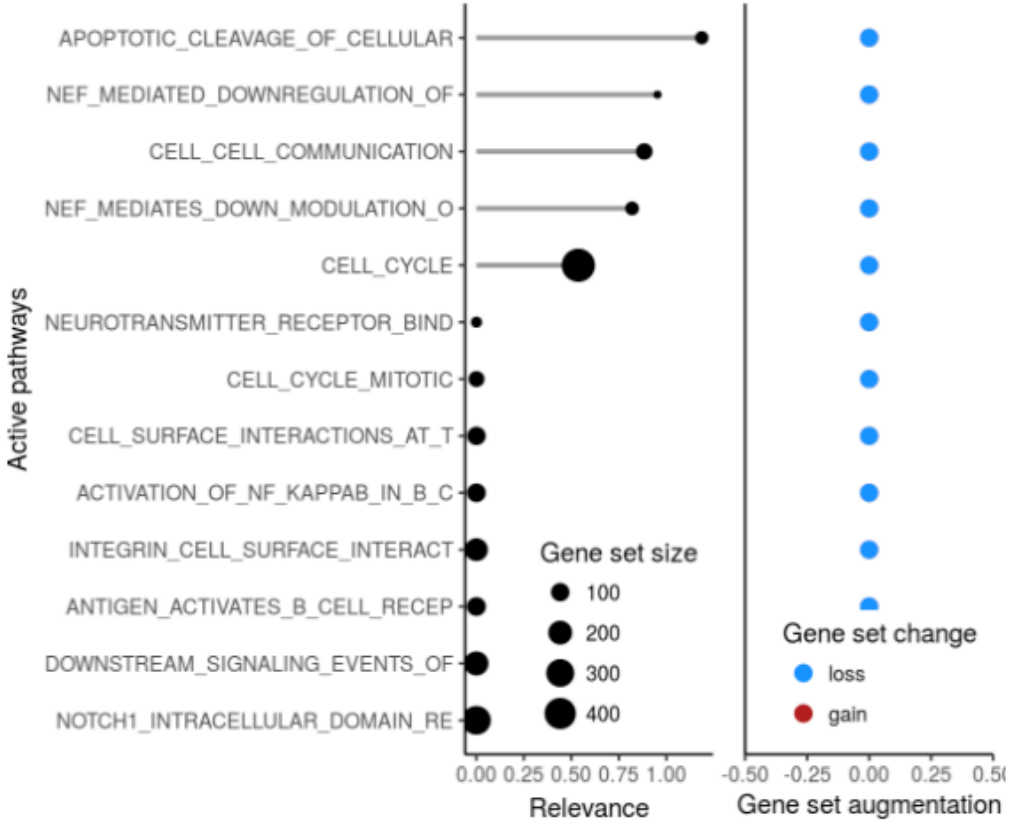
With the model prepared, we then *train* the model with the `train` function.

```
model <- trainSlalom(model, nIterations = 10)
```

```
## pre-training model for faster convergence
## iteration 0
## Model not converged after 50 iterations.
## iteration 0
## Model not converged after 50 iterations.
## iteration 0
## Switched off factor 17
## Switched off factor 18
## Model not converged after 10 iterations.
```

Interrogate results in trained model

```
plotRelevance(m)
```



`topTerms`

`plotRelevance`

`plotTerms`

`plotLoadings`

`addResultstoSingleCellExperiment`

Acknowledgements

- Florian Buettner
- Naruemon (Ploy) Pratanwanich
- John Marioni
- Oliver Stegle