

# Correcting unwanted variation in RNA sequencing data derived from a multi-centre study of leukemia

Anna Quagliari

3rd PhD Student

Walter and Eliza Hall Institute of Medical Research



**ABACBS** Conference 2018

 @annaquagli





A detailed illustration of a blood smear. The background is a deep red, wavy texture representing plasma. Numerous red blood cells (erythrocytes) are visible, appearing as biconcave discs. Several white blood cells (leukocytes) are also present, characterized by their larger size and prominent, multi-lobed nuclei. One large leukocyte is in the foreground, showing a high magnification view of its nucleus and granules. Other smaller leukocytes are scattered throughout the field.

Leukemia:  
Cancer of the blood

**CBF-AML**

# CBF-AML

Diagnosis

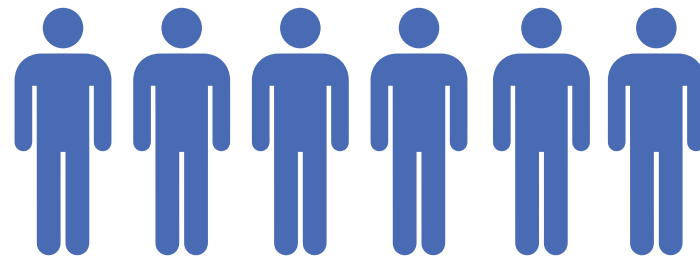


Chemotherapy



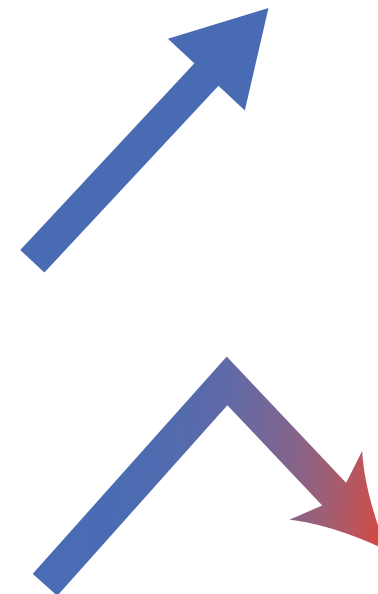
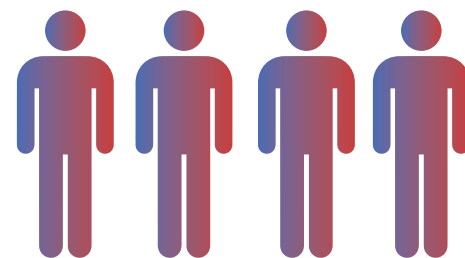
**60%**

Long term remission

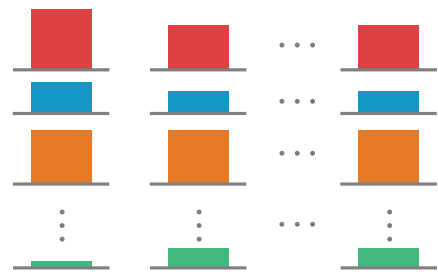


**40%**

Relapse



Are there **genes**, observed at diagnosis,  
associated with different **outcomes**?



# Gene expression from RNA-Seq

**56**

Long term  
remission

**20**

Relapse



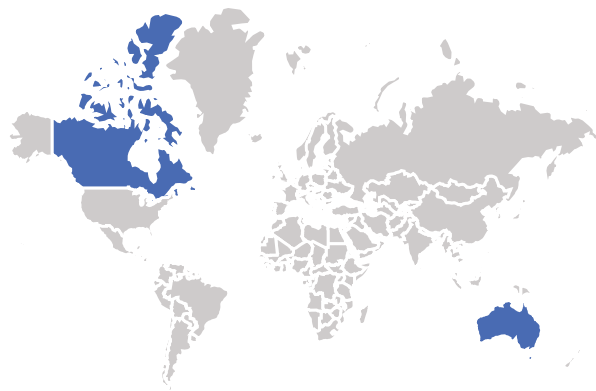
# Gene expression from RNA-Seq

**56**

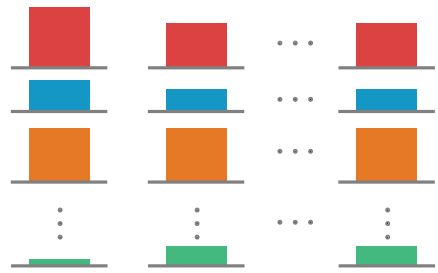
Long term  
remission

**20**

Relapse



From **Australia** and **Canada**  
Sequenced in **3 batches** across 2.5 years



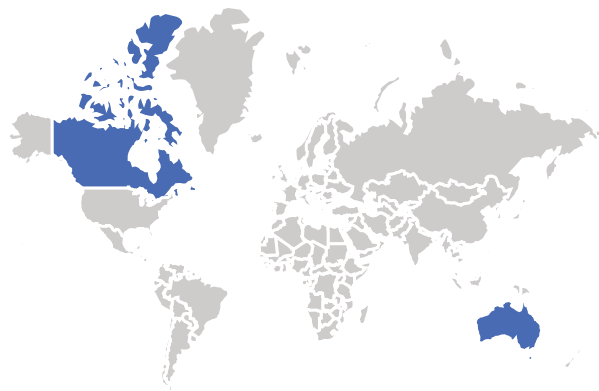
# Gene expression from RNA-Seq

**56**

Long term  
remission

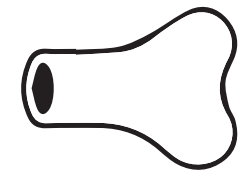
**20**

Relapse



From **Australia** and **Canada**  
Sequenced in **3 batches** across 2.5 years

From  
**Bone Marrow** or **Blood**







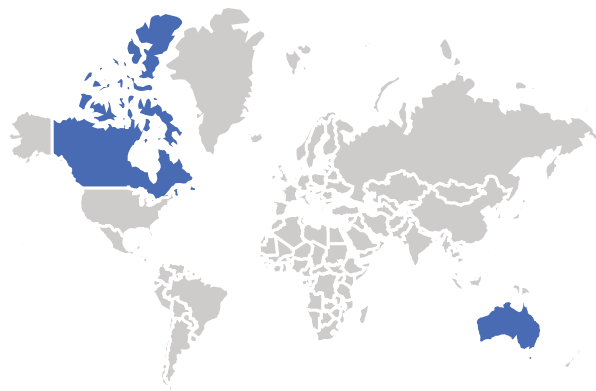
# Gene expression from RNA-Seq

56

Long term  
remission

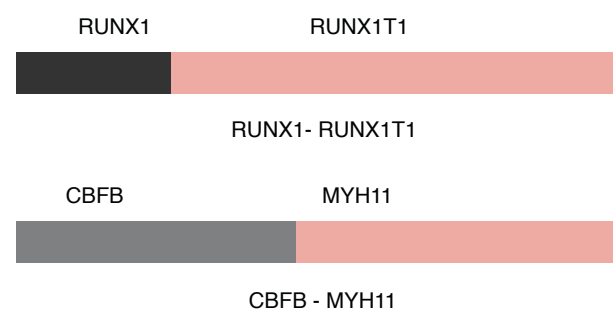
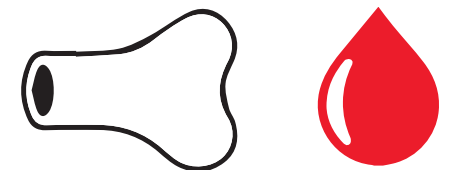
20

Relapse



From **Australia** and **Canada**  
Sequenced in **3 batches** across 2.5 years

From  
**Bone Marrow** or **Blood**



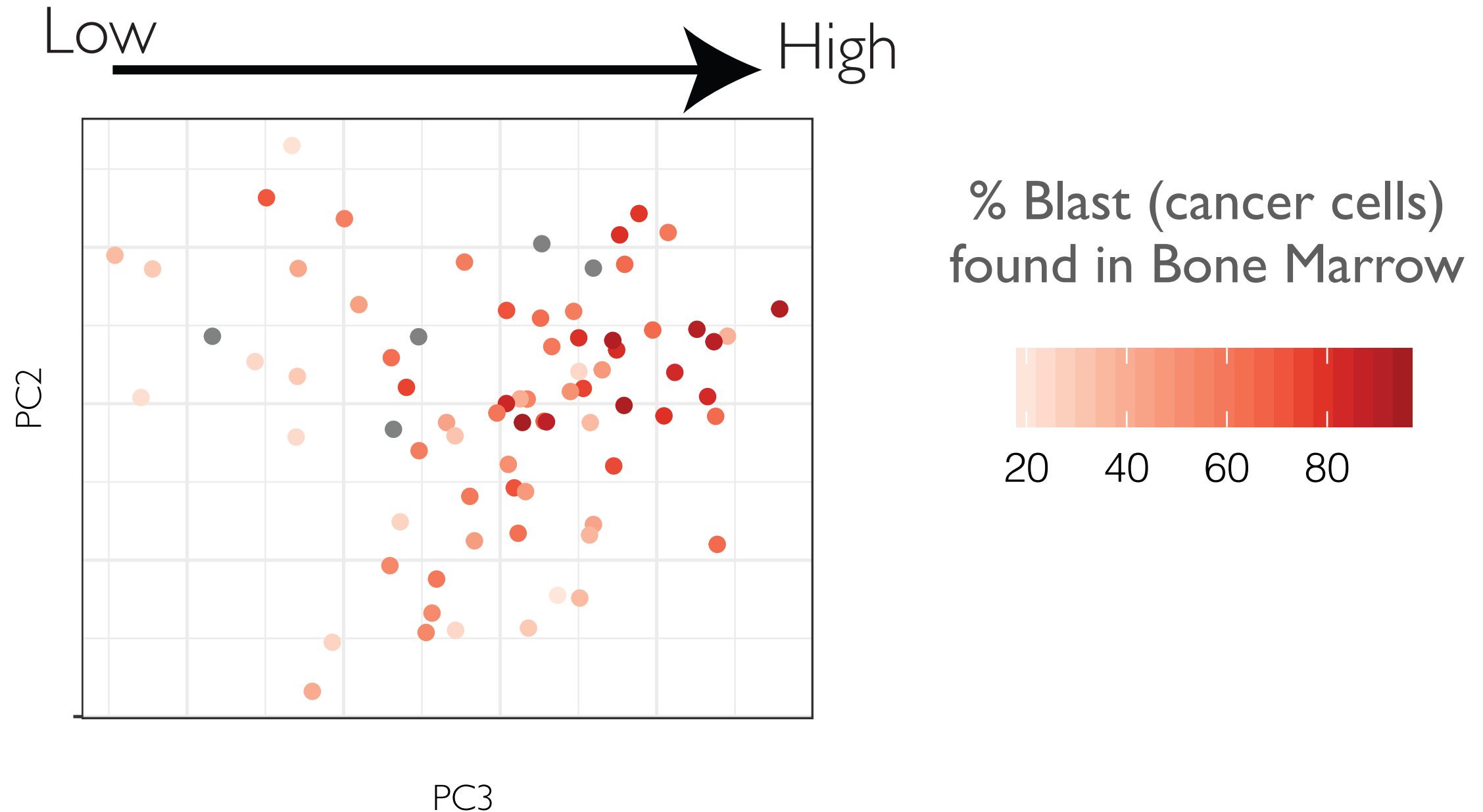
**Two fusion types**

But wait!

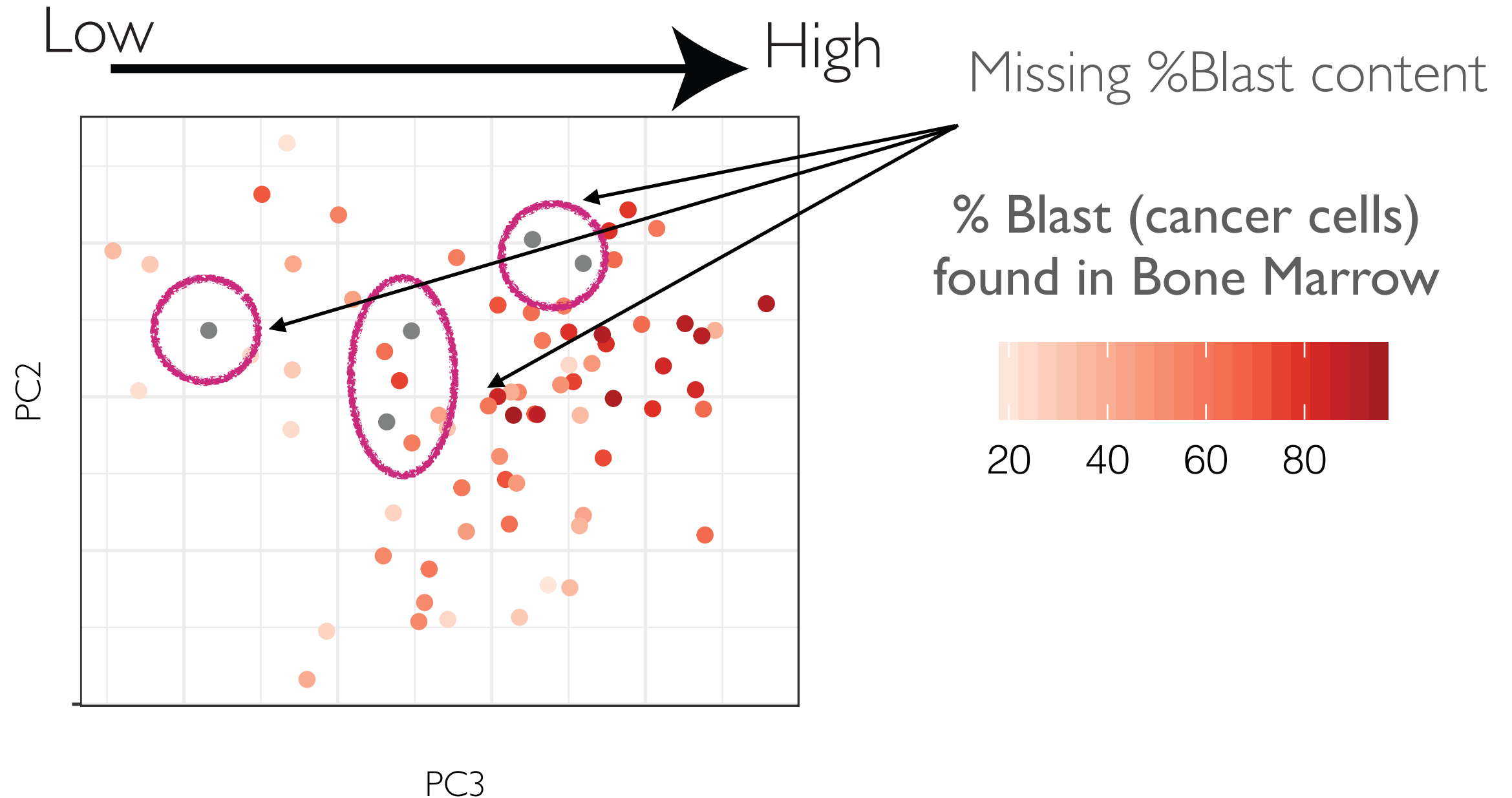
There is more....



# Tumour content



# Tumour content



# That's why we looked towards RUV!

## Removing Unwanted Variation

JA. Gagnon-Bartsch, L. Jacob, T. Speed



```
library(ruv)  
ruv::RUV4
```



$$Y_{m \times n} = X_{m \times p} \beta_{p \times n} + W_{m \times k} \alpha_{k \times n} + \epsilon_{m \times n}$$

↙  
Log2 gene expression matrix  
m samples  
n genes



Factor of interest:

Gene-wise comparison of *Relapse* vs *Long Remission* patients

$$Y_{mxn} = \boxed{X_{mxp}\beta_{pxn}} + W_{mxk}\alpha_{kxn} + \epsilon_{mxn}$$

Matrix with Unwanted Variation estimated  
directly from the data

$$Y_{mxn} = X_{mxp}\beta_{pxn} + W_{mxk}\alpha_{kxn} + \epsilon_{mxn}$$

Estimated using Negative Control genes

$$Y_{mxn} = X_{mxp}\beta_{pxn} + W_{mxk}\alpha_{kxn} + \epsilon_{mxn}$$

$$\beta_{NC} = 0$$

# Can we adjust for tumour content?



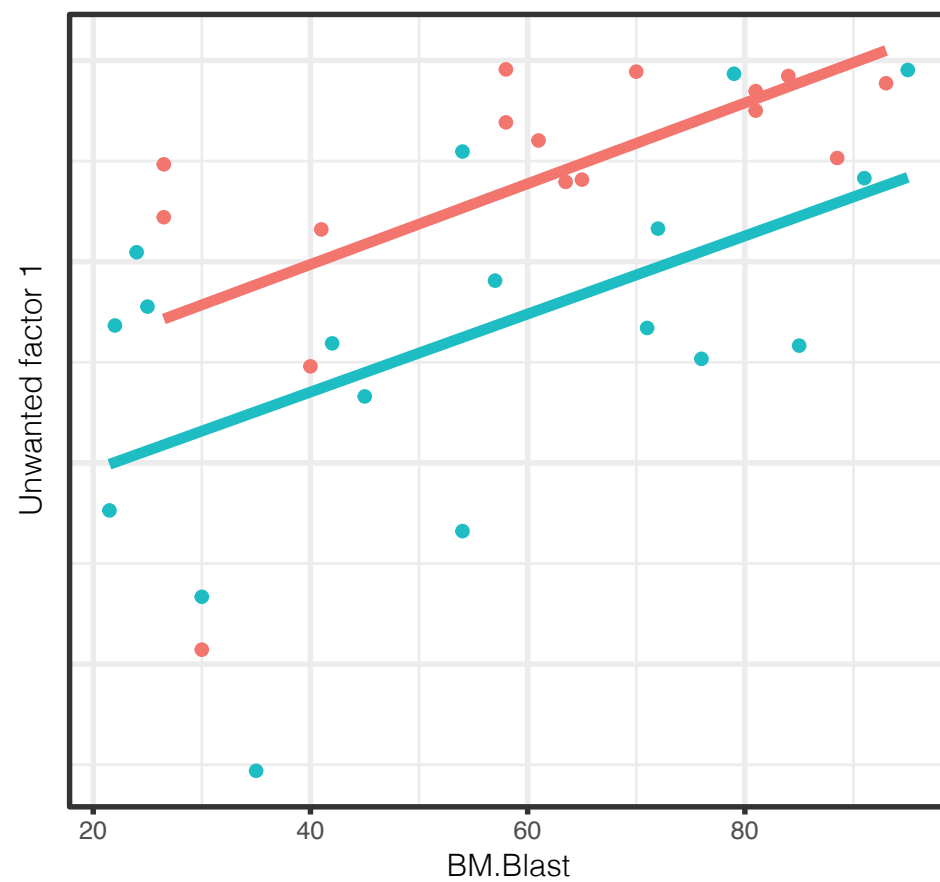
Let's investigate what the columns of  $W$  pick up...



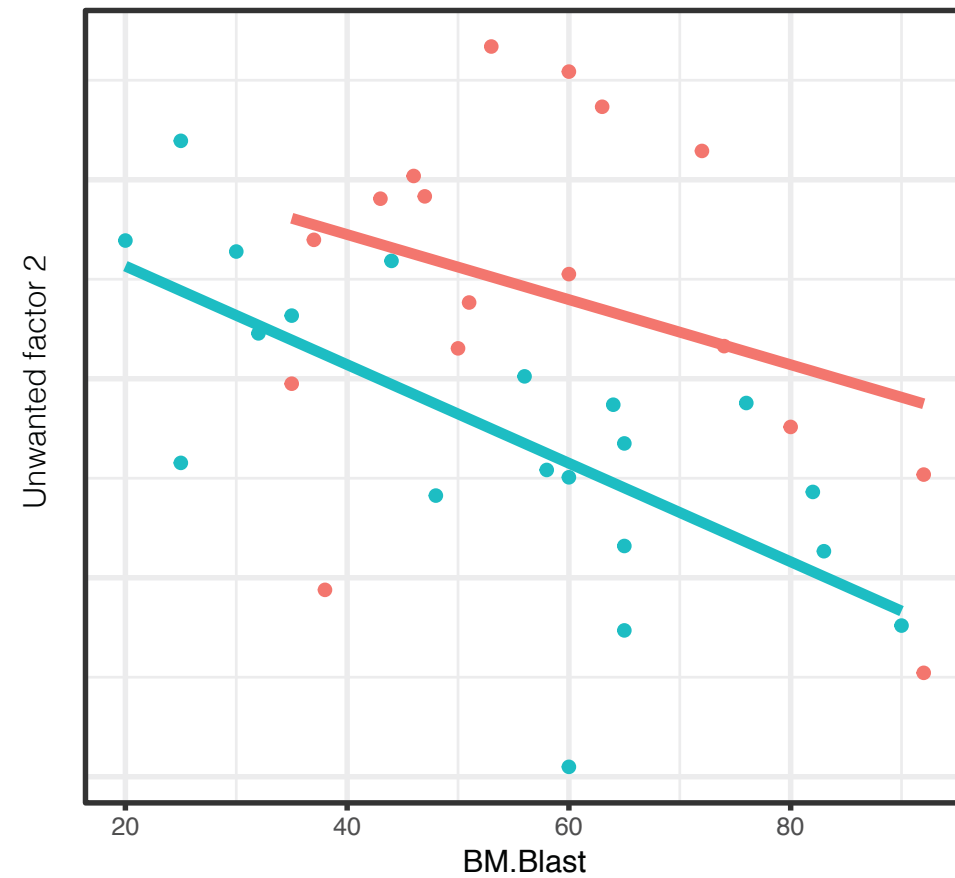
# Bone Marrow Blast (%) vs W columns

Tissue —●— BM —●— Blood

Only Australian samples

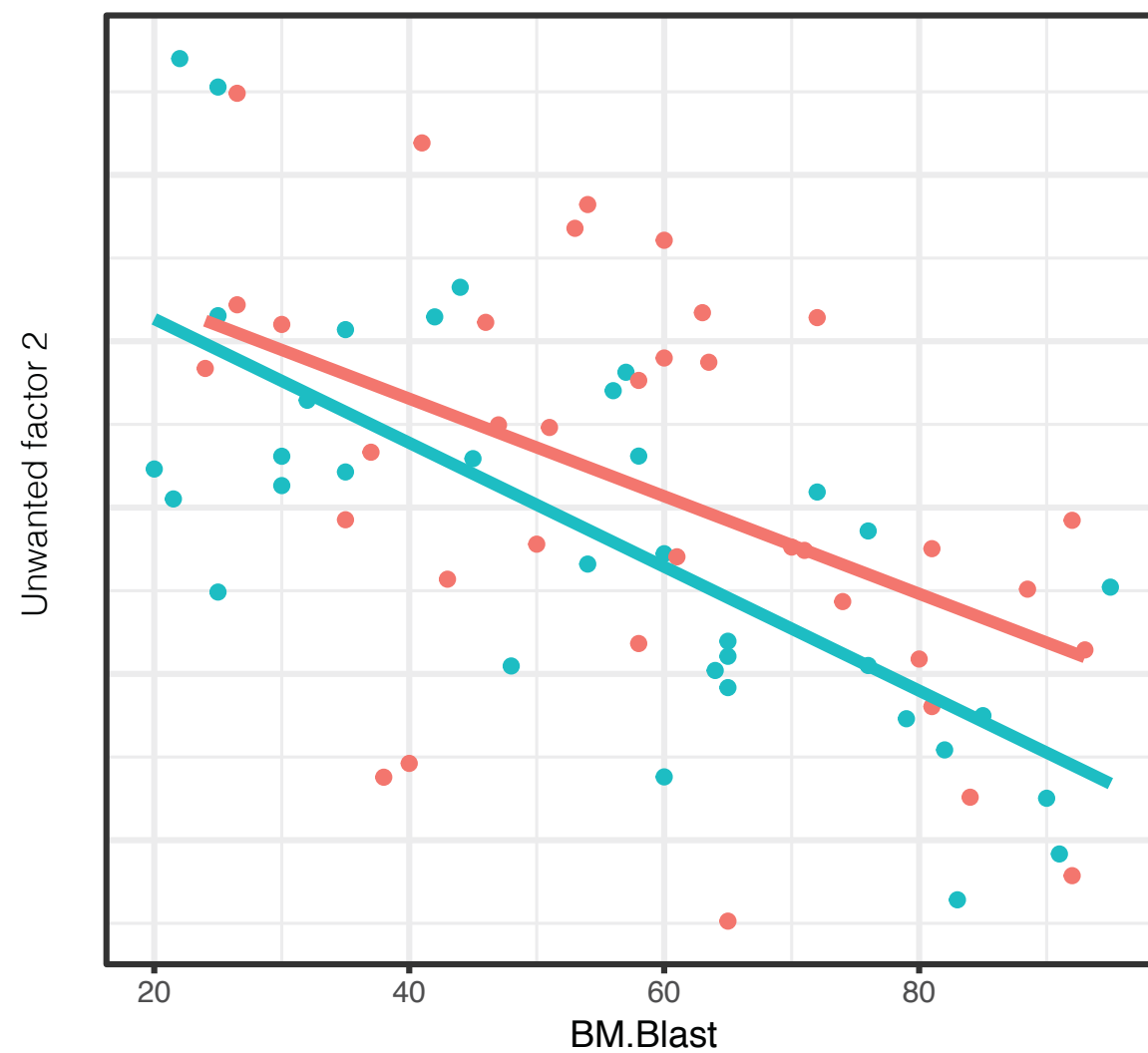


Only Canadian samples



# Bone Marrow Blast (%) vs W columns

Both cohorts



# Conclusions

😍 RUV-4 adjusts for tumour content

# Conclusions

- 😍 RUV-4 adjusts for tumour content
- 😍 Adjustment available for all samples

# Conclusions

- 😍 RUV-4 adjusts for tumour content
- 😍 Adjustment available for all samples
- 😍 RUV-4 adjusts for other sources of heterogeneity



# Conclusions

- 😍 RUV-4 adjusts for tumour content
- 😍 Adjustment available for all samples
- 😍 RUV-4 adjusts for other sources of heterogeneity
- 😊 Found some promising genes previously associated with sensitivity to chemotherapy are identified among the top DE genes.

# Challenges 🤔

- RUV-4 is challenging, more choices to make when using it!
- **Small cohort size** once you account for all the **heterogeneity**
- More research is required to **assess the improvement** provided by RUV-4 as well as the **reproducibility** of the signature in an independent cohort

# Thanks to...

Terry Speed

Ian Majewski

Edward Chew



**Walter+Eliza Hall**  
Institute of Medical Research

**DISCOVERIES FOR HUMANITY**



**Australian Government**



THE UNIVERSITY OF  
**MELBOURNE**