School of Molecular and Cellular Biology
FACULTY OF BIOLOGICAL SCIENCES
UNIVERSITY OF LEEDS

Machine learning and big data for prognosis and prediction in aggressive lymphomas

David R. Westhead

With thanks to

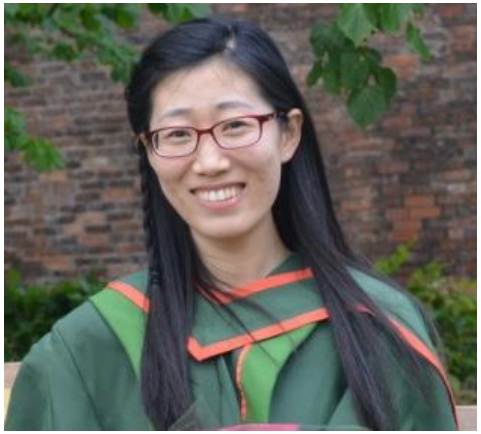
UNIVERSITY OF LEEDS

- Reuben Tooze (Leeds)
- Haematological Malignancy Diagnostic Service (HMDS), St. James Hospital
  - Andrew Jack, Cathy Burton, Sharon Barrans, Sophia Ahmed
- Peter Johnson, Andy Davies (Southampton)
- Ming Du (Cambridge)
- Jude Fitzgibbon (Barts)
- Anna Schuh (Oxford)
- Funding: Bloodwise and MRC

Michael Bentley

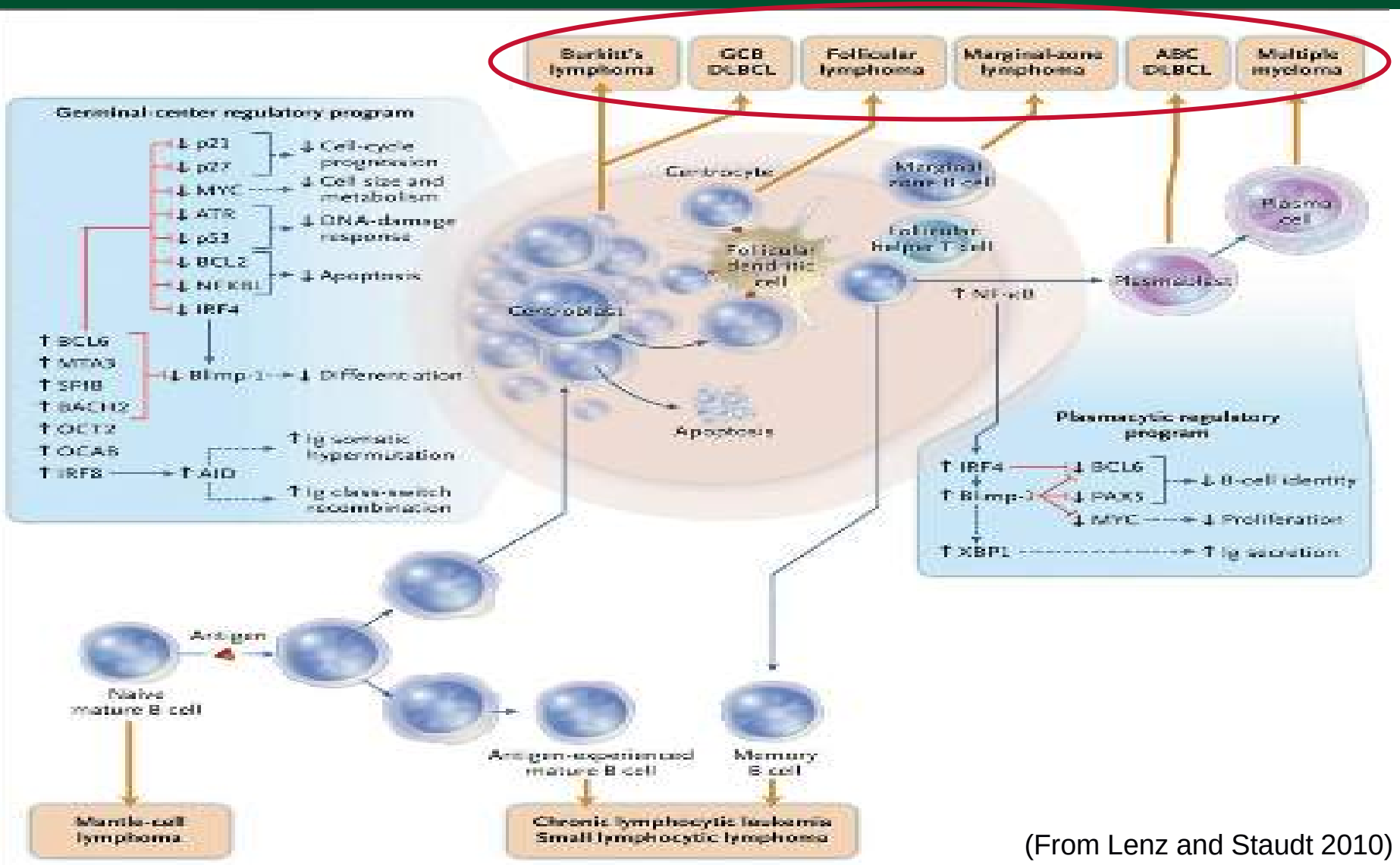
Matt Care

Chulin Sha



B cell related malignancies

UNIVERSITY OF LEEDS



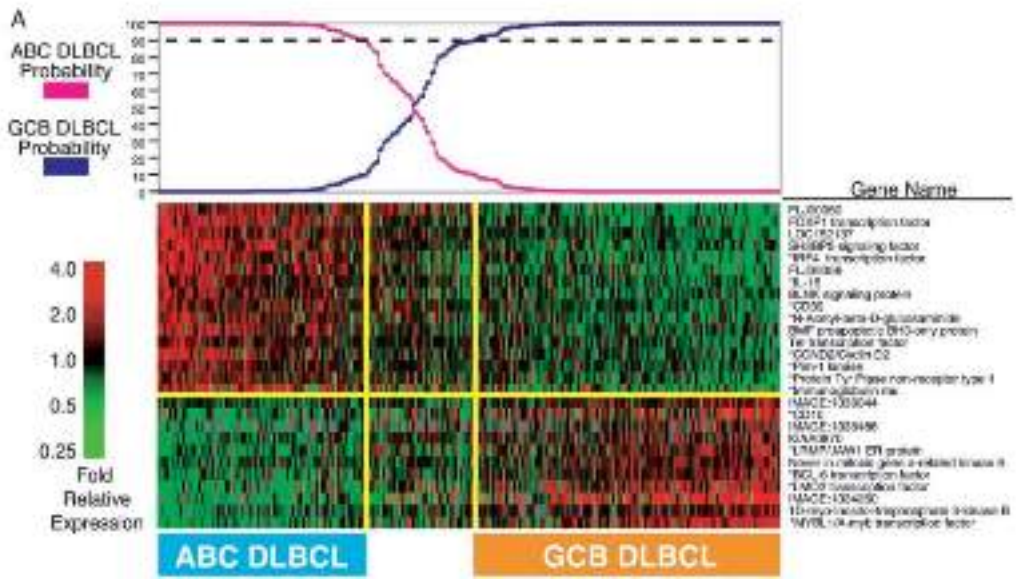
(From Lenz and Staudt 2010)

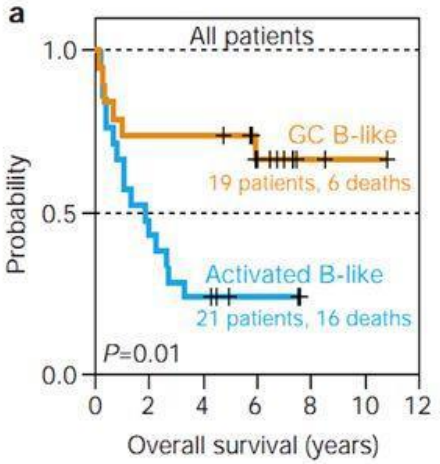
Where this all started: Classification of DLBCL

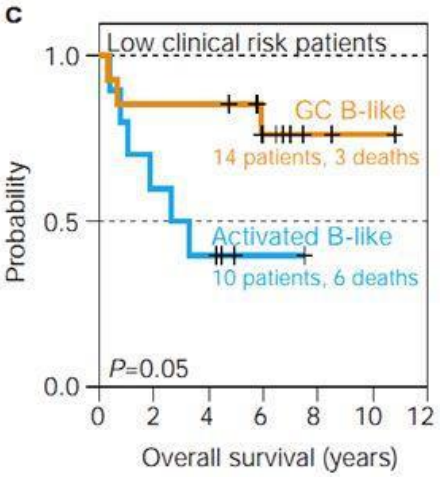
UNIVERSITY OF LEEDS

Alizadeh et al. (2000) recognised two distinct types based on gene expression

- ABC – activated B cell like
- GBC – germinal centre B cell like
- Different survival on (R-)CHOP therapy

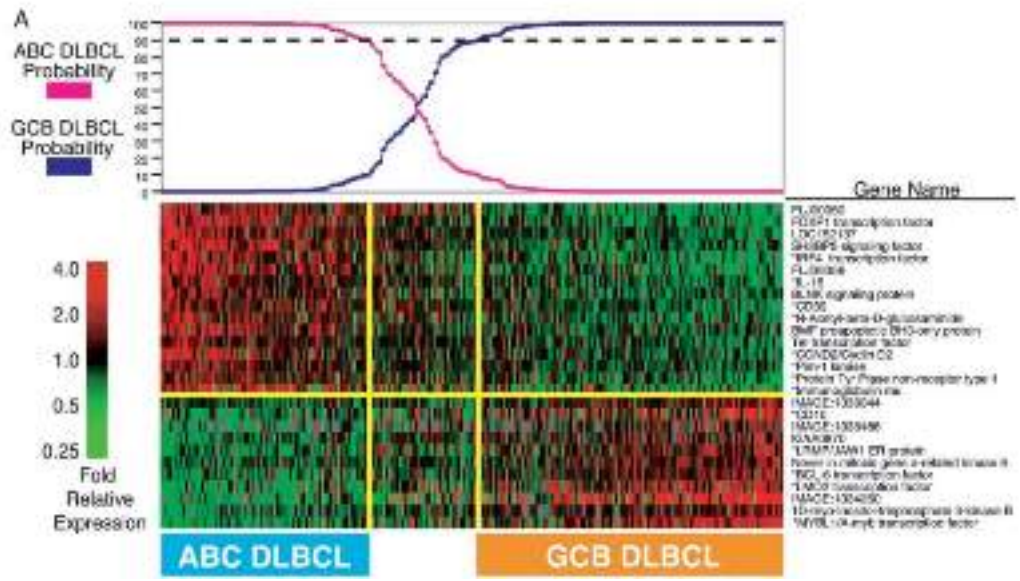






Where this all started

UNIVERSITY OF LEEDS

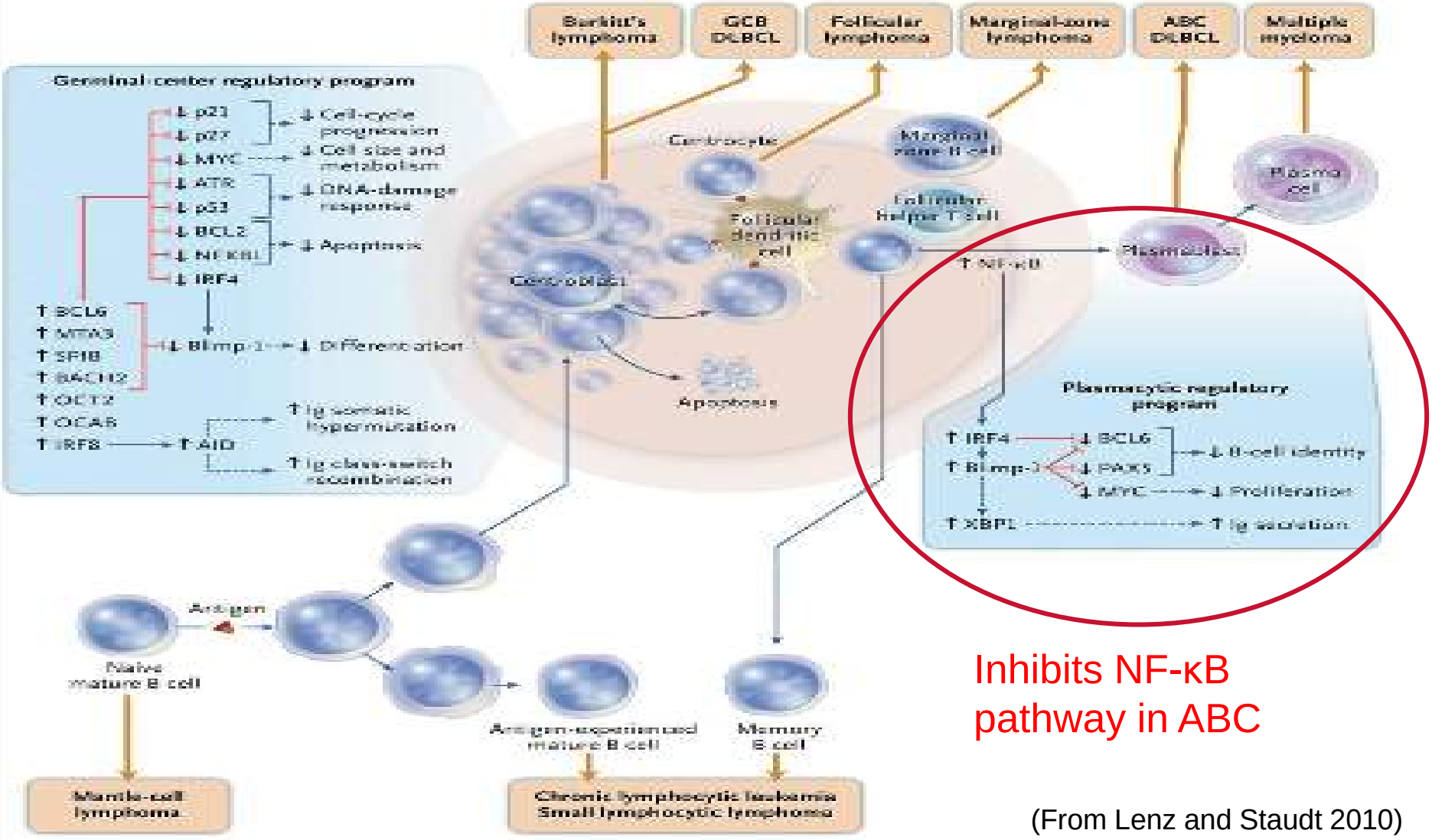


For the purposes of a clinical trial, we were asked if we could recapitulate this work based on gene expression measurements from formalin fixed paraffin embedded (FFPE) tumour biopsies.

Illumina DAZL array platform.

Trial: Bortezomib

UNIVERSITY OF LEEDS



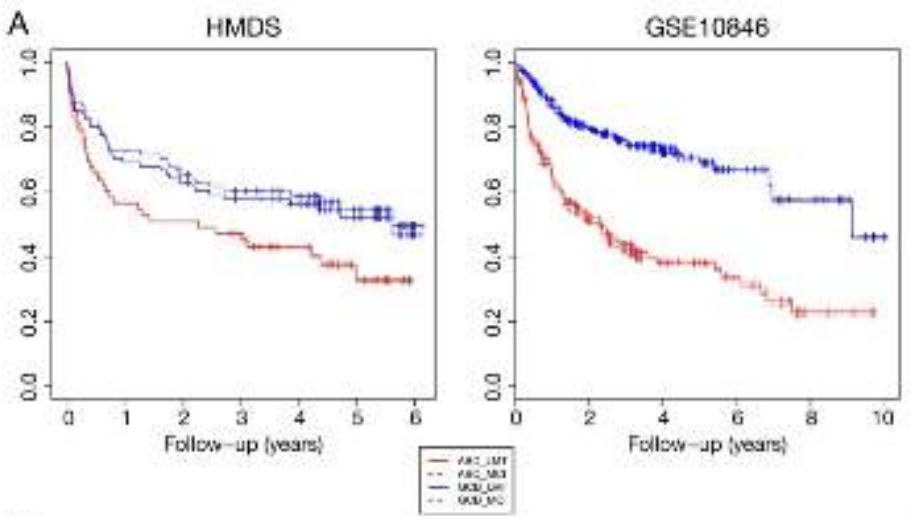
Inhibits NF-κB pathway in ABC

(From Lenz and Staudt 2010)

Gene expression classification works with FFPE samples

UNIVERSITY OF LEEDS

The best single classifier and the best meta classifier produce effective survival separation in both data sets



A great deal of effort: - the DAC classifier

Data quality and normalisation

Training

Transferability – data sets and platforms

Validation on several independent data sets

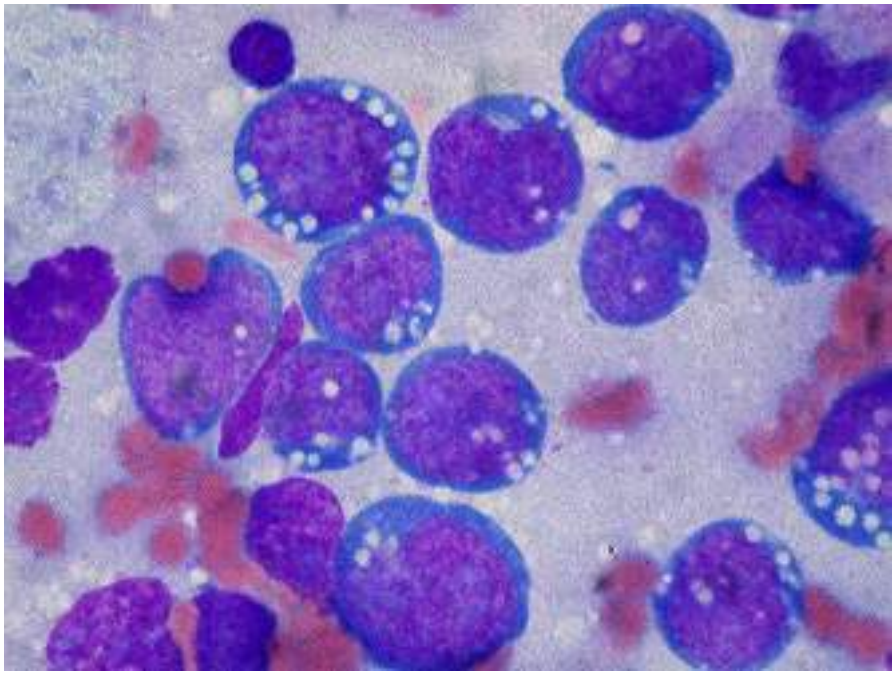
(Care MA, Barrans S, Worrillow L, Jack A, Westhead DR, Tooze RM A microarray platform-independent classification tool for cell of origin class allows comparative analysis of gene expression in diffuse large B-cell lymphoma. PLoS One **8** e55895-, 2013)

Burkitt's lymphoma

UNIVERSITY OF LEEDS

Burkitt's lymphoma

- Related to DLBCL
- More aggressive and fast growing
- Treatment is different
  - Typically R-CODOX-M/IVAC
- Again there are intermediate cases and classification is an issue



From Wikipedia…

Burkitt's lymphoma

UNIVERSITY OF LEEDS

Literature classifiers

- Two seminal works in this area
  - Dave et al. (2006), Hummel et al. 2006.
- Different approaches and gene numbers

Data Set	HGNC matched genes on platform	Genes used in authors' classifier	Classifier genes located in data[1]	Classifier genes available in other data set[2]
Dave	2411	217	214	172
Hummel	12495	58	58	28
Overlap	1913	21	21	-

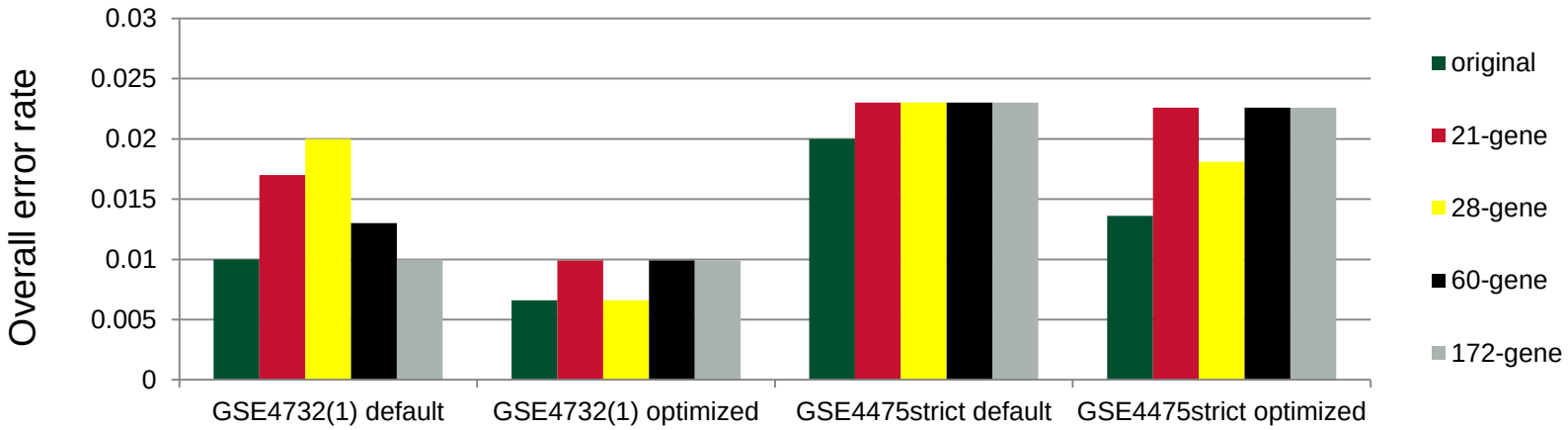
(Note that other classifiers have emerged since – Masque-Solar et al. 2013 – Nanostring – 10 genes.)

Cross-validation results

UNIVERSITY OF LEEDS

Re-capitulating the results of others

- Possible with reasonable accuracy
- Can work with a much reduced gene set!

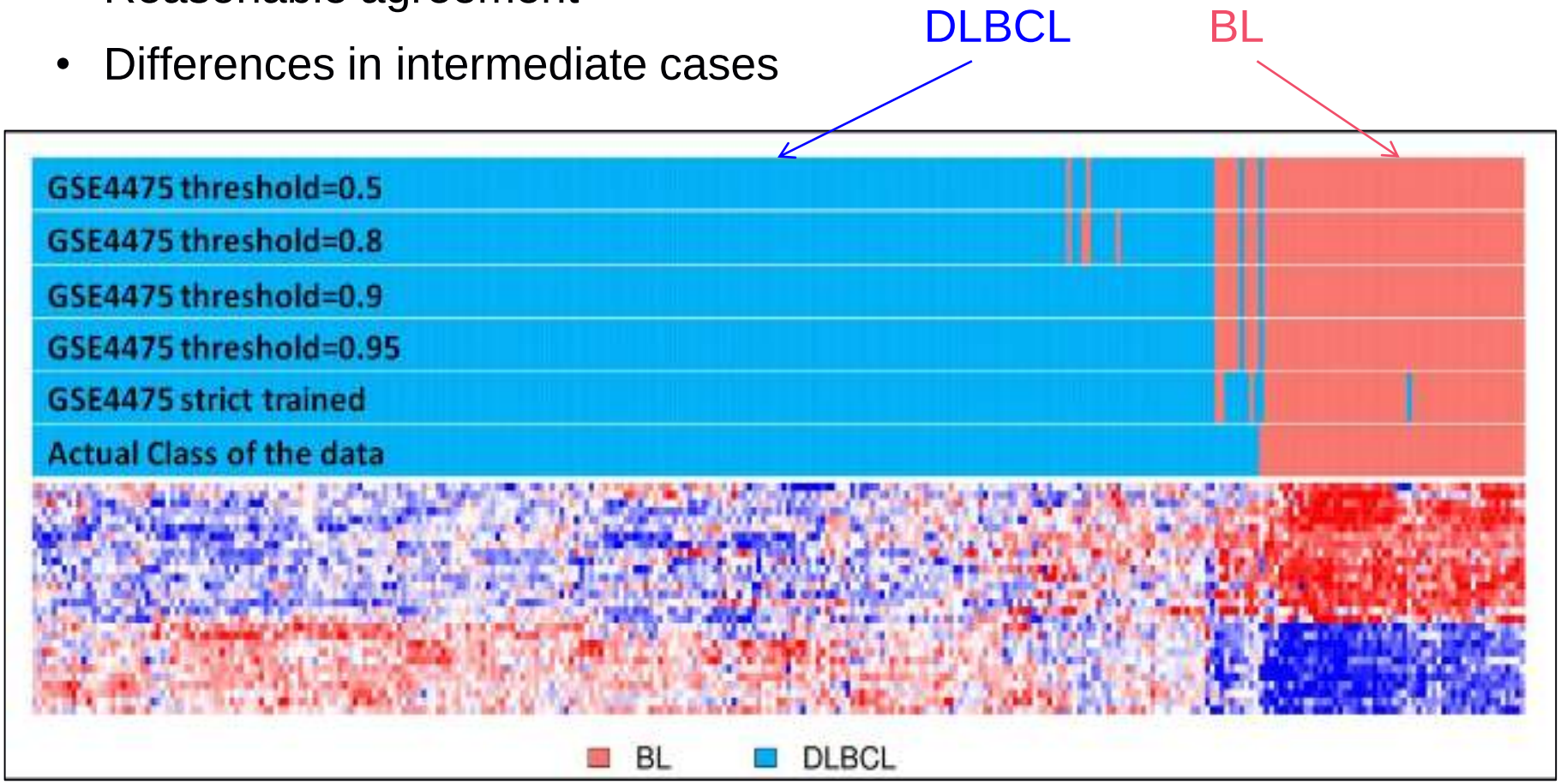


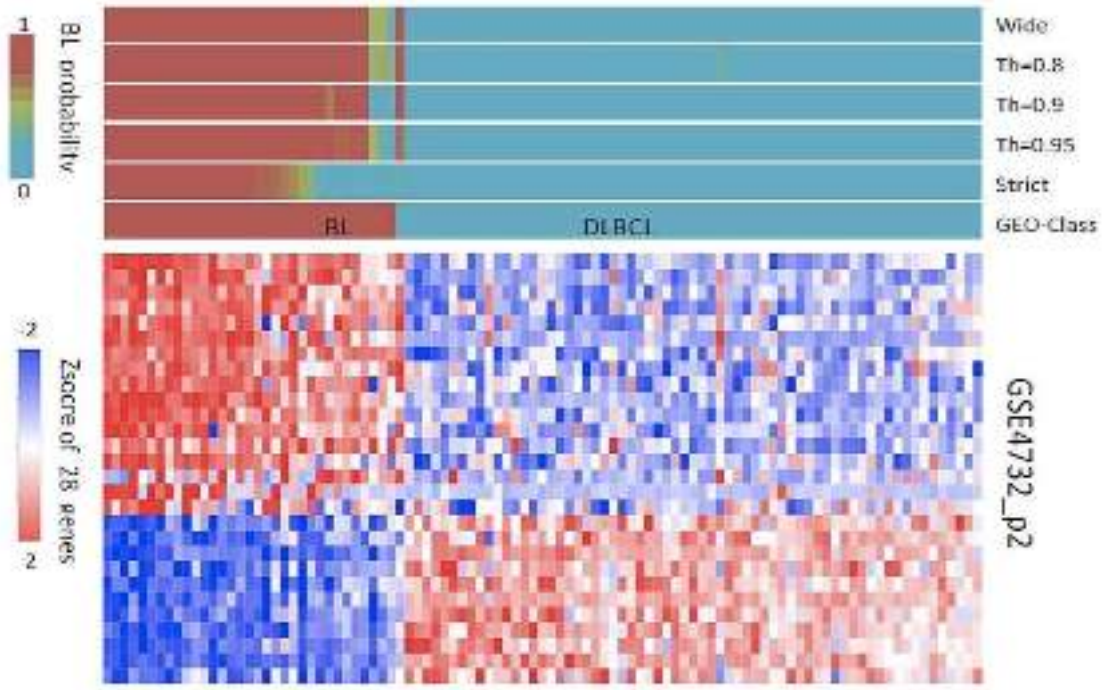
Transferability between data sets

UNIVERSITY OF LEEDS

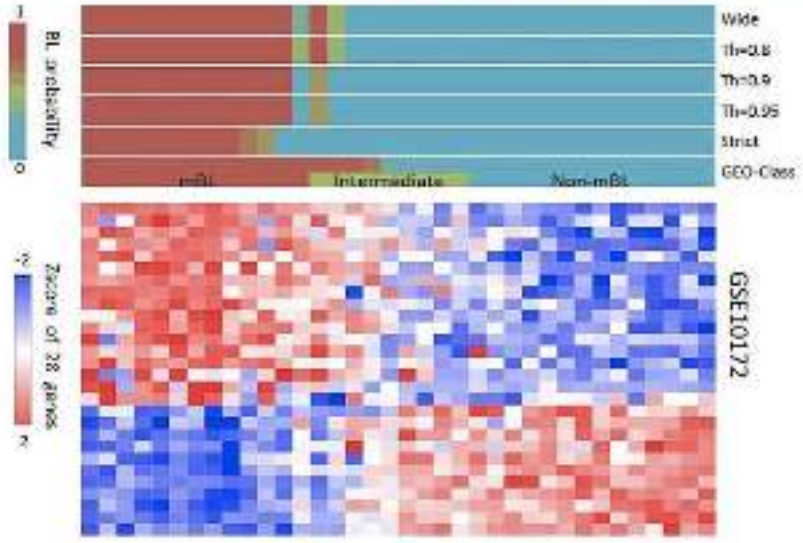
Train on one data set and definition, test on the other

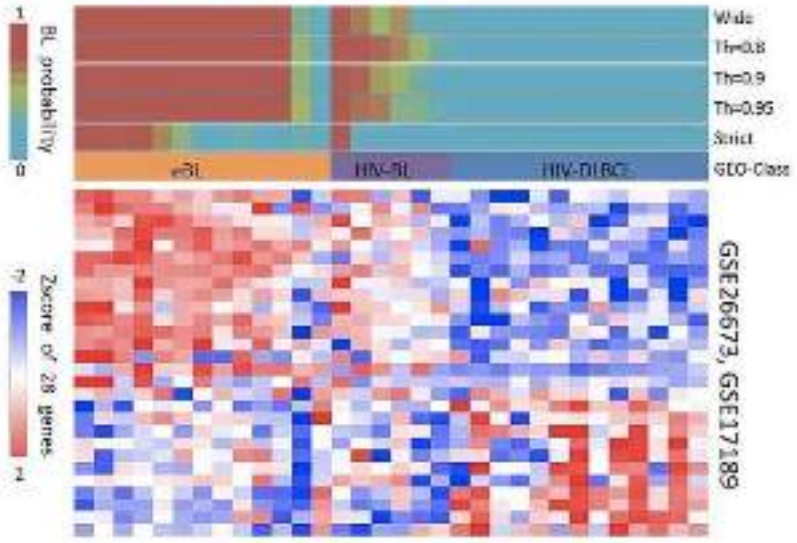
- Reasonable agreement
- Differences in intermediate cases





Applying to other public data sets



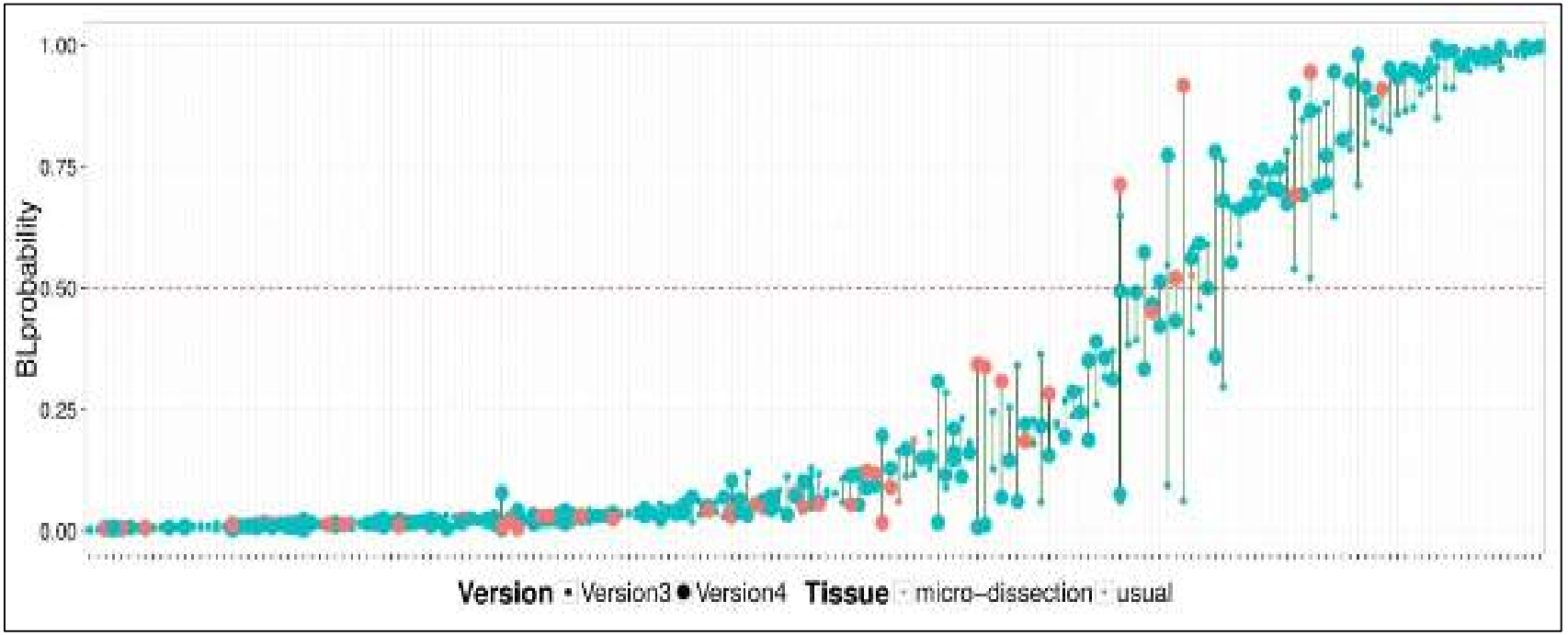


Our own clinical DAZL data

UNIVERSITY OF LEEDS

A careful look at reproducibility

- Different DAZL platforms
- Effect of careful micro-dissection



And with our clinical DAZL data

UNIVERSITY OF LEEDS

Pathology based diagnosis - good agreement

- Most interesting are those DLBCL(path.) diagnosed as BL (exp.)
- For these death rate for BL(exp) is high 74% - different treatment?

Table 5: Classification correlation with current clinical diagnosis

		Diagnosed			
		BL	DLBCL(MYC-R)	DLBCL	
Classified	BL	61 (85%)	14 (30%)	34 (4.5%)	109
	DLBCL	11 (15%)	33 (70%)	720 (95.5%)	764
		72	47	754	

Sha C, Barrans S, Care MA, Cunningham D, Tooze RM, Jack A, Westhead DR Transferring genomics to the clinic: Distinguishing Burkitt and diffuse large B cell lymphomas. **Genome Medicine 7 -, 2015**.

Big data for an interesting group of patients

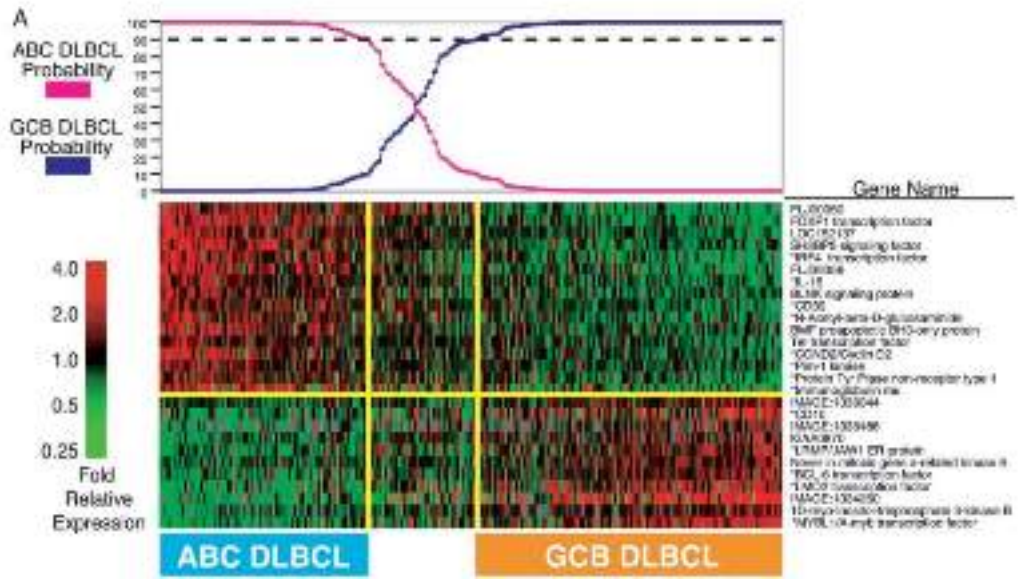
Somatic mutation data now including – but difficult for FFPE

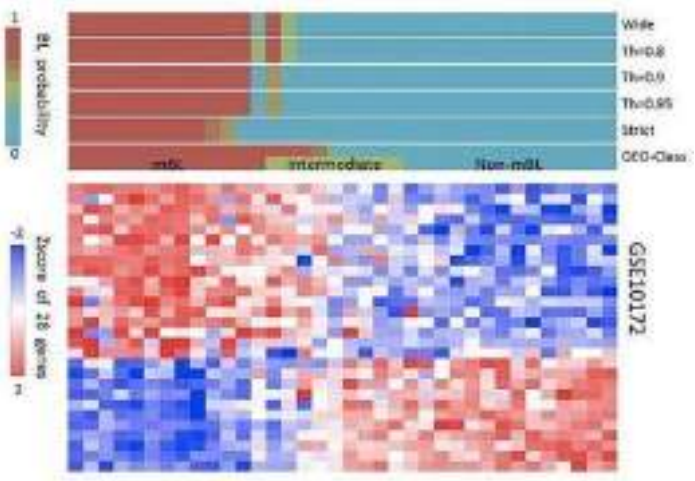
Diagnosed as DLBCL (path) but significant numbers of BLs by our classifier

LMO2…

Grey areas: intermediates cases

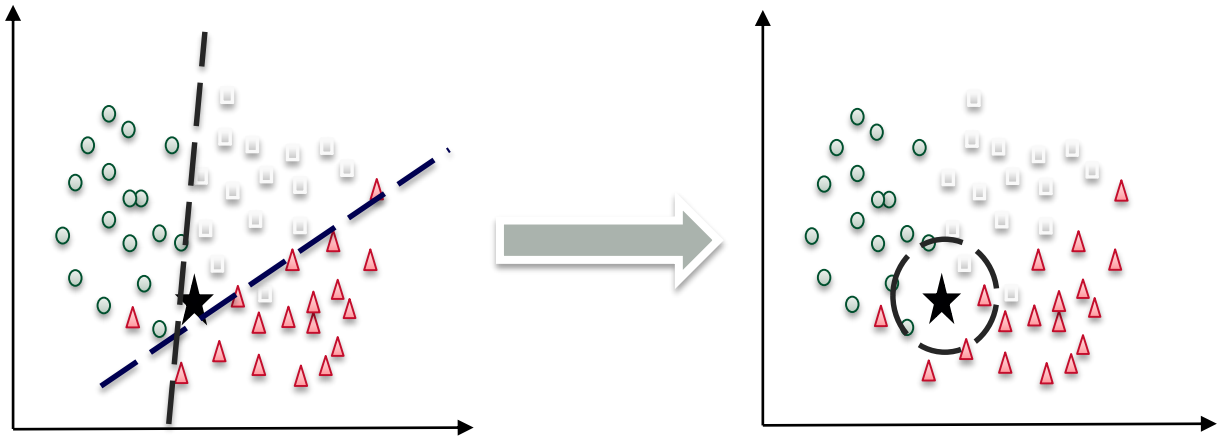
UNIVERSITY OF LEEDS





Similarity searches vs classifications

UNIVERSITY OF LEEDS

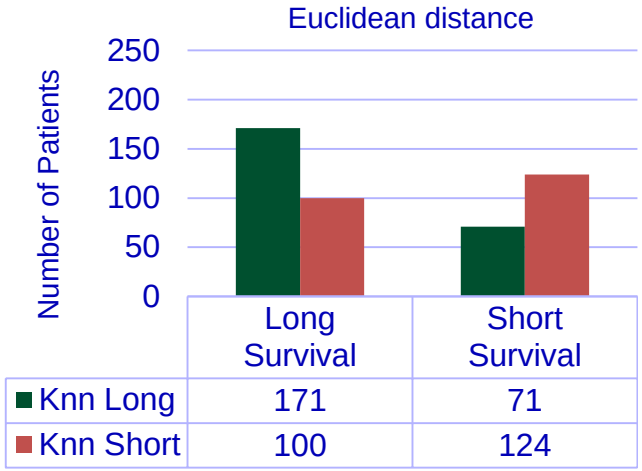


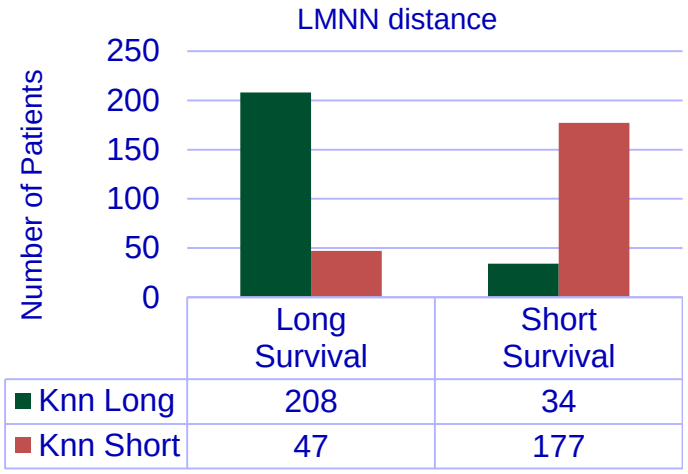
Use ‘similar’ or ‘neighbour’ cases in prognosis and prediction

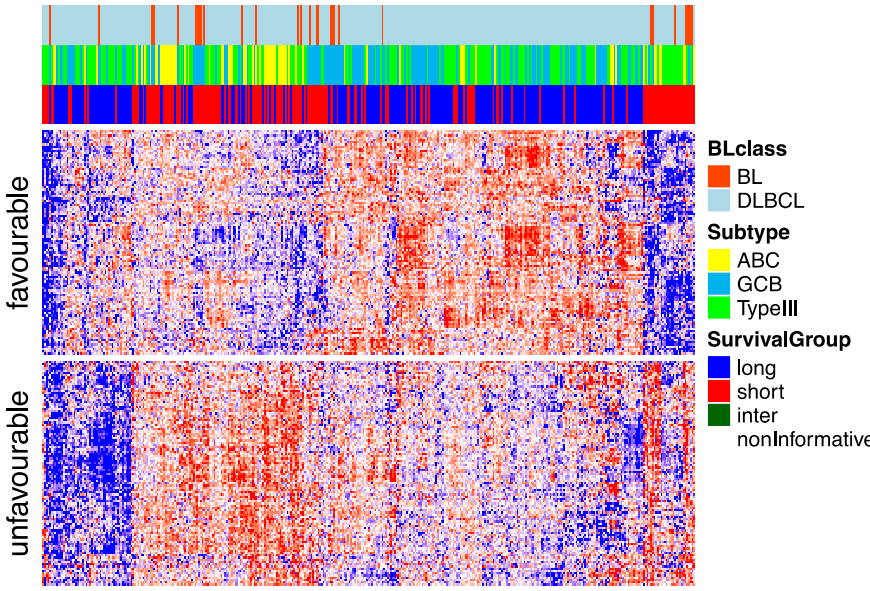
Requires machine learning of a gene expression based similarity or distance metric

Learning a similarity measure

UNIVERSITY OF LEEDS







Conclusions

UNIVERSITY OF LEEDS

Summary

- Gene expression classifiers using data from FFPE can work and can be useful
  - For BL and DLBCL subtypes
- Our classifiers have been evaluated carefully
  - Transfer between data sets – local and public
  - Transfer between different expression measurement platforms
- Watch this space for clinical trial results
- Beginning to progress with category free methods
  - Similarity searches

Thank you for listening

UNIVERSITY OF LEEDS