Effects of Mutations in Histone Modifying Enzymes on Gene Expression Profiles in Non-Hodgkin Lymphoma

Alex Hammel, Andrew Wong,
James Proudfoot, Julia Pon

Non Hodgkin Lymphoma

Cancers of B and T cells of the immune system
Fifth most common cancer in Canada [1]

~35%

Diffuse Large B-cell Lymphoma (DLBCL)

Gene expression profile differences
(histologically indistinguishable) [2]

Germinal Centre B-cell (GCB)

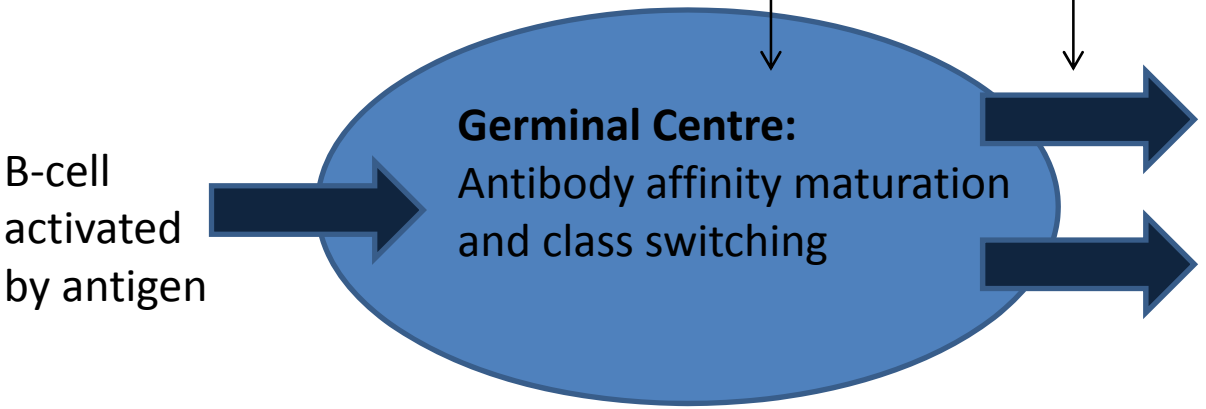
80% 5 year survival [3]

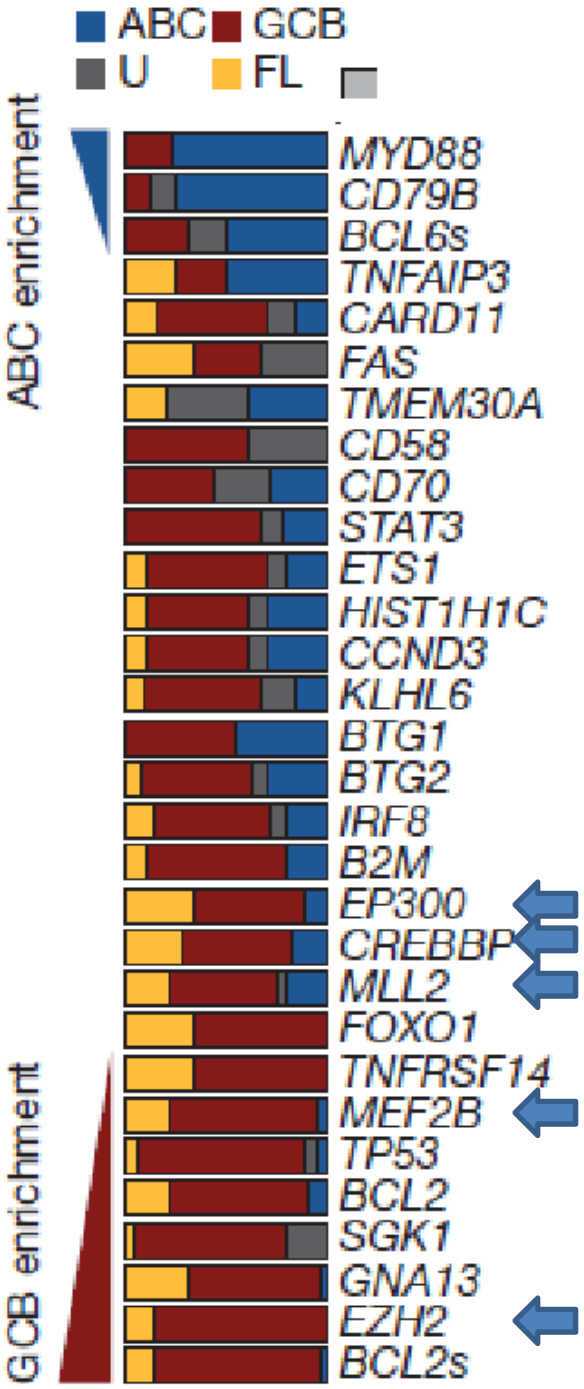
Arises from B-cells in the germinal centre [2]

Activated B-cell (ABC)

50% 5 year survival [3]

Arises from cells poised to exit the germinal centre [2]





Adapted from Morin et al 2011

"Frequent Mutation of Histone Modifying Genes in non-Hodgkin Lymphoma"

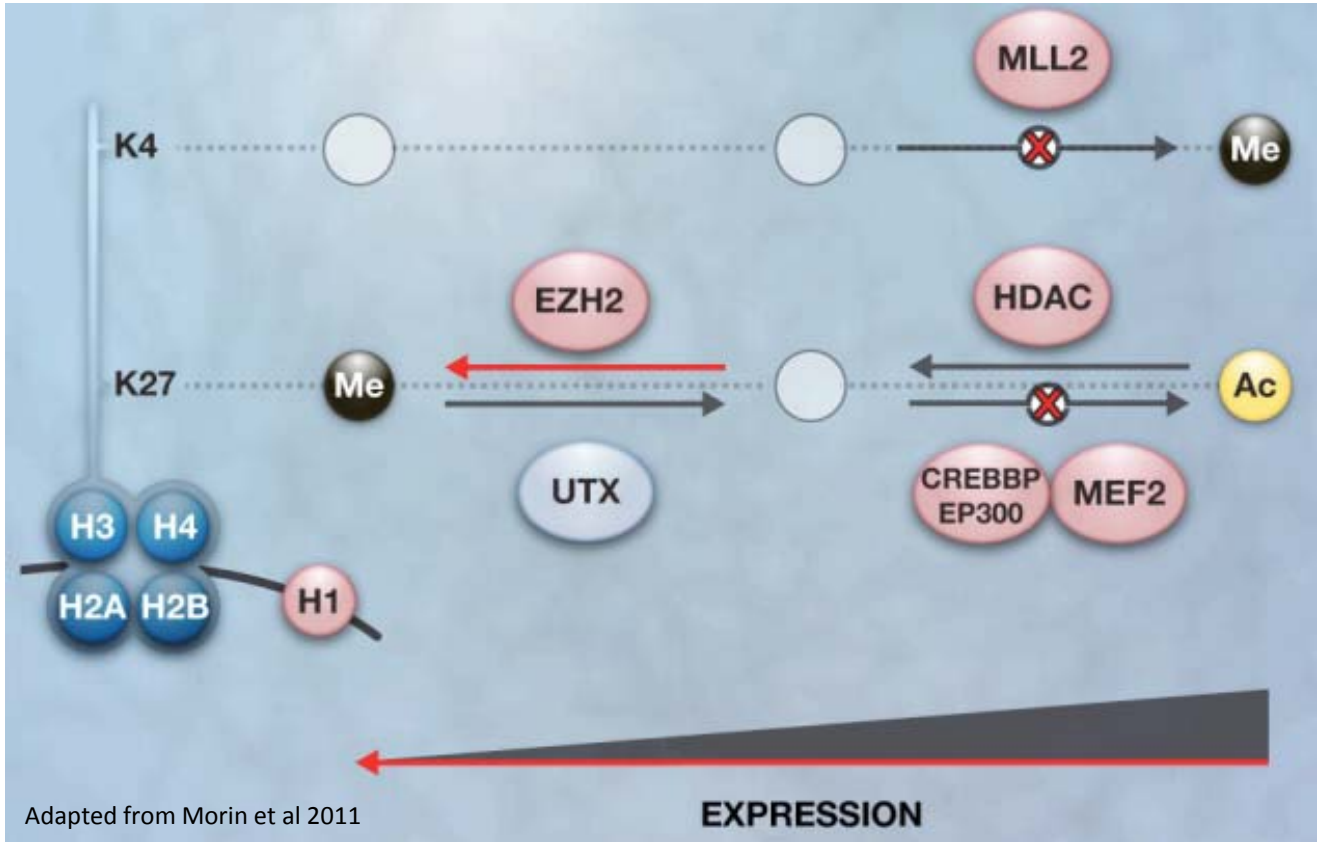
Morin et al, Aug 2011, Nature[4]

Histone modifying gene	% of DLBCL with somatic mutation
CREBBP	24
EP300	8
MLL2	32
MEF2B	11.4
EZH2	22
HDAC7	6.3

The frequency of some of these mutations suggests they provide a survival advantage to lymphoma

Nuclear DNA is organized by wrapping around histone complexes. Histone modification regulates gene expression by affecting how tightly the DNA is wrapped around histones, and providing binding sites for other regulatory proteins.

Hypothesis: Mutations in histone modification genes (pink) tend to decrease gene expression and may regulate common genes.



Adapted from Morin et al 2011

Key questions

1. How does mutation of each histone modification gene change the global gene expression profile?
   - This will suggest molecular interactions and cellular functions of each histone modification gene

2. What are the similarities and differences between the expression profiles associated with mutations in different histone modifying enzymes?
   - targeting a common pathway could be an effective therapy in a large proportion of patients

   - Differences between expression profiles may suggest how cancers with certain mutations could be treated differently.

RNA Sequencing Data

Raw read counts for 35,863 genes in 95 different DLBCL's and 9 samples of non-malignant B-cells from inflammed tonsils

Cancer subtype	Number of cases
GCB DLBCL	53
ABC DLBCL	32
unknown	10

Histone Modification Gene	Cases with mutation
EP300	4
CREBBP	14
MLL2	27
MEF2B	14
EZH2	21
HDAC7	6

Methods

Pre-processing and Data Normalization

- Probes with expression in less than 70 samples were excluded from the analysis.
- Normalization factors for each sample library size was computed via TMM (Trimmed Means of M Components) (Robinson 2010)
- Read counts were then converted to $\log_2$ counts-per-million, with observation-level weights interpolated from an estimated probe-wise mean-variance relationship

Sample Clustering

- Sample clustering and dendrogram bootstrapping was performed using R package pvclust[9] on normalized data. Bootstrapping scale sizes ranged from 0.5 to 1.4 fold of the data set size, with 10000 replicates performed for each.
- Normalized data were subjected to centred, scaled principal component analysis.

Assessing Differential Expression

Differences between tumor types is tested via a one-way ANOVA with the following linear model:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where Y_{ij} is the $\log_2$-cpm normalized data for sample j with tumor type i, μ is an overall mean associated with the gene, τ_i is the effect of tumor type i, and ε_{ij} is a normally distributed error term.

Differences between wildtypeand mutant genes is tested via a two sample t test, as implemented in limma (Smyth, 2004)

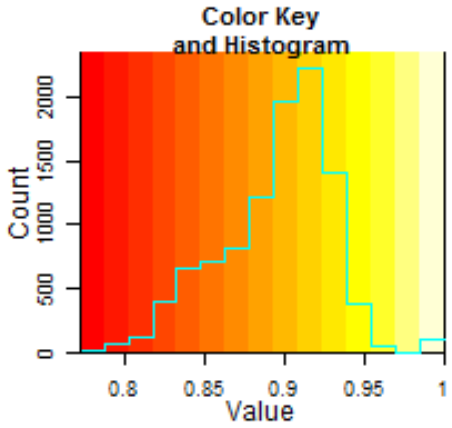
In both cases, test statistics are moderated by an empirical Bayes shrinkage of standard errors

Analysis of Differentially Expressed Gene Lists

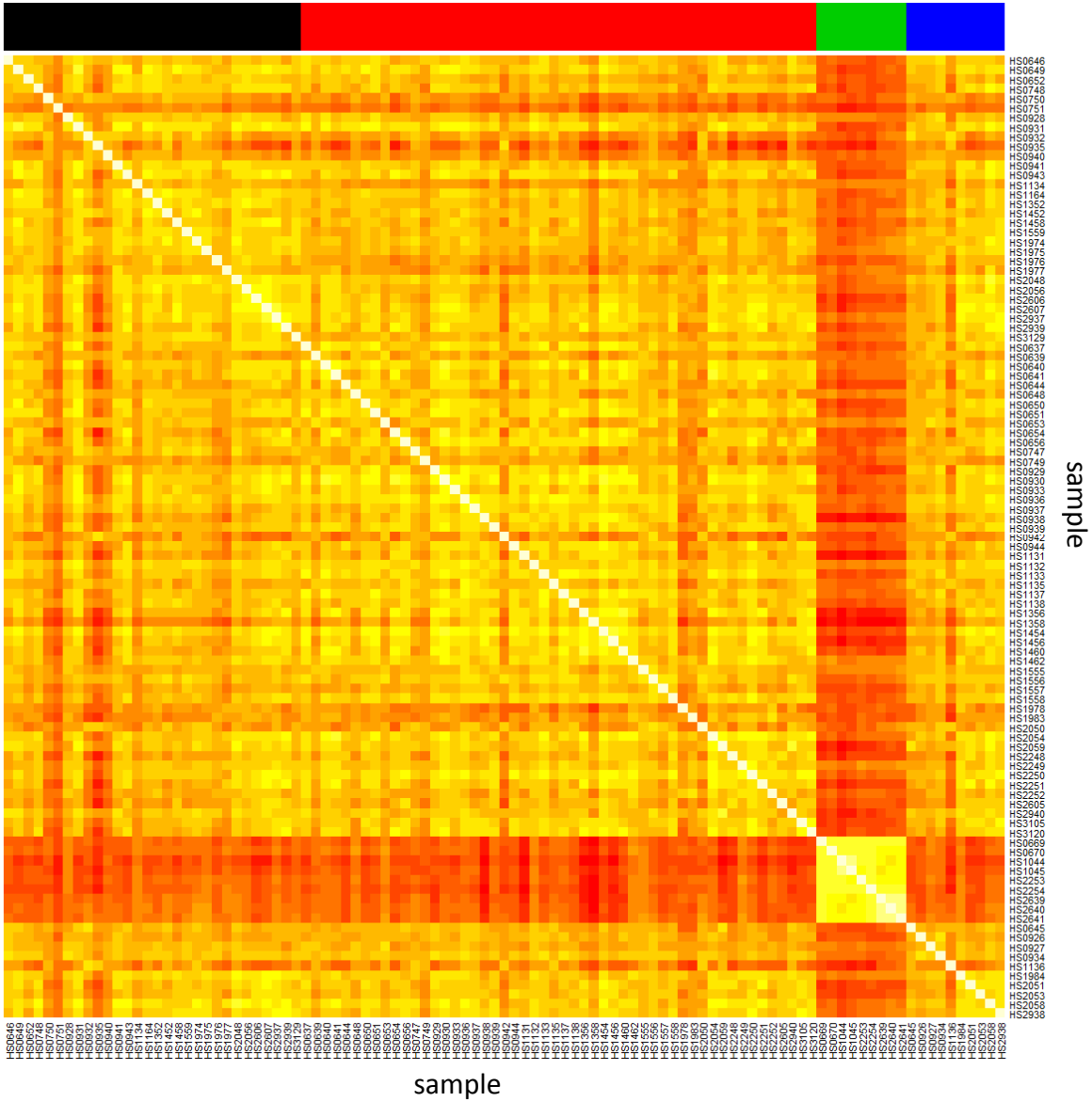
- The intersections of sets of differentially expressed genes were compiled using a Python script
- Pathway enrichment analysis was done using the DAVID Functional Annotation Tool. Pathway information was generated by KEGG.

Preliminary Analysis: Correlation Coefficients Between Samples

Using normalized data



- ABC DLBCL
- GCB DLBCL
- Normal B-cells
- Unknown DLBCL subtype

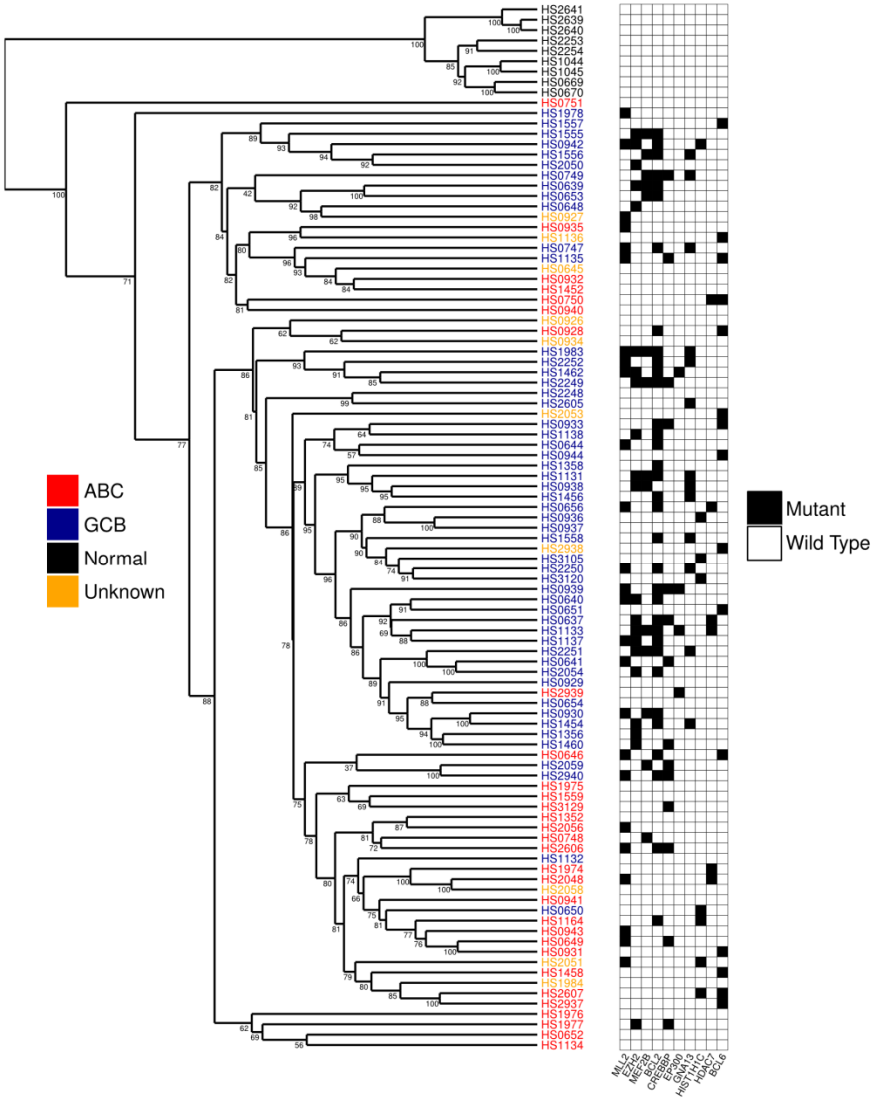


Cancer and non-cancer samples have low correlation, as expected

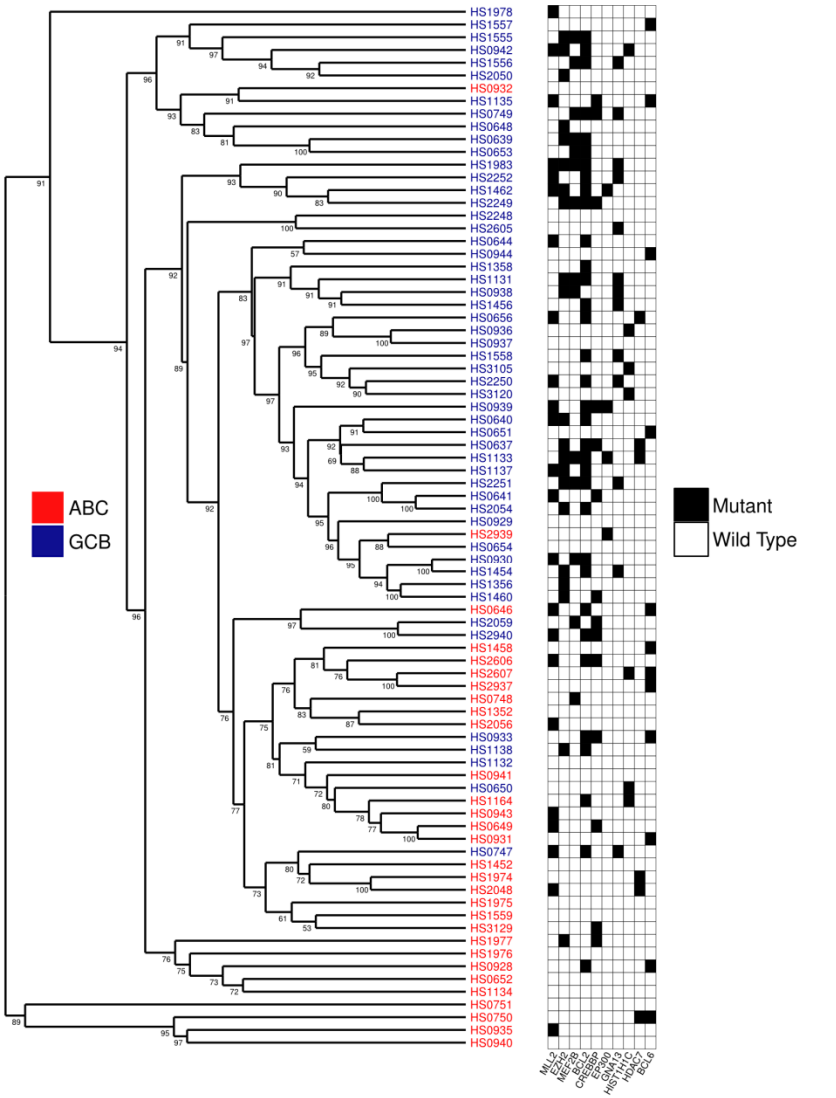
No samples are strong outliers and none will be removed from analysis

Result I: Differential gene expression distinguishes GCB and ABC

Tumor vs non tumor samples cluster using normalized whole dataset



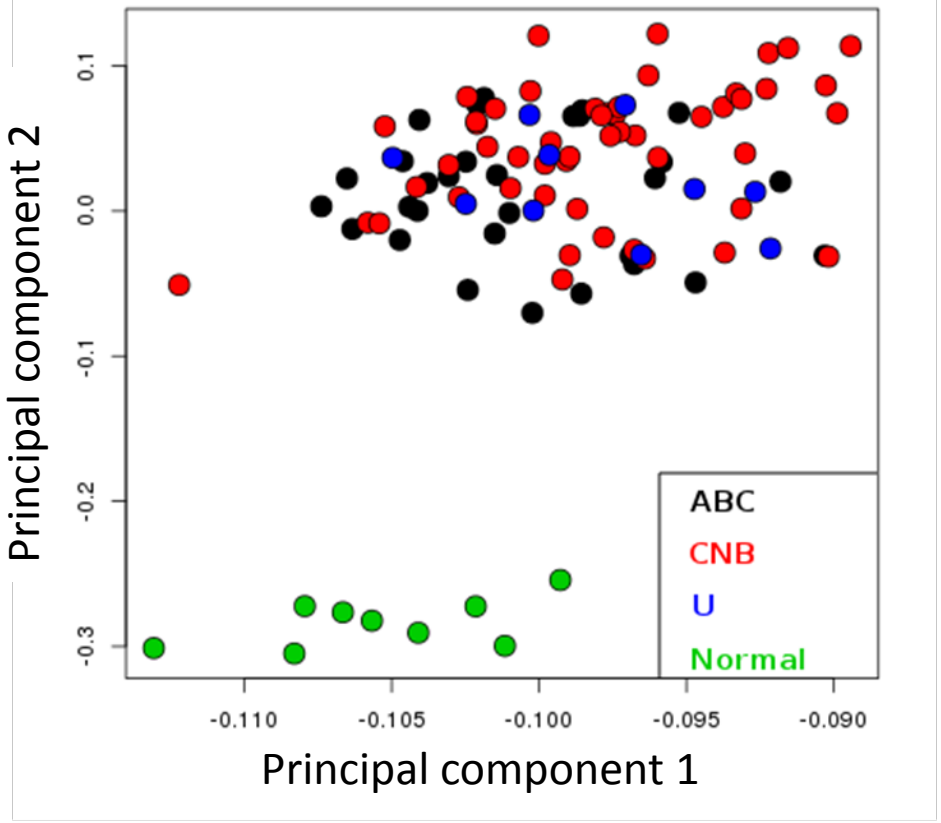
ACB vs GCB samples cluster using normalized data for just ABC and GCB samples



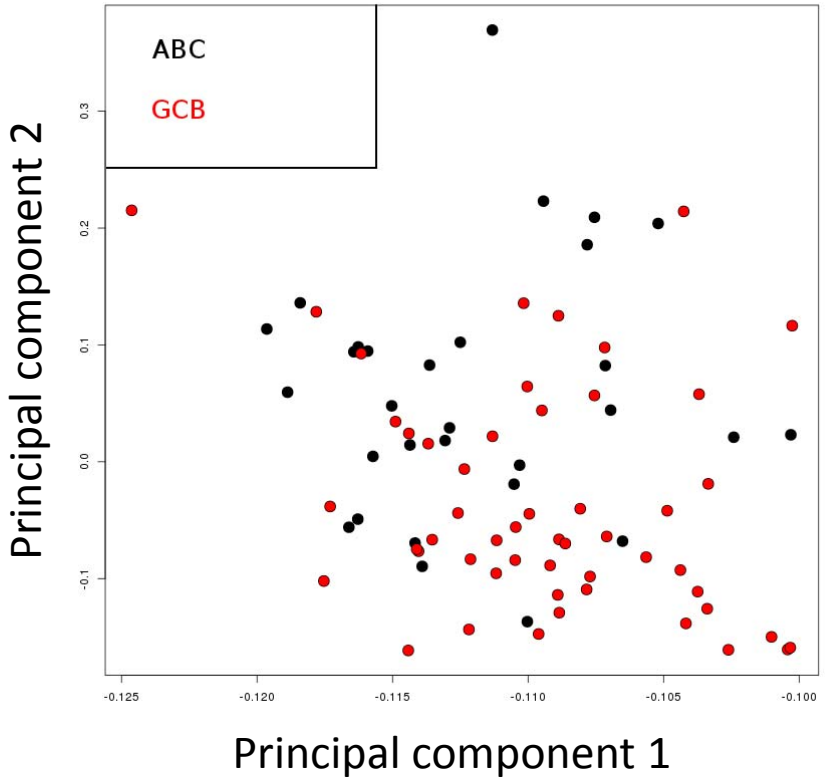
Junctions are labelled with approximately unbiased p-values from multiscale bootstrapping

Principal component analysis (PCA) shows clustering of tumor vs normal samples, but not ABC vs GCB

PCA of normalized data for all samples



PCA of normalized data for just ABC and GCB samples



In both cases, ~80% of the variance was accounted for by the first principal component.

5,781 genes were differentially expressed in ABC vs GCB (FDR 0.05)

Differentially expressed genes:

Include three of the somatically mutated genes investigated here:

- EZH2
- MEF2B
- HDAC7

Include other regulators of histone modification:

- MEF2C
- HDAC1, 2, 3, 11
- CBP/p300 interacting transactivator
- Creb regulated transcriptional coactivators 1 and 2
- H1 and H2A histone families
- Histone cluster 1

Show enrichment for key signalling pathways

Signaling pathway	Modified Fisher exact P-value for enrichment
TGF-beta	0.0013
MAPK	0.019
Wnt	0.042
EGFR	0.08
p53	0.087

Show the greatest enrichment for:

Pathway	Modified Fisher Exact P-value for enrichment
Cell cycle	$1.4x10^{-5}$
ECM receptor interaction	$1.1x10^{-4}$
Colorectal cancer	$1.1X10^{-4}$
Focal adhesions	$3.1x10^{-4}$
Base excision repair	$6.1x10^{-4}$
Mismatch repair	$7x10^{-4}$

Prognosis is poorer for ABC than GCB cases. This may be because of:

- Altered focal adhesion and ECM receptor pathways may contribute to changes in cell motility and altered dependency on substrate for growth, affecting the likelihood of metastasis
- Altered repair pathways may affect mutation rate, affecting the rate of accumulation of driving mutations or treatment resistance
- Alterations in numerous proliferation and apoptosis pathways affect tumor cell abundance

Differential expression of histone modifiers between GCB and ABC:

•Supports that altered histone modification, including that resulting from the somatic mutations in histone modifiers investigated here, plays a role in clinical differences between GCB and ABC tumor subtypes.

Results validate ABC vs GCB clustering, as reported in literature:[2,8]

- 15 of the 19 top differentially expressed genes from literature[8] were differentially expressed in our analysis
- Only one gene from the literature's 19 top genes[8], RASL11A, was present in our top 50 up and down regulated genes.

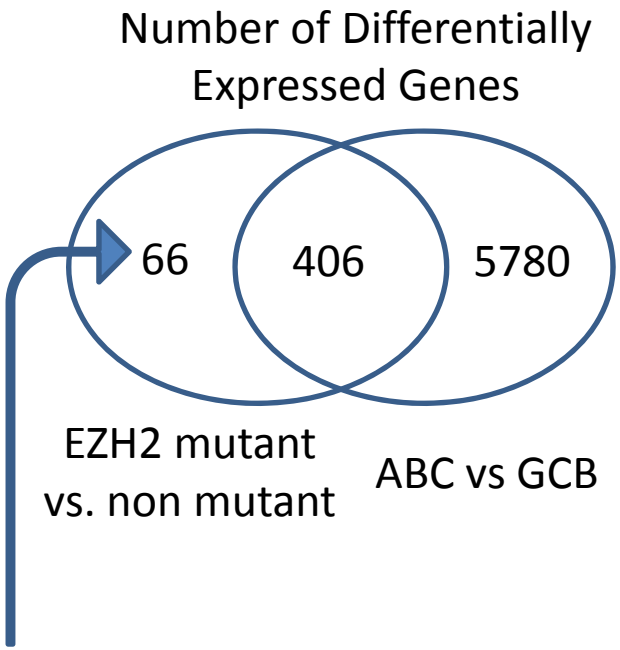
Result II: Differentially expressed genes in EZH2 mutant tumors show strong overlap with genes differentially expressed between tumor subtypes

- 473 genes showed differential expression between samples with mutant and wildtype EZH2 (FDR 0.05).
- The top three pathways enriched in these genes **(shownto the right)** are among the top 4 pathways enriched in GCB vs ABC differentially expressed genes

However:

- All but one EZH2 mutation is in GCB.
- Using just GCB samples only two genes were differentially expressed between EZH2 mutant and wildtype tumors (FDR 0.1), neither of which is well annotated
- Effects of EZH2 mutation are confounded with other characteristics of GCB tumors
- EZH2 may or may not actually contribute to the differential expression between GCB and ABC
- Need to compare expression in EZH2 mutant and wildtype cells with otherwise identical genetic backgrounds

Overlap of EZH2 and ABC vs GCB differentially expressed genes:



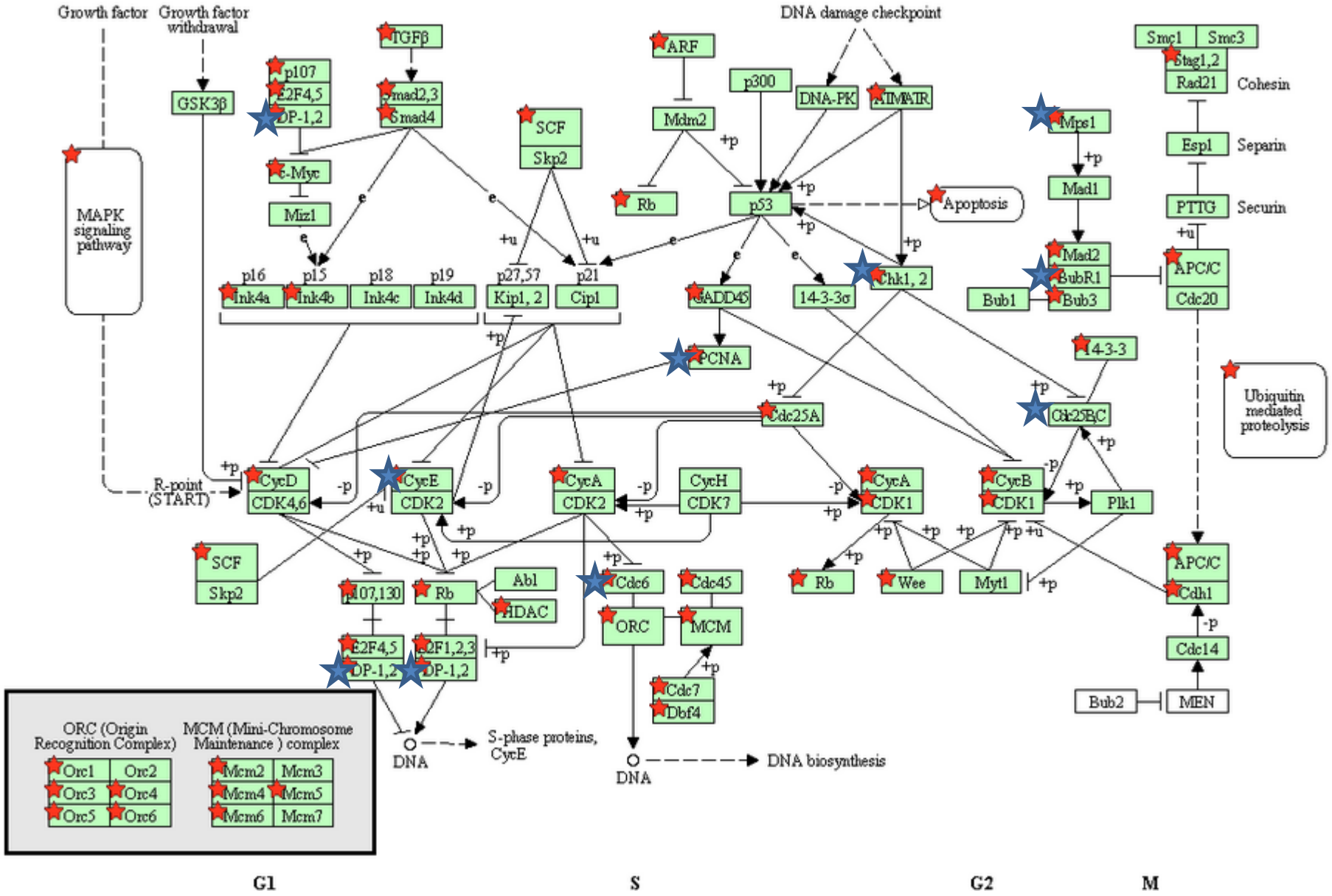
These 66 genes show no significant pathway enrichment and include only three genes clearly related to cell proliferation or apoptosis (IGF-1, PPIF, MST4)

Result III: No genes were significantly differentially expressed in MEF2B, MLL2, CREBBP, EP300 or HDAC7 mutant samples

- At a FDR of 0.05, MEF2B and EP300 had only 5 and 2 genes, respectively, differentially expressed. Other mutations investigated had none.
- MEF2B and EP300 are more frequently mutated in GCB; when only GCB samples were used, no genes were differentially expressed in MEF2B or EP300 mutants.
- Like EZH2, these genes are confounded with GCB subtype.

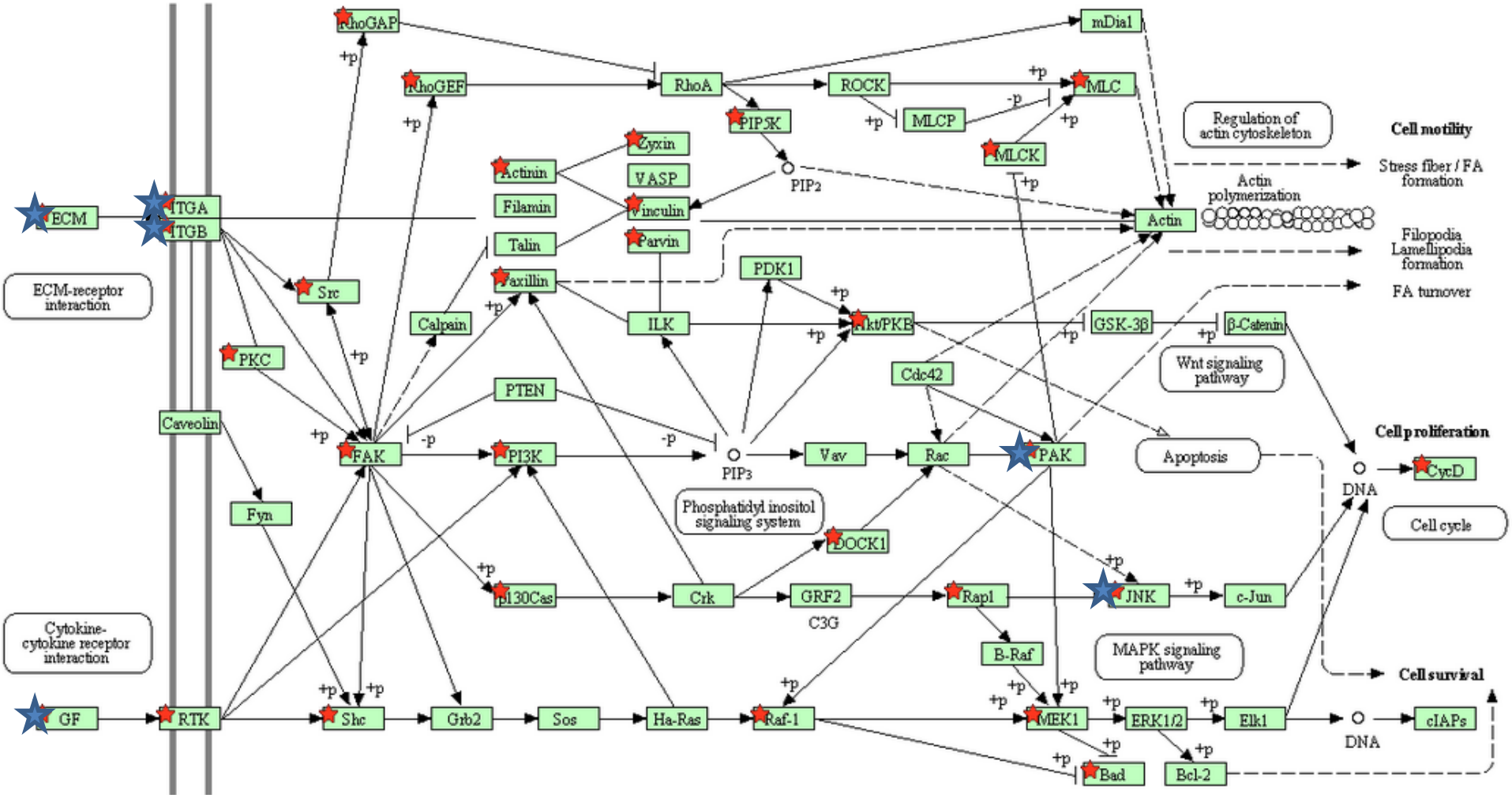
Differentially expressed in EZH2 mutants
Differentially expressed in GCB vs ABC

Cell Cycle Regulation Pathway



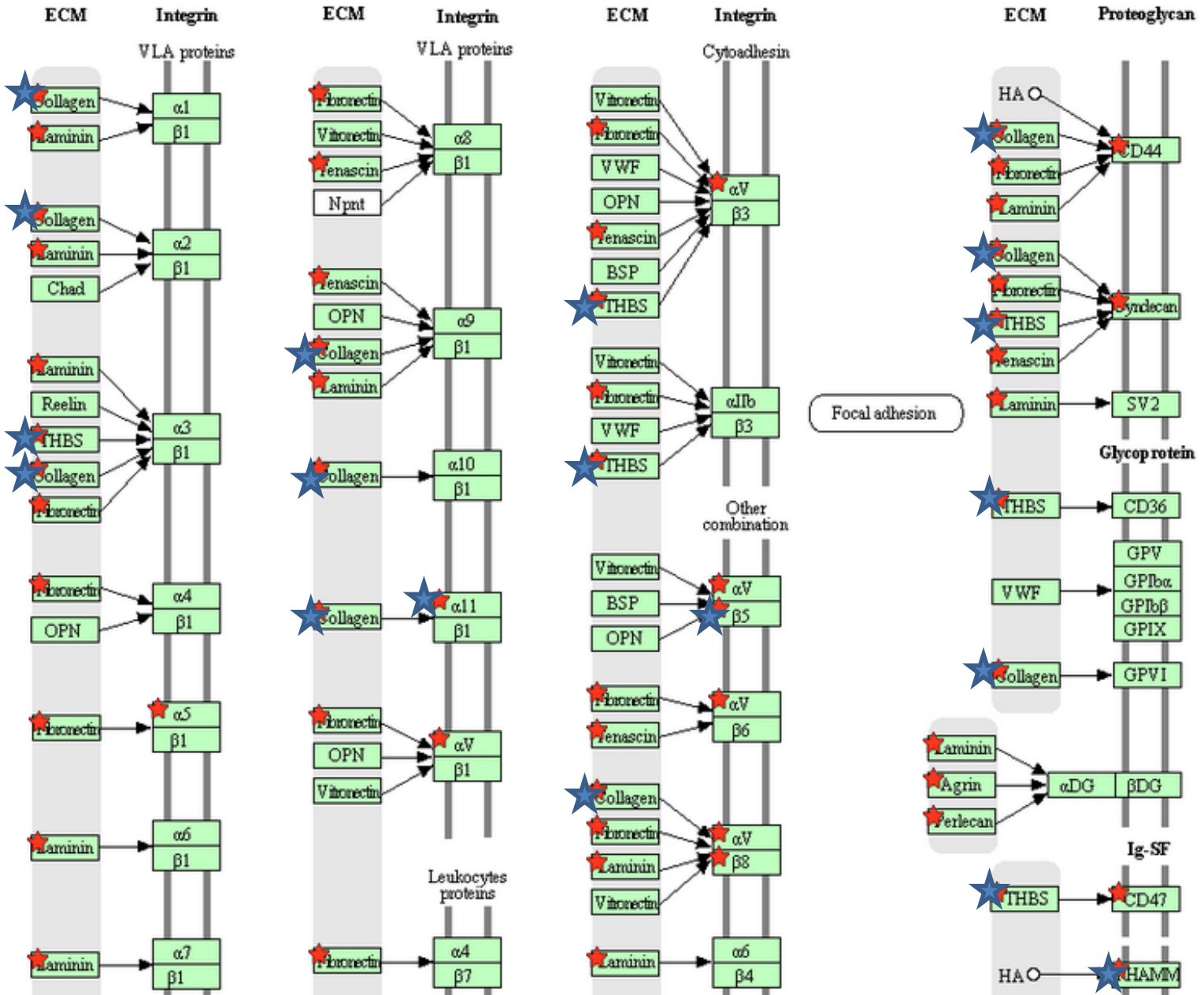
Differentially expressed in EZH2 mutants
Differentially expressed in GCB vs ABC

Focal Adhesion Pathway



Differentially expressed in EZH2 mutants
Differentially expressed in GCB vs ABC

ECM Receptor Interactions



Conclusions

- Clustering of GCB vs ABC based on differential gene expression was validated
- Although many genes were significantly differentially expressed genes in EZH2 mutant vs non mutant tumors, EZH2 status is strongly confounded with tumor subtype
- No genes were significantly differentially expressed in MEF2B, MLL2, CREBBP, EP300 or HDAC7 mutant samples

Future Directions

- NOISeq (a purely empirical analysis tool) or DESeq (based on Negative Binomial GLM's) may be more appropriate for assessing differential expression and can be applied to the raw count data, but are computationally expensive.
- Examining differential gene expression in samples genetically identical except for a single mutation (such as a DLBCL cell line transfected with a mutant gene), would remove confounding variables.
- Identification of DNA binding sites for histone modifying enzymes would help distinguish true differentially expressed genes from false positives

References:

1. Canadian Cancer Society's Steering Committee: Canadian Cancer Statistics 2010. Toronto: Canadian Cancer Society, 2010. April 2010, ISSN 0835-2976
2. Alizadeh AA, Eisen MB, Davis RE, et al.Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-11.
3. Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-B-cell lymphomas. N Engl J Med 2008; 359:2313- 23.
4. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011. doi: 10.1038/nature10351.
5. Godley LA, Le Beau MM. The histone code and treatments for acute myeloid leukemia. N Engl J Med. 2012; 366(10):960-1.
6. Robinson and Oshlack. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 2010: **11**:R25 doi:10.1186/gb-2010-11-3-r25
7. Smyth, G. K Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 2004: Volume **3**, Article 3.
8. Linton K, Howarth C, Wappett M et al. 2012. Microarray Gene Expression Analysis of Fixed Archival Tissue Permits Molecular Classification and Identification of Potential Therapeutic Targets in Diffuse Large B-Cell. J Mol Diagn. Lymphoma.
9. Suzuki R and Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006; 22(12):1540-2.

Thank you to Ryan Morin and Emilia Lim for assisting us in access the dataset, and to the authors of Morin et al. (2011) for their involvement in producing this dataset