

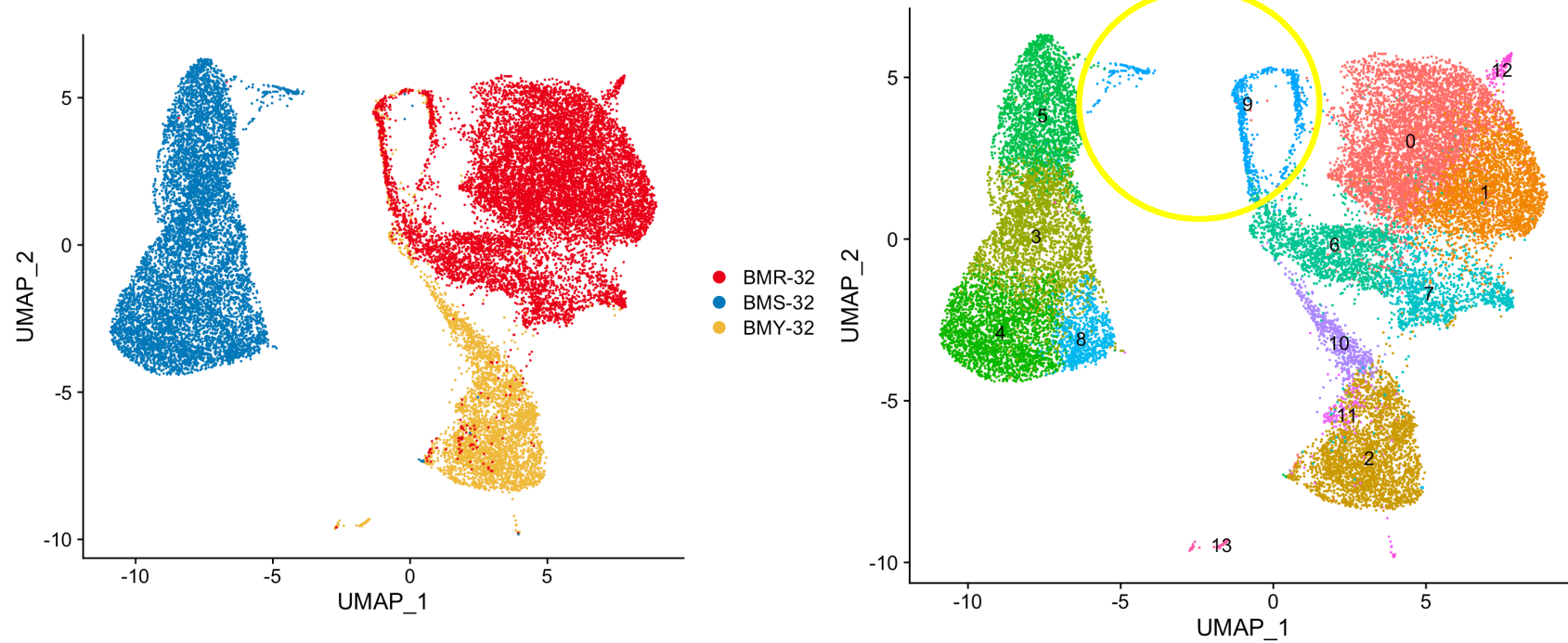
# Looking into the BMY-BMR overlapped cell cluster

Zinger Yang Loureiro

October 4, 2021

## Cluster 9 contains cells from all samples

Iteration 3, min.feature = 2000, resolution = 0.5 (14 clusters)

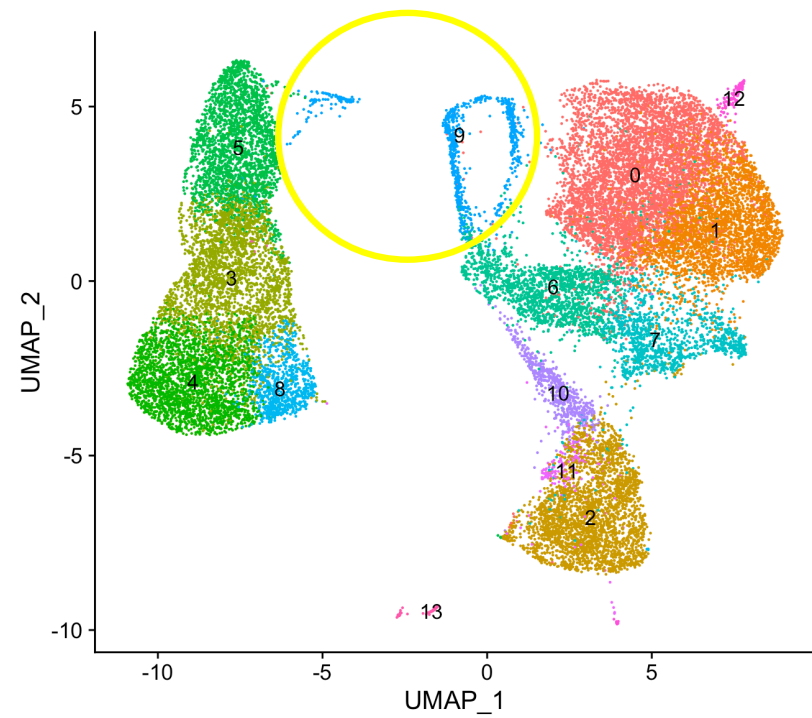
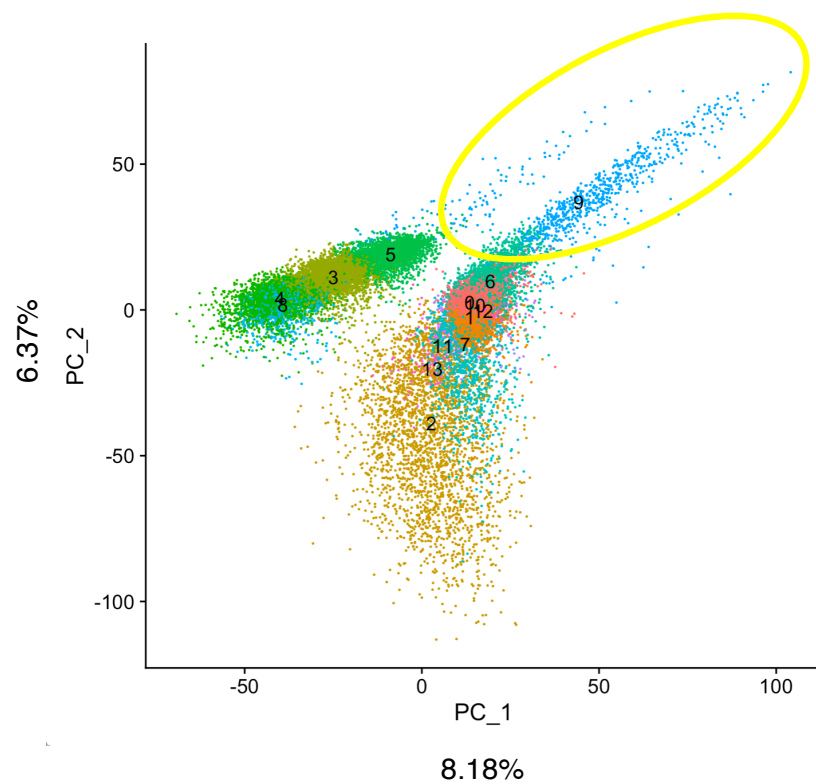


### Cluster 9 cells contribution to each sample

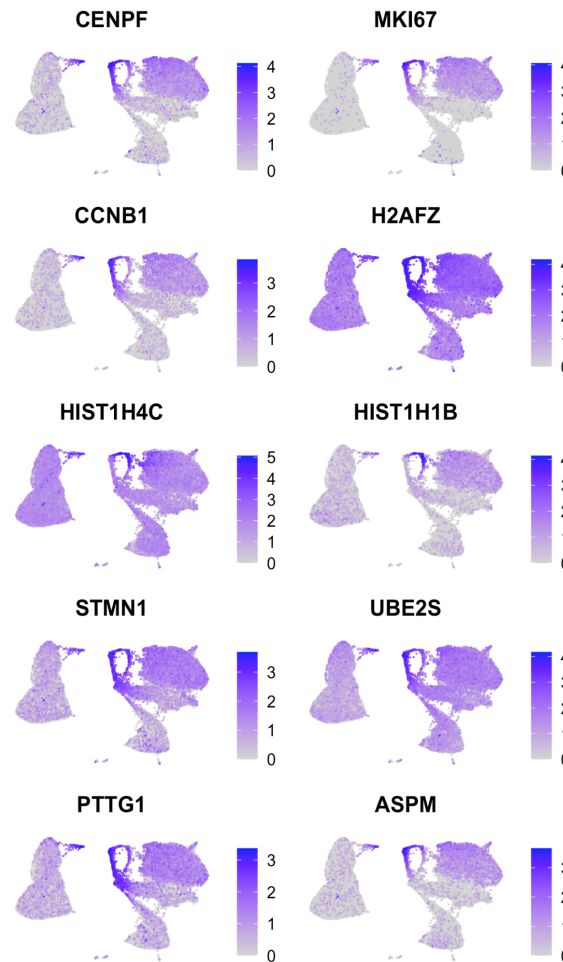
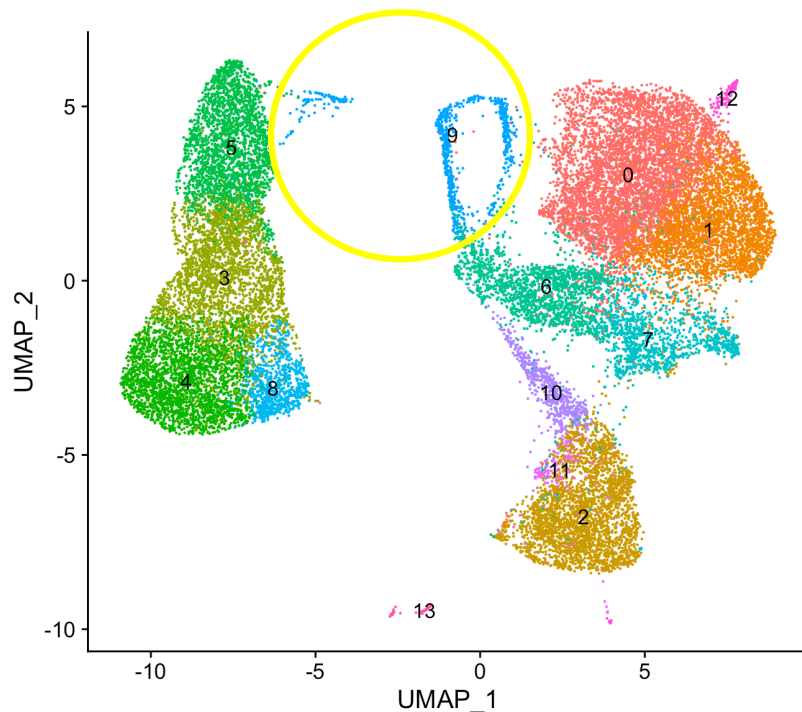
- BMY-32 148 of 3947 cells (3.75%)
- BMR-32 537 of 12037 cells (4.46%)
- BMS-32 119 of 8688 cells (1.37%)

## Cluster 9 in PCA vs UMAP projections

PCA projects cluster 9 as the outliers, primary contributor to PC1



## Top markers of the Cluster 9, “KI67-high” cluster are cell proliferation genes



Top 10 markers of cluster 9 by differential expression analysis:

- CENPF: Centromere Protein F
- MKI67: Proliferation Marker KI67
- CCNB1: Cyclin B1
- H2AFZ: Histone H2A.Z
- HIST1H4C: Histone H4
- HIST1H1B: Histone H1B
- STMN1: Stathmin 1, prevents assembly and promotes disassembly of microtubules
- UBE2S: Ubiquitin Conjugating Enzyme E2 S
- PTTG1: Securin, an anaphase-promoting complex substrate
- ASPM: Assembly Factor For Spindle Microtubules

DE Test: Wilcoxon Rank Sum test, Seurat package  
Features: limited to ones detected in >25% cells in the cluster

What is the identity of the KI67-high cluster?  
Are there shared features beyond cell cycle genes?

Cluster marker genes identified by DE: 219 genes

Possibilities:

1. The BMY, BMR cells cluster together due to shared strong expression of cell-proliferation genes
2. There are shared biology beyond cell-proliferation.

Approach:

1. Set enrichment (cell cycle genes strongly enriched)
2. Set enrichment of genes not involved in cell cycle

DE Test: Wilcoxon Rank Sum test, Seurat package  
Features: limited to ones detected in >25% cells in the cluster  
Threshold: log2 fold change > 1

## Set enrichment analysis returns cell cycle gene sets

Enrichment: GSEA MSigDB GO biological processes, set size < 1000, top 50 displayed

Gene Set Name CELL\_DIVISION CHROMOSOME\_SEGREGATION MITOTIC\_NUCLEAR\_DIVISION ORGANELLE\_FISSION  
SISTER\_CHROMATID\_SEGREGATION NUCLEAR\_CHROMOSOME\_SEGREGATION MITOTIC\_SISTER\_CHROMATID\_SEGREGATION  
DNA\_CONFORMATION\_CHANGE CELL\_CYCLE\_PHASE\_TRANSITION DNA\_METABOLIC\_PROCESS DNA\_PACKAGING  
PROTEIN\_DNA\_COMPLEX\_SUBUNIT\_ORGANIZATION MICROTUBULE\_CYTOSKELETON\_ORGANIZATION SPINDLE\_ORGANIZATION  
MICROTUBULE\_CYTOSKELETON\_ORGANIZATION\_INVOLVED\_IN\_MITOSIS REGULATION\_OF\_CELL\_CYCLE\_PROCESS  
REGULATION\_OF\_CHROMOSOME\_ORGANIZATION MICROTUBULE\_BASED\_PROCESS REGULATION\_OF\_CHROMOSOME\_SEGREGATION  
CHROMOSOME\_SEPARATION MITOTIC\_SPINDLE\_ORGANIZATION CENTROMERE\_COMPLEX\_ASSEMBLY REGULATION\_OF\_MITOTIC\_CELL\_CYCLE  
CHROMATIN\_ORGANIZATION REGULATION\_OF\_CHROMOSOME\_SEPARATION KINETOCHORE\_ORGANIZATION  
METAPHASE\_ANAPHASE\_TRANSITION\_OF\_CELL\_CYCLE CHROMATIN\_ASSEMBLY\_OR\_DISASSEMBLY NUCLEOSOME\_ASSEMBLY  
NUCLEOSOME\_ORGANIZATION REGULATION\_OF\_MITOTIC\_NUCLEAR\_DIVISION REGULATION\_OF\_MITOTIC\_SISTER\_CHROMATID\_SEGREGATION  
CHROMATIN\_REMODELING\_AT\_CENTROMERE REGULATION\_OF\_CELL\_CYCLE\_PHASE\_TRANSITION  
NEGATIVE\_REGULATION\_OF\_METAPHASE\_ANAPHASE\_TRANSITION\_OF\_CELL\_CYCLE REGULATION\_OF\_NUCLEAR\_DIVISION  
NEGATIVE\_REGULATION\_OF\_CHROMOSOME\_ORGANIZATION DNA\_RECOMBINATION HISTONE\_EXCHANGE CHROMOSOME\_CONDENSATION  
DNA\_REPLICATION MEIOTIC\_CELL\_CYCLE NEGATIVE\_REGULATION\_OF\_CELLULAR\_COMPONENT\_ORGANIZATION  
DNA\_REPLICATION\_INDEPENDENT\_NUCLEOSOME\_ORGANIZATION NEGATIVE\_REGULATION\_OF\_NUCLEAR\_DIVISION  
PROTEIN\_LOCALIZATION\_TO\_CHROMOSOME NEGATIVE\_REGULATION\_OF\_CELL\_CYCLE CELLULAR\_RESPONSE\_TO\_DNA\_DAMAGE\_STIMULUS  
MEIOTIC\_CELL\_CYCLE\_PROCESS POSITIVE\_REGULATION\_OF\_CELL\_CYCLE

## The 73 genes not contributing to the cell cycle gene sets are mostly genes involved in splicing

30 genes contributed to gene sets below. 43 genes not contributing to any gene sets.

Gene Set Name	# Genes in Overlap (k)	FDR q-value
GOBP_RNA_SPLICING_VIA_TRANSESTERIFICATION_REACTIONS	17	2.46E-15
GOBP_RNA_SPLICING	17	4.83E-14
GOBP_MRNA_METABOLIC_PROCESS	20	1.70E-13
GOBP_MRNA_PROCESSING	17	2.01E-13
GOBP_NUCLEAR_TRANSPORT	11	4.73E-08
GOBP_MRNA_EXPORT_FROM_NUCLEUS	7	1.94E-06
GOBP_RNA_EXPORT_FROM_NUCLEUS	7	8.20E-06
GOBP_MRNA_TRANSPORT	7	1.02E-05
GOBP_RNA_3_END_PROCESSING	7	1.02E-05
GOBP_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS	10	1.02E-05
GOBP_MRNA_3_END_PROCESSING	6	1.88E-05
GOBP_ESTABLISHMENT_OF_RNA_LOCALIZATION	7	4.83E-05
GOBP_NUCLEAR_EXPORT	7	4.83E-05
GOBP_RNA_LOCALIZATION	7	1.22E-04
GOBP_NUCLEOBASE_CONTAINING_COMPOUND_TRANSPORT	7	1.93E-04
GOBP_PROTEIN_CONTAINING_COMPLEX_LOCALIZATION	7	3.94E-04
GOBP_REGULATION_OF_MRNA_METABOLIC_PROCESS	7	1.11E-03
GOBP_RIBONUCLEOPROTEIN_COMPLEX_SUBUNIT_ORGANIZATION	6	1.60E-03
GOBP_REGULATION_OF_RNA_SPLICING	5	2.60E-03
GOBP_RNA_CATABOLIC_PROCESS	7	3.84E-03
GOBP_RESPONSE_TO_LEUKEMIA_INHIBITORY_FACTOR	4	4.95E-03
GOBP_REGULATION_OF_MRNA_SPLICING_VIA_SPLICEOSOME	4	1.14E-02
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS	5	1.26E-02
GOBP_PEPTIDE_BIOSYNTHETIC_PROCESS	8	1.79E-02
GOBP_MRNA_METABOLIC_PROCESS	5	2.20E-02
GOBP_NCRNA_PROCESSING	6	2.53E-02
GOBP_RESPONSE_TO_INTERLEUKIN_12	3	2.87E-02
GOBP_SPLICEOSOMAL_SNRNP_ASSEMBLY	3	3.11E-02
GOBP_REGULATION_OF_MRNA_PROCESSING	4	3.12E-02
GOBP_ORGANIC_CYCLIC_COMPOUND_CATABOLIC_PROCESS	7	3.87E-02
GOBP_AMIDE_BIOSYNTHETIC_PROCESS	8	4.46E-02

SRSF7	serine and arginine rich splicing factor 7
SRSF3	serine and arginine rich splicing factor 3
MRPL15	mitochondrial ribosomal protein L15
PFKP	phosphofructokinase, platelet

SNRPA1	small nuclear ribonucleoprotein polypeptide A'
LMNB1	lamin B1
HNRNPDL	heterogeneous nuclear ribonucleoprotein D like

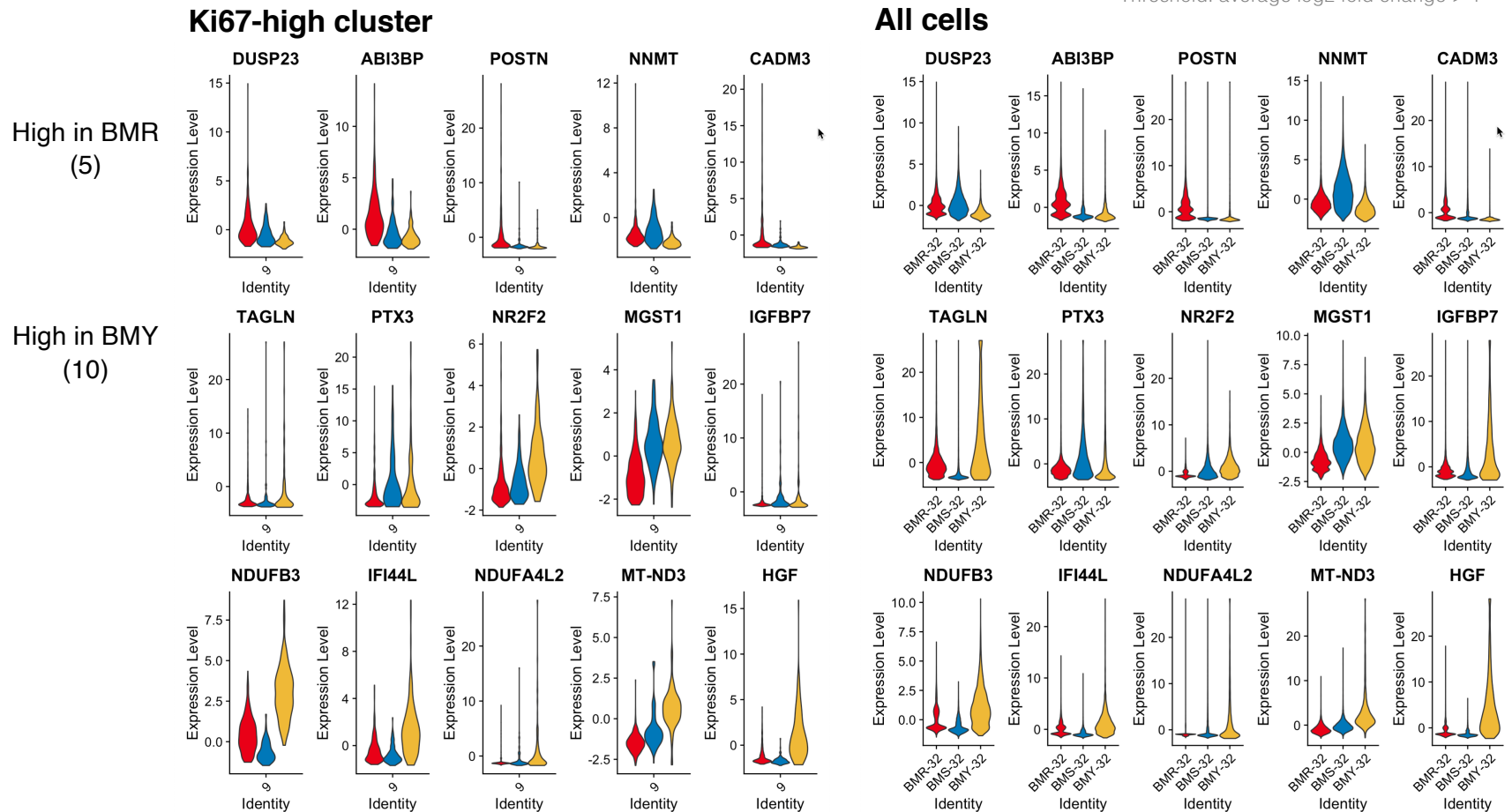
## The 43 genes not contributing to gene sets

ARHGAP11A	Rho GTPase activating protein 11A
ARL6IP1	ADP ribosylation factor like GTPase 6 interacting protein 1
BZW1	basic leucine zipper and W2 domains 1
C12orf75	chromosome 12 open reading frame 75
CKAP2L	cytoskeleton associated protein 2 like
COMMD4	COMM domain containing 4
CXCL3	C-X-C motif chemokine ligand 3
CYCS	cytochrome c, somatic
DEPDC1	DEP domain containing 1
DIAPH3	diaphanous related formin 3
DIRAS3	DIRAS family GTPase 3
DNAJC9	DnaJ heat shock protein family (Hsp40) member C9
DTYMK	deoxythymidylate kinase
EMP2	epithelial membrane protein 2
FABP5	fatty acid binding protein 5
GNPNAT1	glucosamine-phosphate N-acetyltransferase 1
HNRNPAB	heterogeneous nuclear ribonucleoprotein A/B
HSPE1	heat shock protein family E (Hsp10) member 1
HSPH1	heat shock protein family H (Hsp110) member 1
ILF2	interleukin enhancer binding factor 2
JPT1	Jupiter microtubule associated homolog 1
KCNMA1	potassium calcium-activated channel subfamily M alpha 1
MT2A	metallothionein 2A
ODC1	ornithine decarboxylase 1
PAICS	phosphoribosylaminoimidazole carboxylase and phosphoribosylaminoimidazolesuccinocarboxamide synthase
PBK	PDZ binding kinase
PGP	phosphoglycolate phosphatase
PHGDH	phosphoglycerate dehydrogenase
PHLDA2	pleckstrin homology like domain family A member 2
PPP1R14B	protein phosphatase 1 regulatory inhibitor subunit 14B
PTMA	prothymosin alpha
SAP30	Sin3A associated protein 30
SHCBP1	SHC binding and spindle associated 1
SMS	spermine synthase
STEAP1	STEAP family member 1
STEAP1B	STEAP family member 1B
STIP1	stress induced phosphoprotein 1
TFAM	transcription factor A, mitochondrial
TFPI2	tissue factor pathway inhibitor 2
TIMM10	translocase of inner mitochondrial membrane 10
TMPO	thymopoietin
TNFRSF12A	TNF receptor superfamily member 12A
TOMM40	translocase of outer mitochondrial membrane 40



# DE genes between BMR and BMY in KI67-high cluster are differentially expressed in the overall samples

DE Test: Wilcoxon Rank Sum test, Seurat package  
Features: limited to ones detected in >25% cells in the cluster  
Threshold: average log2 fold change > 1



## Summary

- Besides cell cycle and splicing, I did not observe obvious shared biology across BMR and BMY in the KI67-high cluster.
- BMY cells in KI67-high cluster retains features consistent with other BMY cells. Vice versa for BMR cells.