

Towards detection of differential RNA editing events in transcriptomics datasets

Andrew Matveenko

Supervised by:

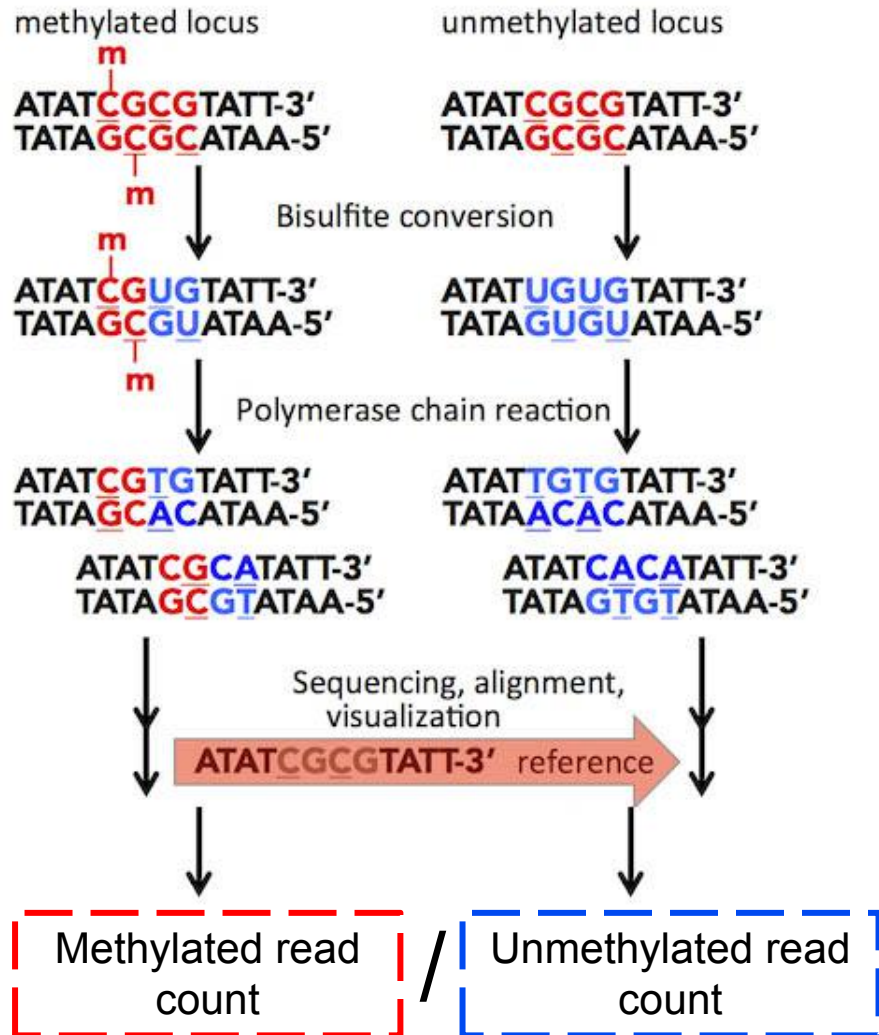
Anastasia Samsonova

Alexander Kanapin

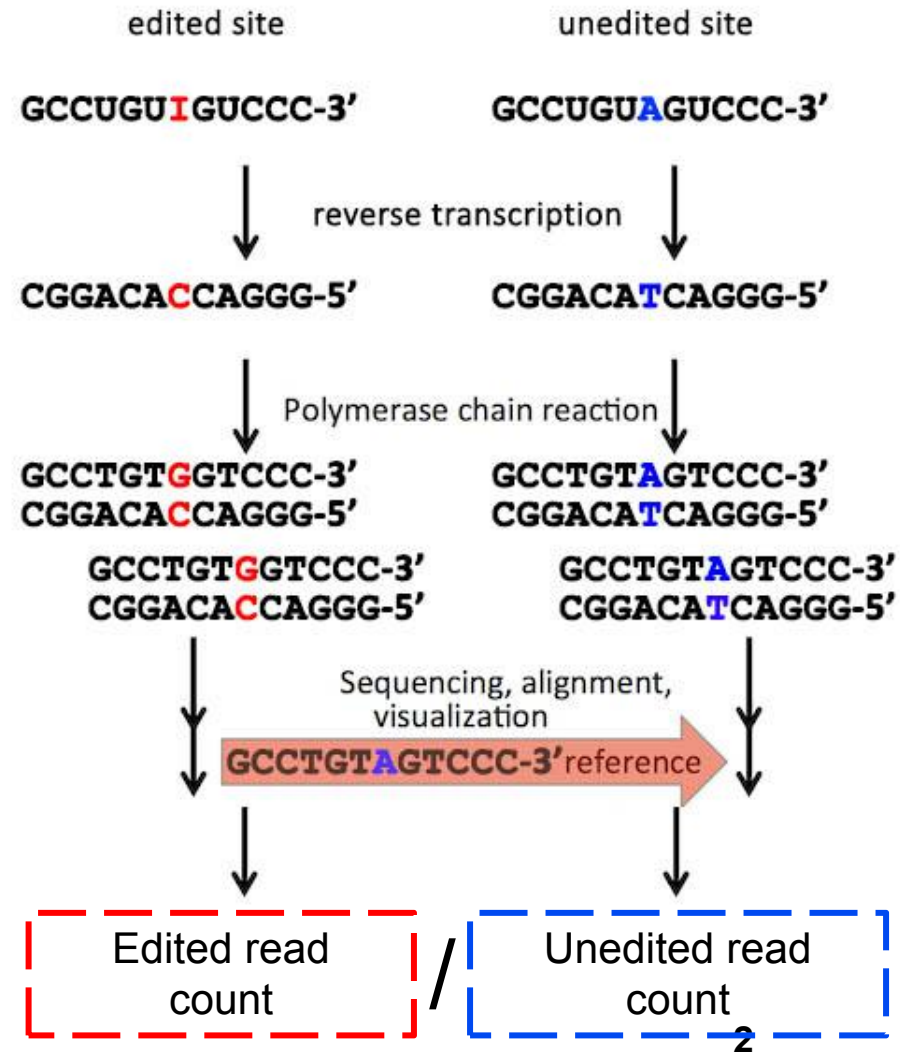
2019

RNA editing and Bisulfite sequencing

Bisulfite sequencing



RNA editing



RNA editing data is similar to Bisulfite sequencing (BSSeq)

	BSSeq	RNA editing
Event	C to 5mC conversion	A to I editing
Detection	C to T substitution	A to G substitution
Sequencing	Whole-genome	Total RNA
Regions	CpG	Various transcripts



RNA editing data is similar to Bisulfite sequencing (BSSeq)



Goals of the project:

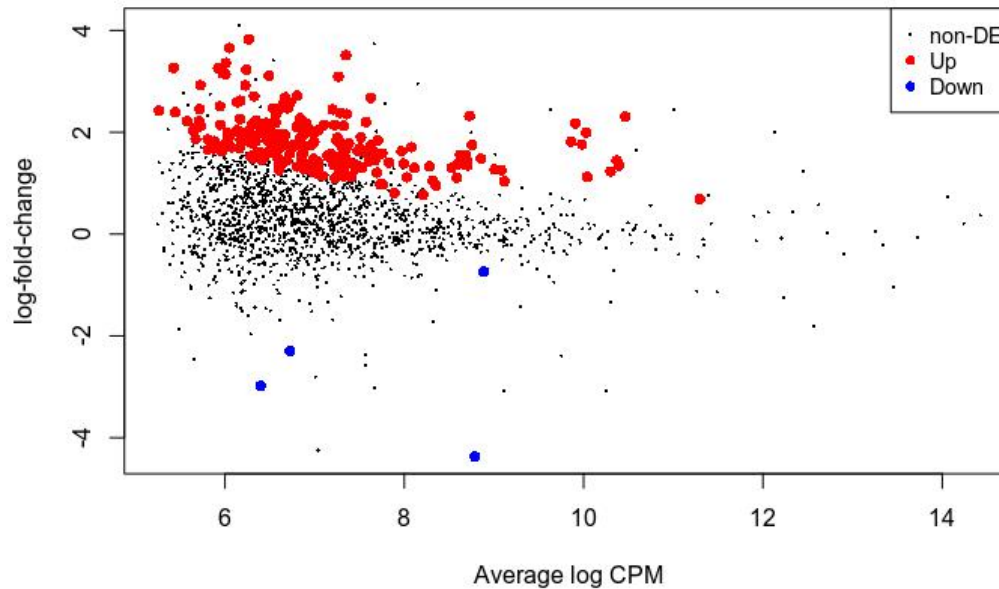
- 1. Literature review of existing methods for BSSeq differential methylation analysis
- 2. Experimental evaluation of existing BSSeq approaches and algorithms on RNA-seq data
- 3. Propose a prototype of a statistical framework and an algorithm for estimation of statistically significant changes in editing efficiency on a single nucleotides and/or gene level

DM analysis tools that can be used for DEd analysis

tool	pub date	stat method
edgeR (BSSeq)	2016	Logistic regression
methylKit	2011	Logistic regression
eDMR	2013	Logistic regression
BSmooth(bsseq)	2012	Smoothing
BiSeq	2013	Smoothing
DSS	2012	Beta-binomial
methylSig	2014	Beta-binomial
DSS-single	2015	Beta-binomial
DSS-general	2015	Beta-binomial
HMM-Fisher	2014	HMM
HMM-DM	2014	HMM

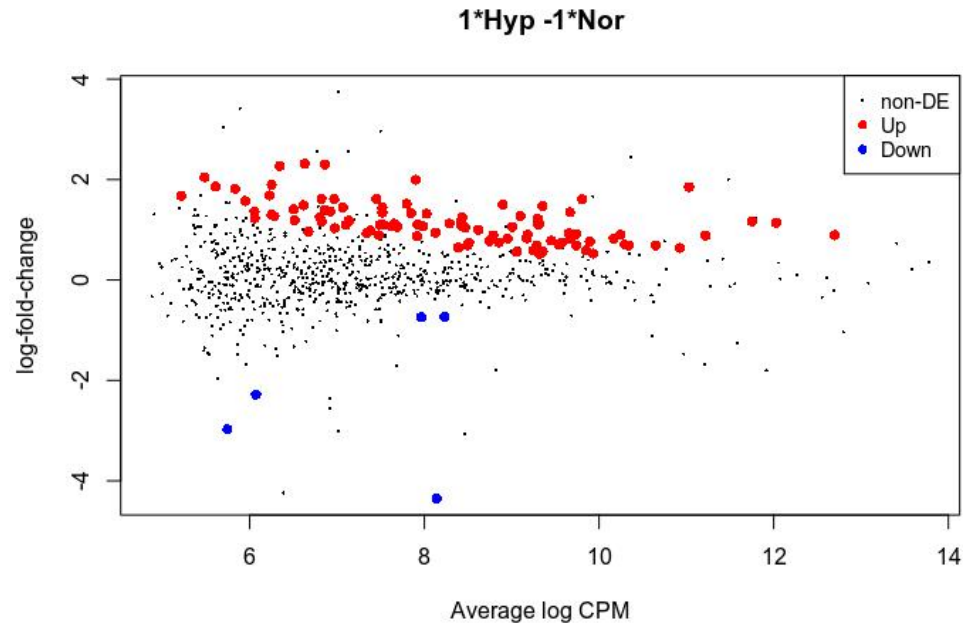
edgeR DM pipeline: single site DEd analysis

1*Hyp -1*Nor



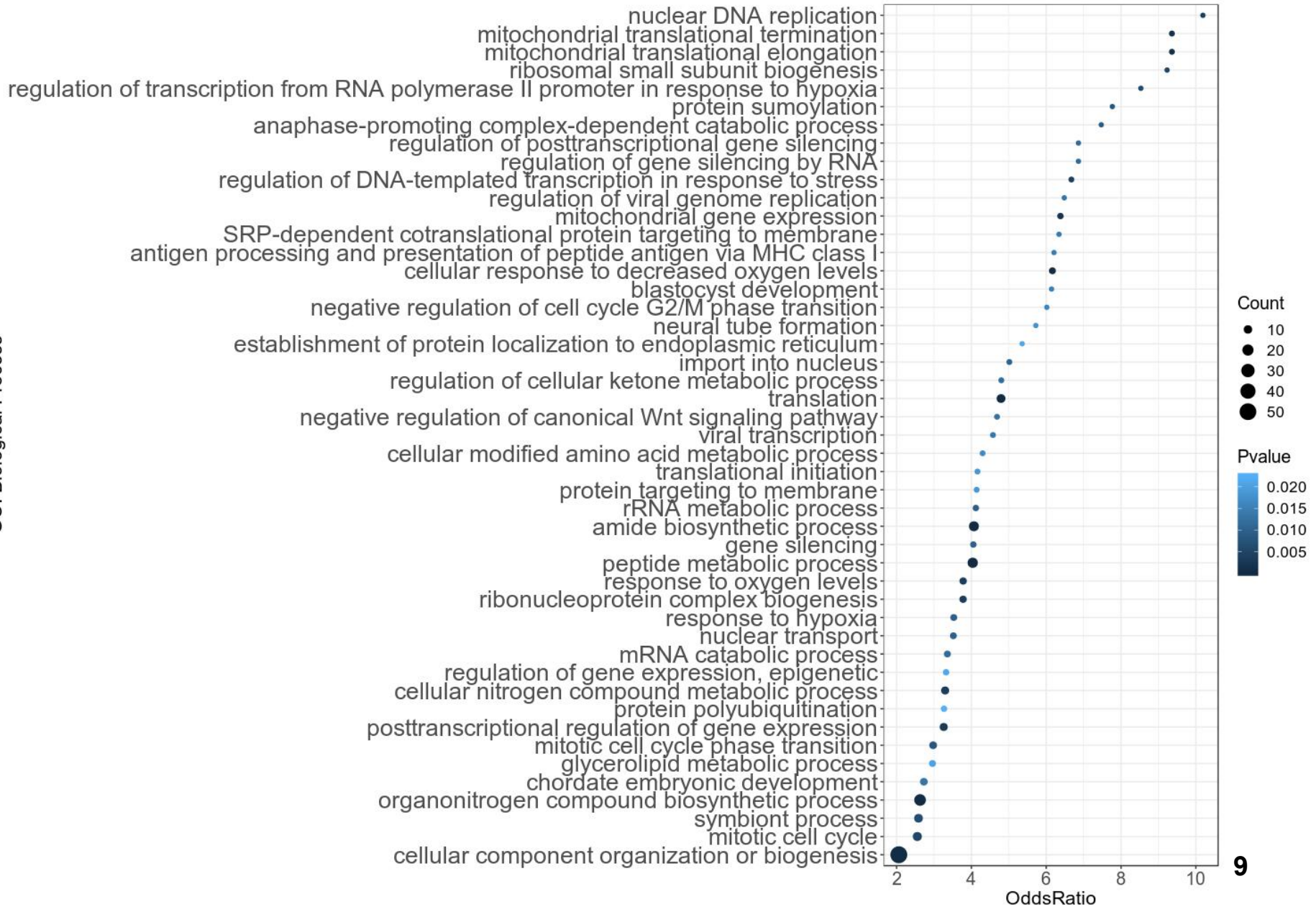
site	gene	logFC	logCPM	PValue	FDR
chr8-47977549	MCM4	2,308883	10,463645	1,16713E-52	1,97245E-49
chr8-47977073	MCM4	2,175716	9,904611	4,30624E-17	3,63877E-14
chr11-65777096	AP5B1	3,09172	7,266155	1,31694E-14	7,41874E-12
chr16-22285539	EEF2K	2,673666	7,626522	3,45825E-14	1,46111E-11
chr8-47977687	MCM4	1,816463	9,862521	1,18281E-13	3,99791E-11
chr4-56460147	PAICS	1,99636	10,030811	4,58359E-12	1,29105E-09
chr12-50930530	METTL7A	2,321698	8,726173	3,21048E-11	7,75101E-09
chr4-56460713	PAICS	1,346137	10,39574	4,26424E-11	9,00822E-09
chr4-56459960	PAICS	1,124809	10,040059	1,04243E-10	1,95746E-08

edgeR DM pipeline: analysis of DEd genes

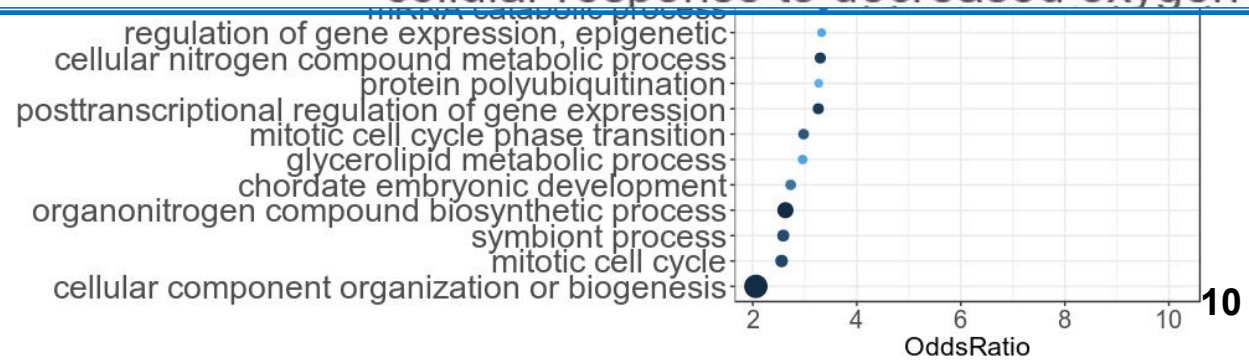
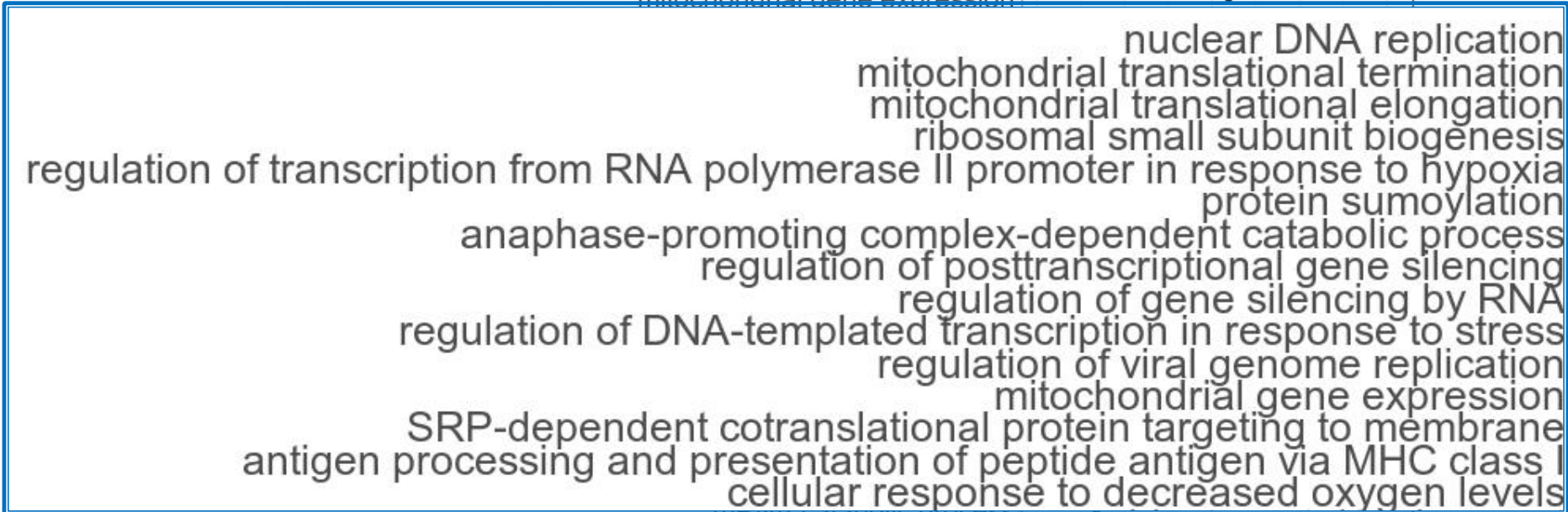
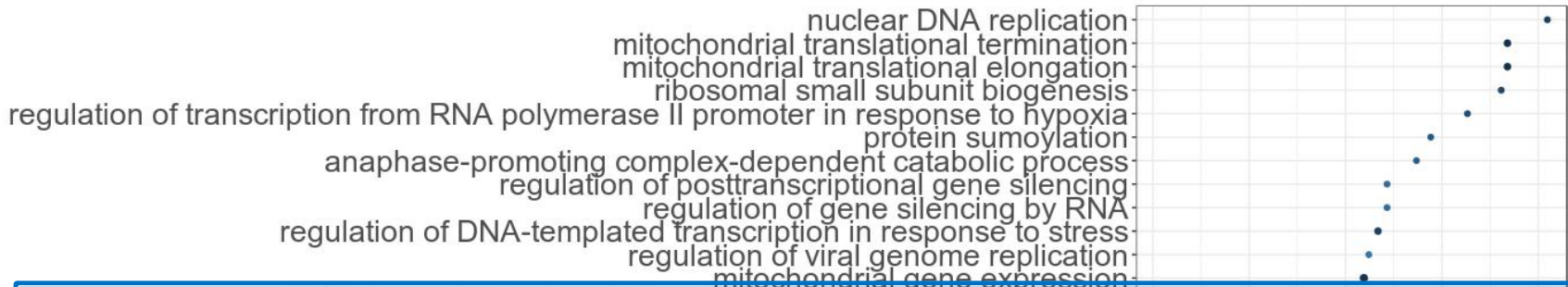


genes	logFC	logCPM	LR	PValue	FDR
MCM4	1,85076316	11,03018079	447,1932322	2,94408E-99	2,42887E-96
METTL7A	1,141094043	12,02728118	278,1665284	1,88428E-62	7,77266E-60
LIMD1	1,606858484	9,806994857	112,689622	2,5234E-26	6,93934E-24
PLEKHA2	1,347415666	9,666191094	73,1284901	1,21478E-17	2,50549E-15
GPATCH11	1,502883739	8,897046972	68,11080736	1,54561E-16	2,55026E-14
RPS19	1,117727572	9,315822032	57,43543936	3,49257E-14	4,80229E-12
LINC00963	1,473146625	9,35190027	54,98142462	1,21674E-13	1,43402E-11
BLCAP	1,995362948	7,90536899	48,95833445	2,61458E-12	2,69628E-10
MRPS16	1,272622246	9,101524593	47,77256877	4,78642E-12	4,38756E-10

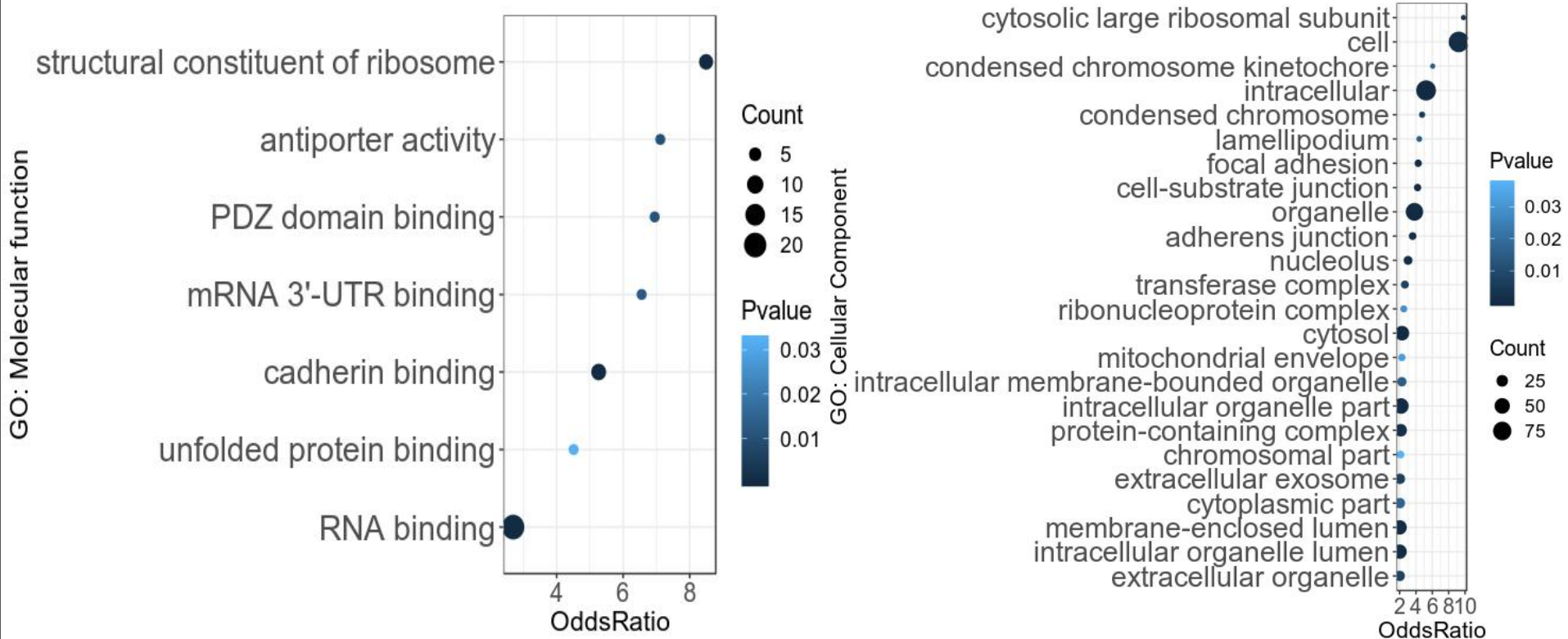
GO annotation of DEd genes



GO annotation of DEd genes



GO annotation of DEd genes



DM analysis tools that can be used for DEd analysis

tool	DM single-site	DM region analysis
edgeR (BSSeq)	+	+
methylKit	+	+
eDMR	-	+
DSS	+	+
methylSig	+	+



Conclusions

- At least one tool for differential methylation analysis (edgeR DM pipeline) can be used for differential RNA editing analysis
- RNA editing is enhanced during hypoxia in genes acting in ribosome biogenesis, mitochondrial translation, and transcriptional regulation associated with hypoxia