performance and reproducibility

EPICENTRE SCRIPTSEQ

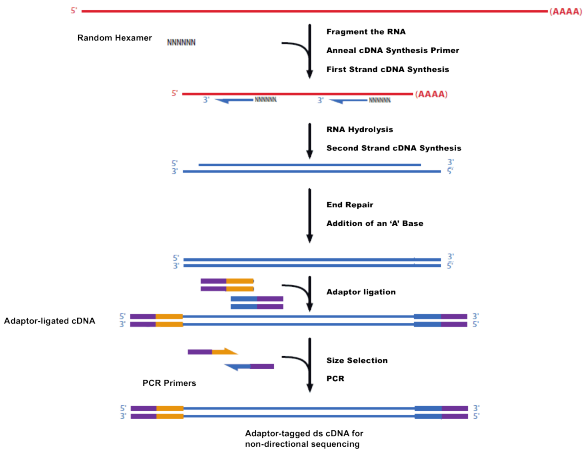
(incidental) design

- sample ‘keratinocyte’
- prepared twice using epicentre’s scriptseq, once using illumina’s mrna-seq (rev D)
- spiked synthetic RNA (ERCCs) into each aliquot
- polyA selection
- 15 million raw reads per prep
- hierarchical alignment with bwa to contam, genome (hg18), and junction files
- processed with in-house software

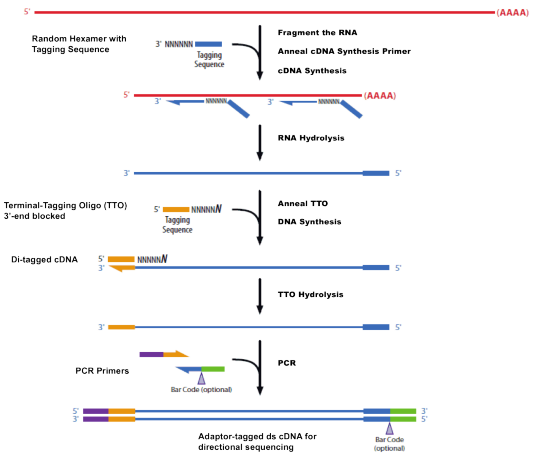
pragmatics

	epicentre	illumuna
mrna input	50 ng	100 ng
size selection	✖	✔
column cleanups	2	5
ethanol precipitation	✖	✔
strand specific	✔	✖
pcr cycles	10	15
barcodes	✔	✖
time	4 hours	2 days
cost	$295	$325

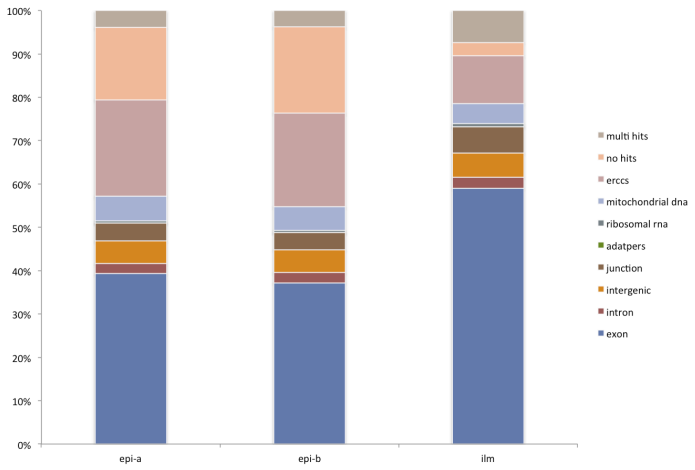
ilm



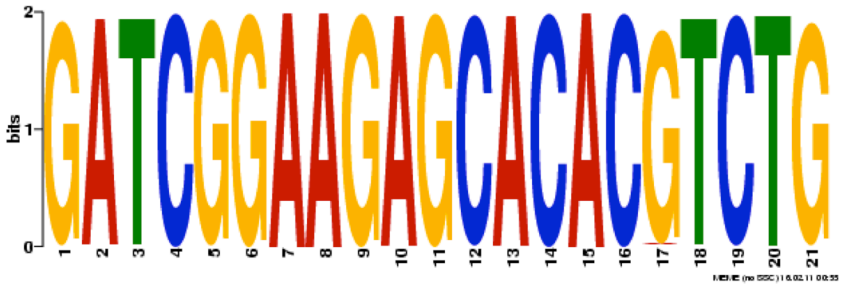
epi



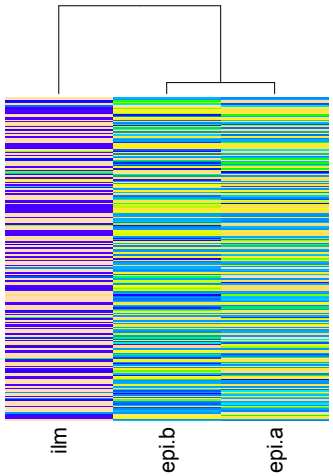
mapping statistics



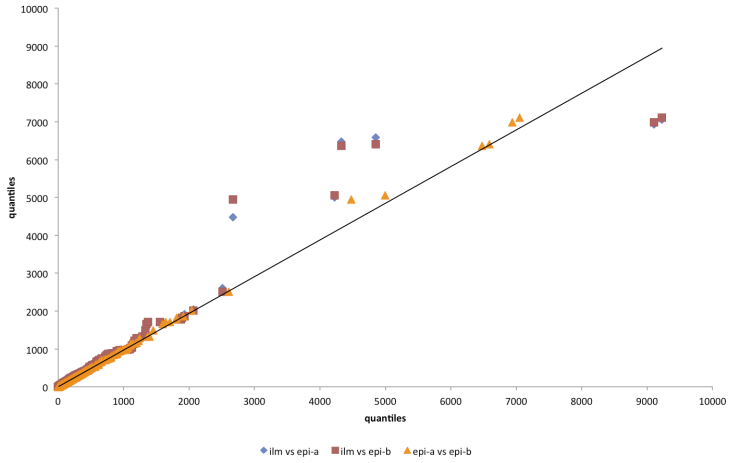
synthetic motif found in > 40% epi no-hits



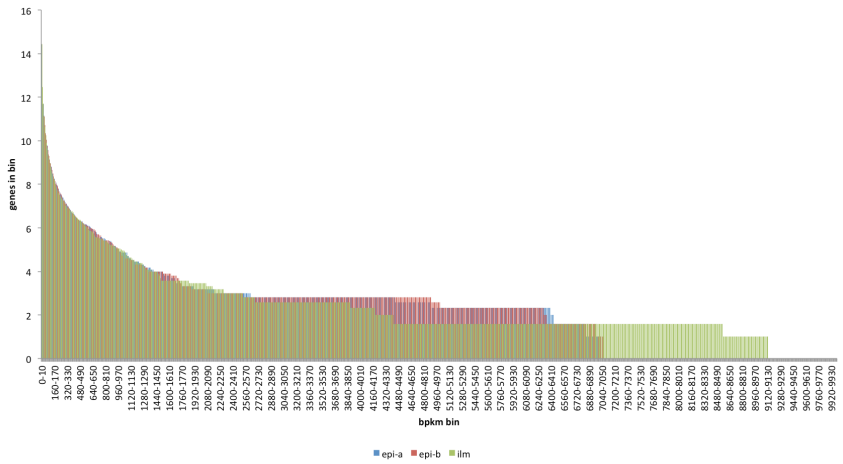
does the data cluster by prep?



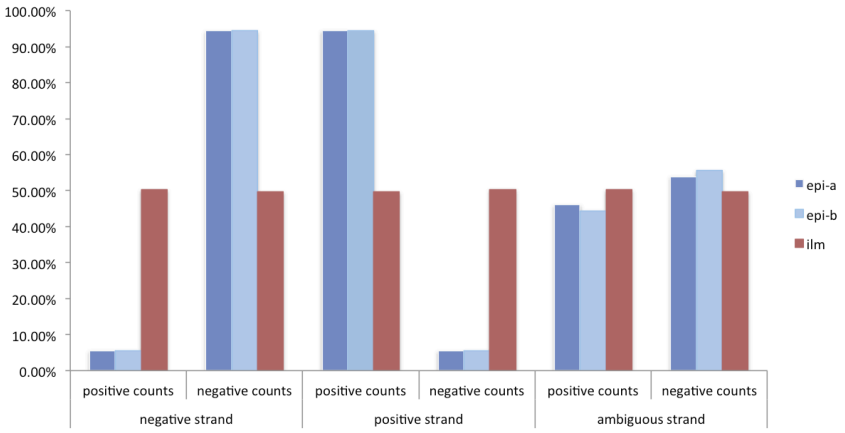
qq-plot



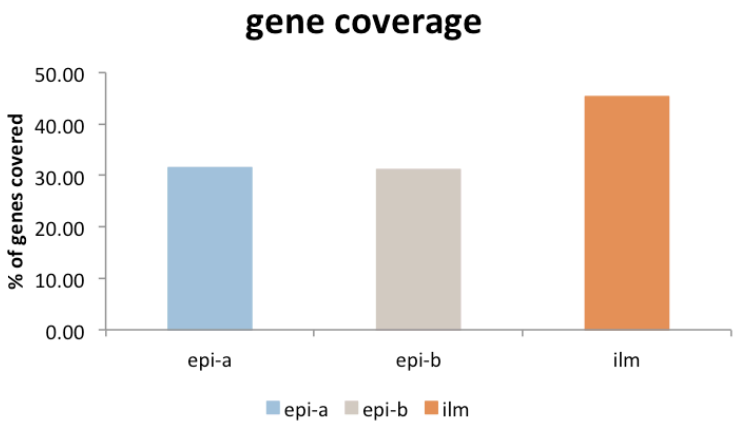
bpkm histogram



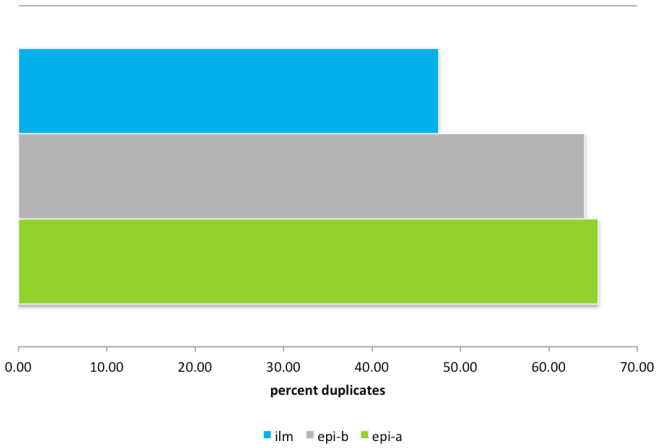
is epi stranded?



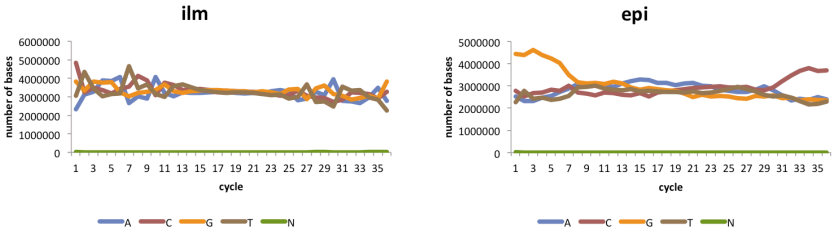
% of all genes covered



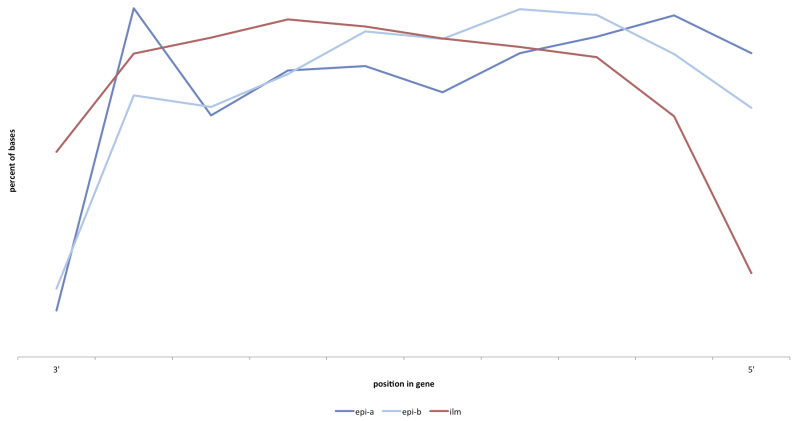
duplicates?



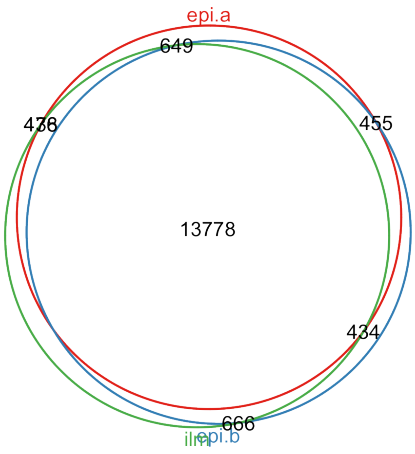
base composition biases

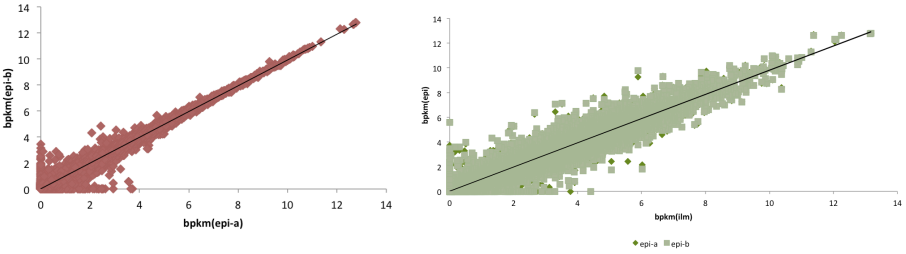


coverage across gene

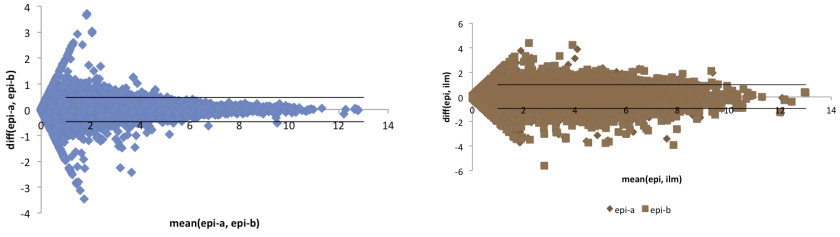


genes detected

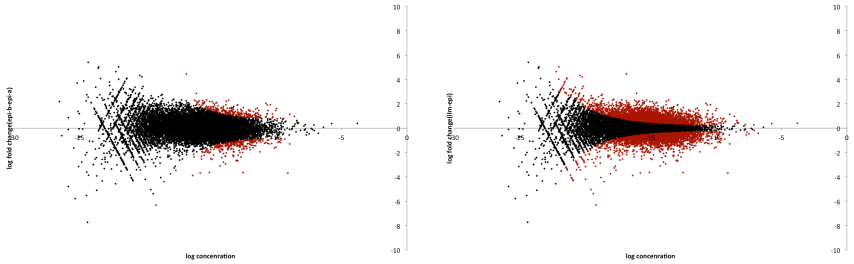




bland-altman tribute



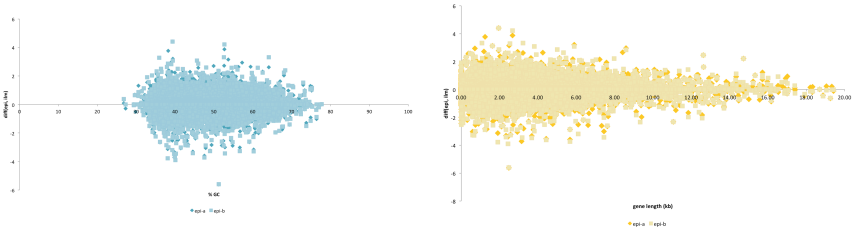
FC plots using common dispersion



can the difference be explained

by GC content?

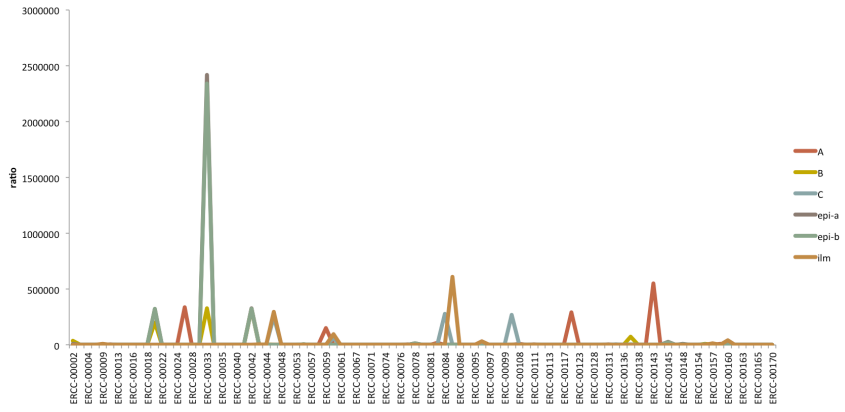
by gene length?



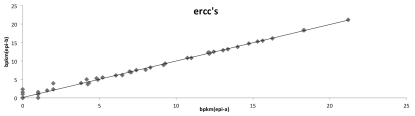
ercc design

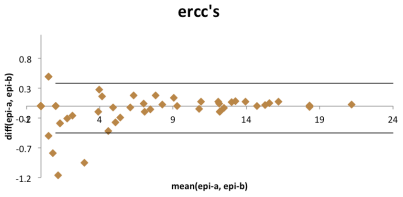
	subpool a	subpool b	subpool c
number errc's	32	32	32
ercc overlap	0	0	0
average length (bp)	881	915	922
average gc (%)	43.09	45.22	44.82
average abundance	46368	32812	43477

which subpool is in which sample?



ercc reproducibility





expected vs observed