BrainScan miRNA pipeline

RK

mirTools

- Oleg's choice?
- available only via hosted web-only interface
- contacted authors w.r.t. getting source code after call on 07/22 but no reply
- explored alternatives...

desired workflow

1. (f) read filtering, counting, & QC
2. (db) map to database such as miRBase
3. (map) map to genome to find novel hits and mRNA contamination
4. (DE) differential expression analysis ?

miRNA software

name	year	kind	workflow	notes
mirTools	2010	web-only	f ➜ db ➜ map	uses SOAP & miRBase
miRNAkey	2010	java GUI & perl	f ➜ db ➜ DE	uses miRBase & SEQ-EM keeps multi-map reads
miRanalyzer	2009	java cmd	db ➜ map	uses Bowtie, weka, miRBase, & Vienna (secondary-struc)
miRExpress	2009	C++ cmd	f ➜ db	uses miRBase
miRDeep	2008	perl cmd	f ➜ map	uses BLAST & Vienna discards multi-map reads: 5+

and several others incl. DSAP (web-only) and SeqBuster (Java GUI) ...

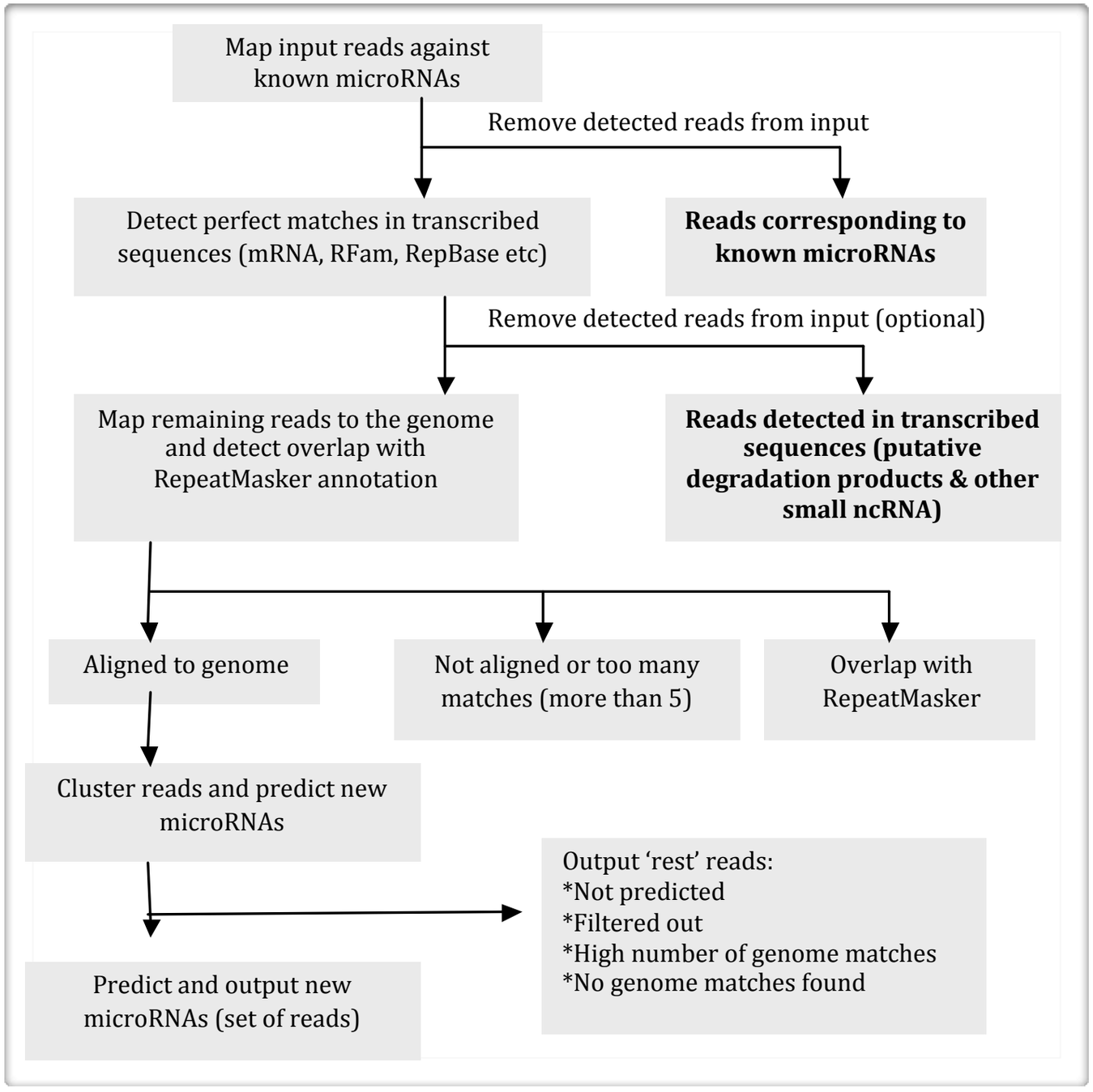
miRNA software

name	year	kind	workflow	notes
mirTools	2010	web-only	f ➜ db ➜ map	uses SOAP & miRBase
miRNAkey	2010	java GUI & perl	f ➜ db ➜ DE	uses miRBase & SEQ-EM keeps multi-map reads
miRanalyzer	2009	java cmd	db ➜ map	uses Bowtie, weka, miRBase, & Vienna (secondary-struc)
miRExpress	2009	C++ cmd	f ➜ db	uses miRBase
miRDeep	2008	perl cmd	f ➜ map	uses BLAST & Vienna discards multi-map reads: 5+

and several others incl. DSAP (web-only) and SeqBuster (Java GUI) ...

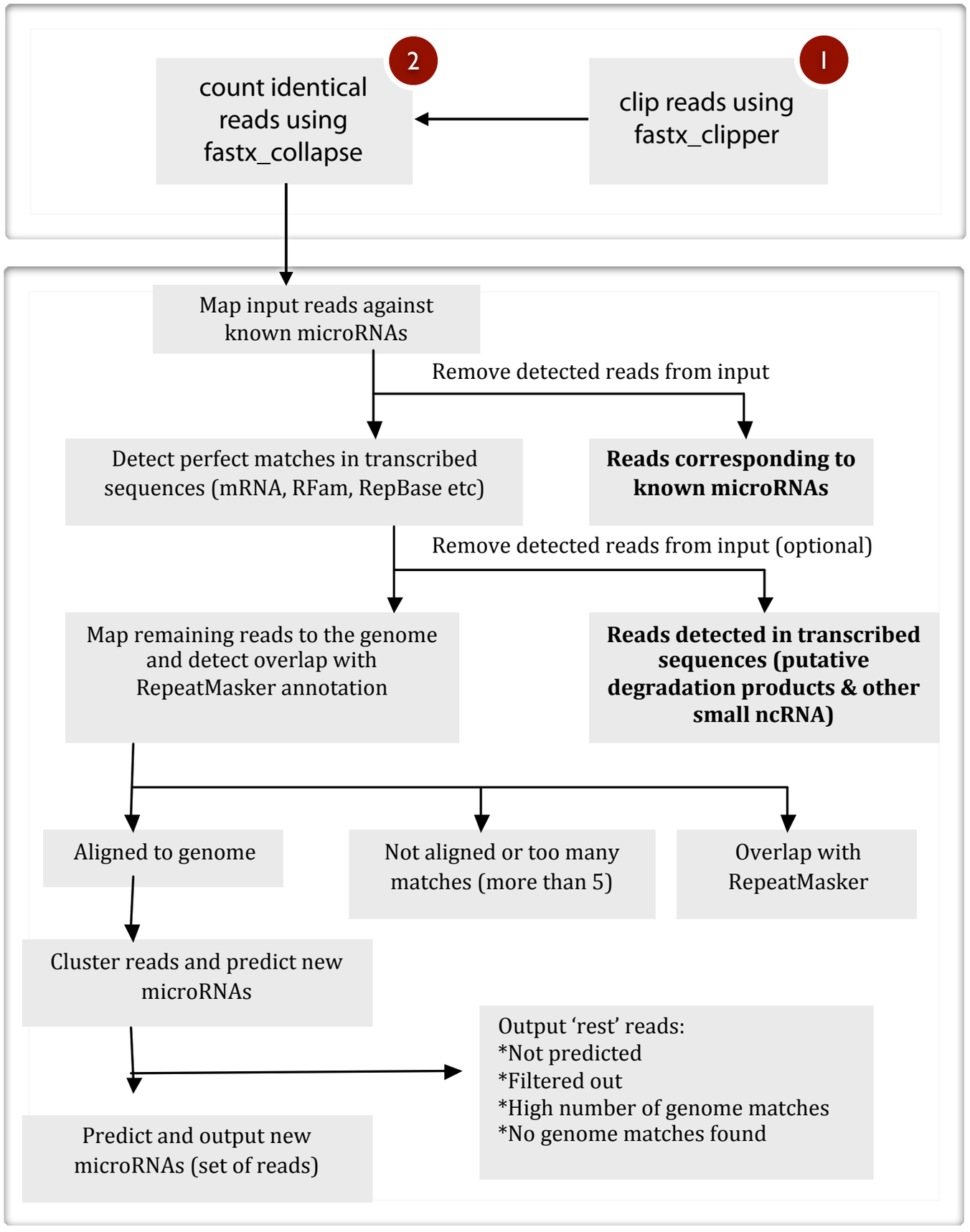
miRanalyzer

- analysis workflow:
- Hackenberg *et al.* 2009 NAR



fastx

miRanalyzer



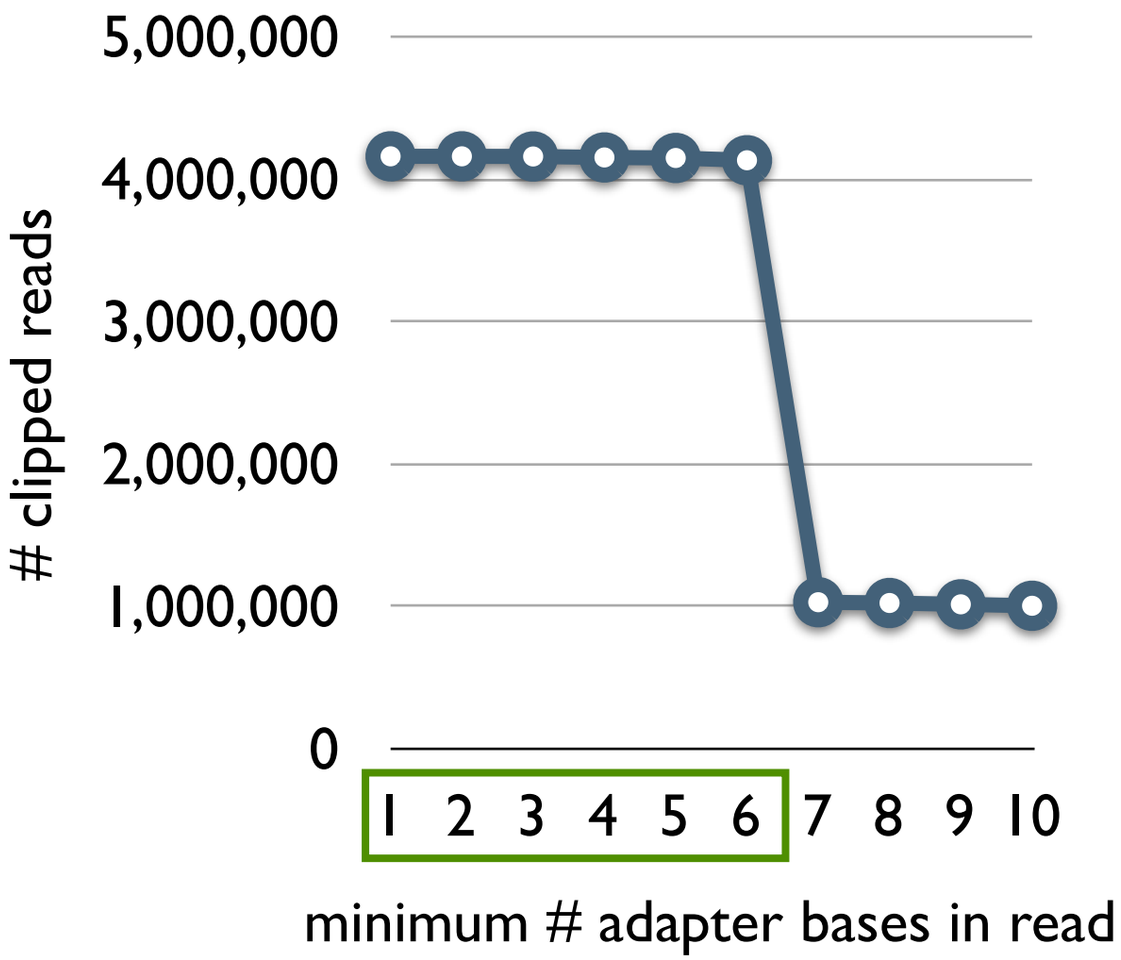
fastx_clipper

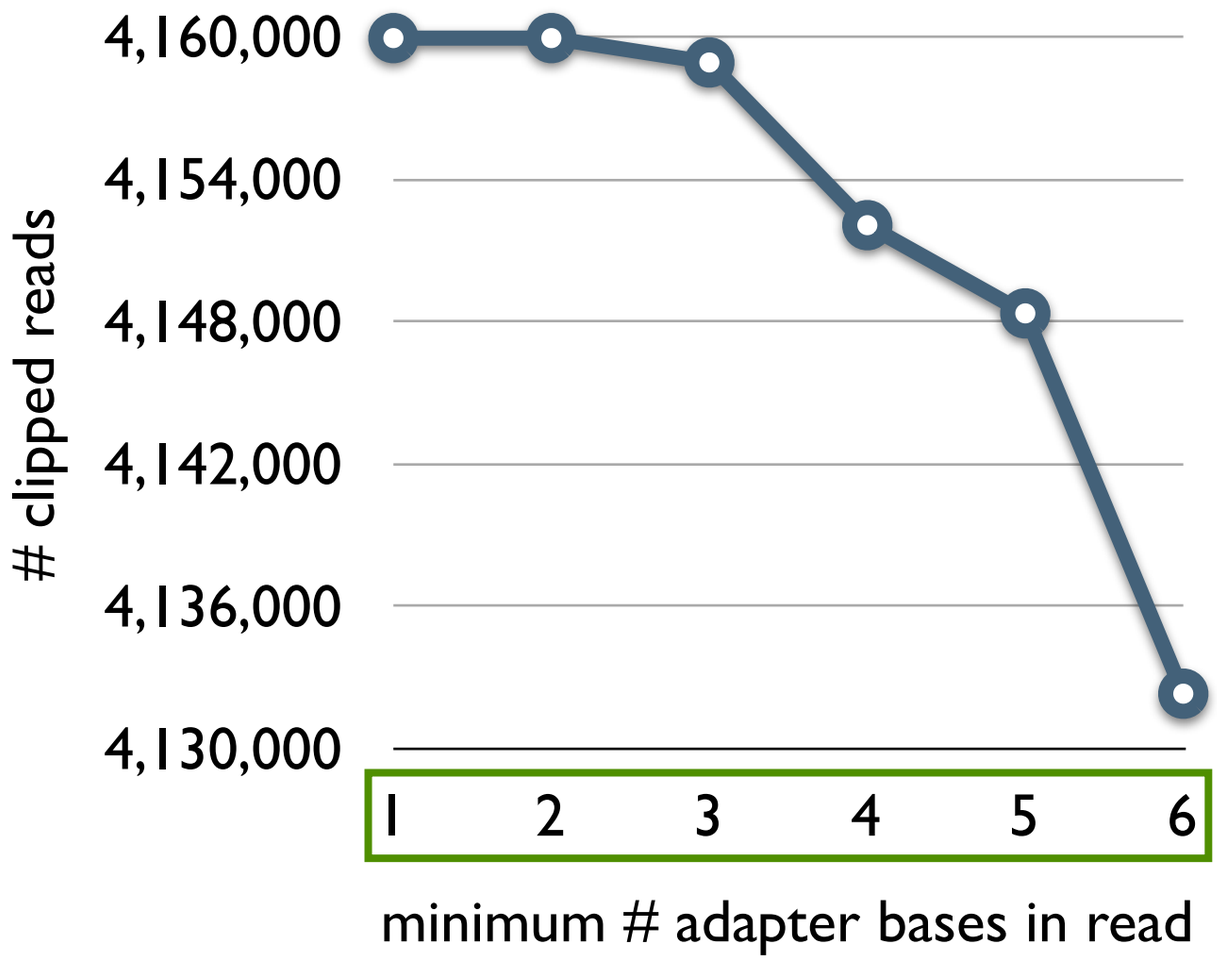
- remove 3’ adapter sequences
- miRNAkey default adapter sequence is:
  ATCTCGTATGCCGTCTTCTGCTTGC
- matches Illumina 3' smallRNA Adapter:
  TCGTATGCCGTCTTCTGCTTG

obtained from:

http://intron.ccam.uchc.edu/groups/tgcore/wiki/013c0/Solexa_Library_Primer_Sequences.html

fastx_clipper



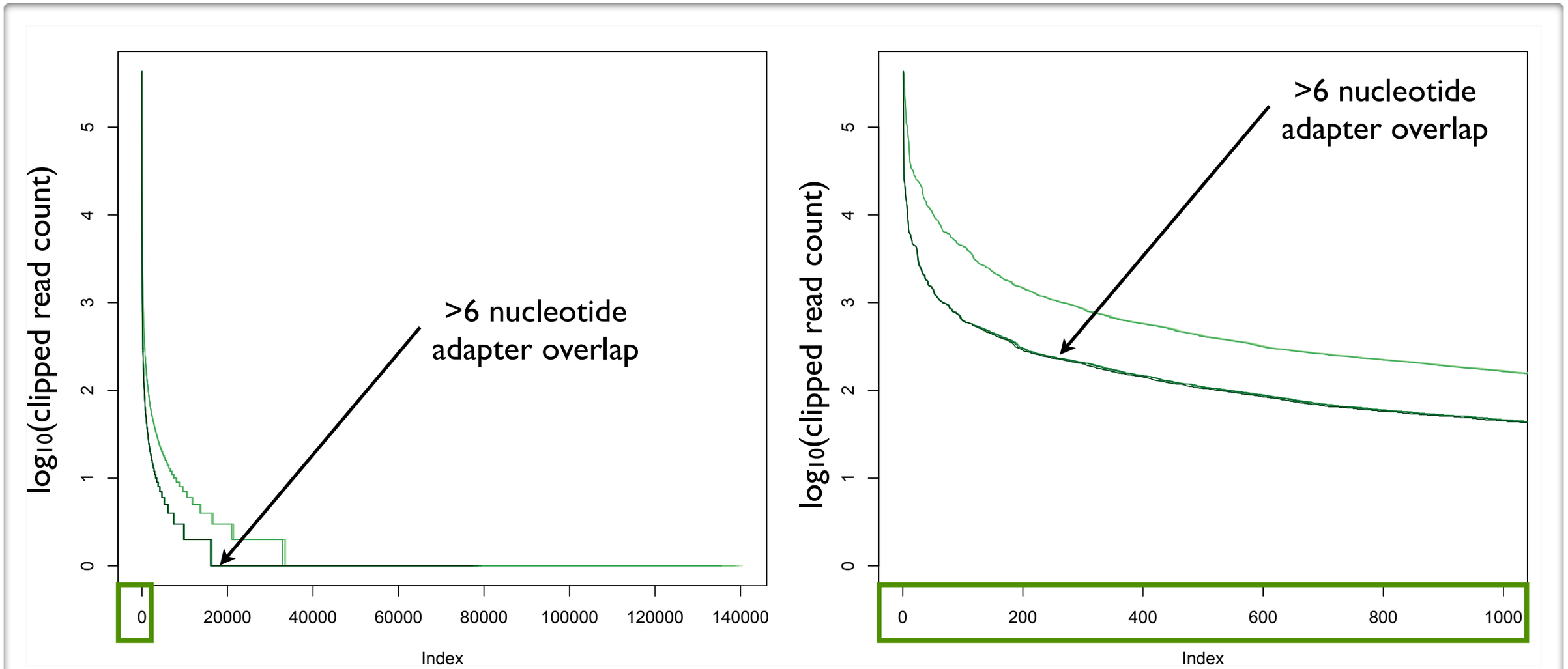


fastx_clipper

- read must contain at least 5 bases from the adaptor
- Clipping Adapter: ATCTCGTATGCCGTCTTCTGCTTGC
  Min. Length: 16
  Non-Clipped reads - retained.
  Input: 7,436,344 reads.
  Output: 4,148,344 reads.
  Output: 2,538,632 non-clipped reads.
- discarded 327,781 too-short reads.
  discarded 421,587 adapter-only reads.

fastx_collapser

- counts and collapses identical sequences



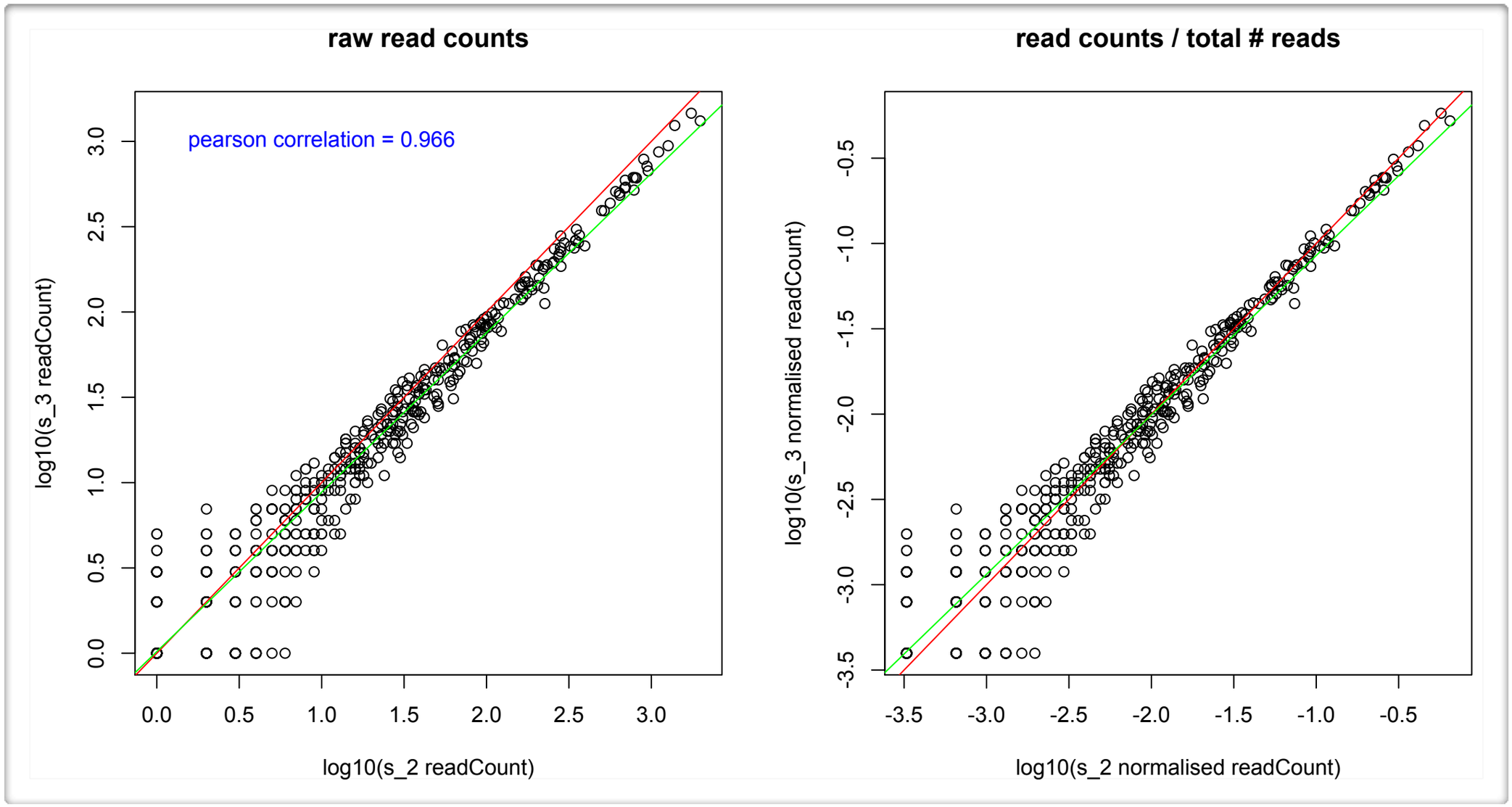
preliminary results

- miRanalzyer results for the 2 test datasets on Louise (using miRBase mapping):

result	s_2	s_3	s_3 / s_2	s_2 ∩ s_3	s_2 ∪ s_3
# unique reads	324,850	261,173	0.80	N/A	N/A
# unique mature miRNAs	538	530	0.99	481	587
# unique hairpin miRNAs	353	332	0.94	264	421
# unique mature* miRNAs	121	117	0.97	108	130

preliminary results

- 481 known miRNAs detected in s2 & s3:



to do

- use this pipeline?
- if(TRUE)
  - choose parameters for trimming
  - choose parameters for miRanalyzer
  - test on real data
  - install on server visible to all BrainSeqers