How many right vs wrong?

CHRIV from LYS4 to SNU56 – R1

* Total 25 trinity transcripts
* 6 of 7 genes annotated correctly
  + FMN1 has upstream start but regular start not annotated
* 7 transcripts which are gene fragments
* 1 correctly annotated AS
* Rest of stuff is random noncoding fragments – very low level

CHRVI from CDC4 to MSH4 – R2

* Total 26 trinity transcripts
* 7 genes seem annotated appropriately
* 4 antisense transcripts
  + 3 done correctly, one very low and therefore very fragmented.
* 11 crappy random fragments

CHRVII from RSC1 to ERG25 – R3

* Total 21 trinity transcripts
* 5 genes – 3 annotated correctly
  + SPR3 not expressed and therefore not annotated
  + ERG25 rare example of start site being different between biological replicates. Only the furthest stop site annotated
* snR48
  + annotated only as longer form from IP, steady state trimmed version not in annotation
* 4 correctly annotated non-coding RNAs
* 11 crappy random fragments

CHX from SLG1 to YOR011W-A

* Total 28 trinity transcripts
* 6 genes
  + 5 annotated correctly - TIR2 appears to have multiple start sites and trinity has only captured the furthest one upstream
* 4 Noncoding RNA annotated correctly
* 17 crappy random fragments

Conclusions

* Trinity does a good job of generating correct transcripts however it also generates lots of additional wrong fragments
  + Up to 50% of total is crappy
  + Of the good 50%, maybe 90% is correct
* Start sites are hard especial when multiple are present at one loci