Outline

Riboprofiling samples

Trimming reads

Removing contaminants

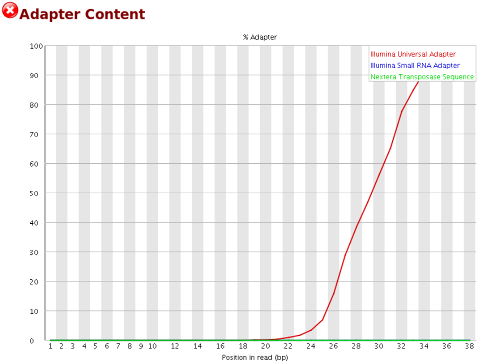
Genome mapping STAR

Comparing stats with references

Riboprofiling samples

	6h	18h	120h
Reads	42,130,444	45,982,977	44,077,266
Sequence length	50	50	50
%GC	59	55	56

Trimming



- min length: 25
- min adapter alignment length: 5
- unclipped discarded
- first base discarded

Trimming - Cutadapt

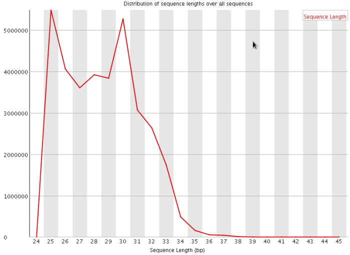
–quality-cutoff=10 Trim low-quality bases from 3’ ends of each read before adapter removal.

id	input	output	discarded TooShort	reads with Adapter
H006	42,130,444	34,439,444 (81.7%)	6,933,352 (16.5%)	41,271,051 (98.0%)
H018	45,982,977	39,441,134 (85.8%)	5,682,240 (12.4%)	44,907,869 (97.7%)
H120	44,077,266	32,501,482 (73.7%)	10,638,847 (24.1%)	42,985,050 (97.5%)

Sequence length distribution after trimming - Cutadapt

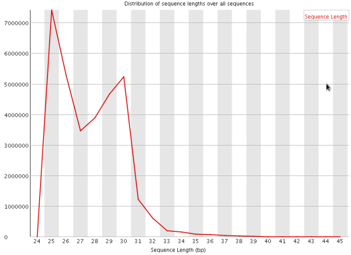
Sequence Length Distribution

H006



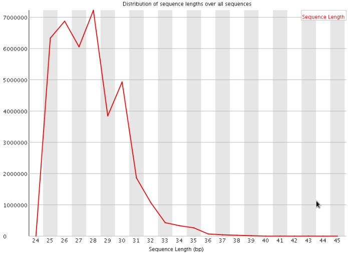
Sequence Length Distribution

H120



Sequence Length Distribution

H018



time	sequence length
H006	25-44
H018	25-44
H120	25-44

Removing contaminants - rRNA

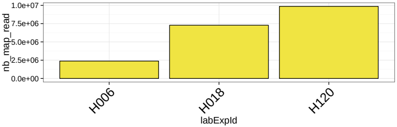
- STAR: without split mapping; max 10 multimaps

id	reads processed	uniquely mapped	multiple loci	too many loci	discarded too short
H006	34,439,444	29,084,185 (84.45%)	4,775 (0.04%)	158 (0.00%)	15.51%
H018	39,441,134	11,902,992 (30.18%)	73,911 (0.19%)	4,052 (0.01%)	69.62%
H120	32,501,482	12,525,154 (38.54%)	39,688 (0.12%)	10,985 (0.03%)	61.29%

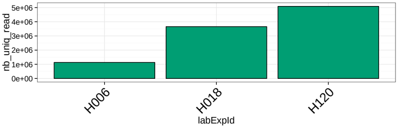
Genome mapping STAR - max 10 multimaps

- Unaligned reads from rRNA mapping
- –outFilterMatchNmin 16
- **max 10 multimaps**

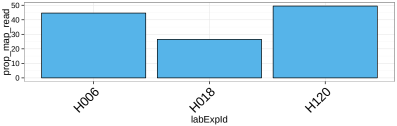
Number of mapped reads



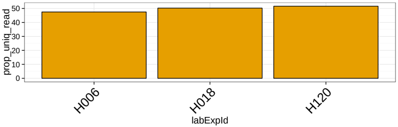
Number of uniquely mapped reads



Proportion of mapped reads



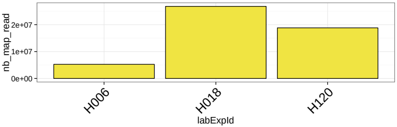
Proportion of uniquely mapped reads



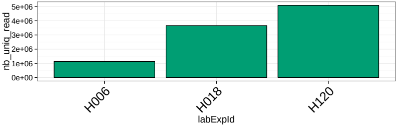
Genome mapping STAR - max 100 multimaps

- Unaligned reads from rRNA mapping
- –outFilterMatchNmin 16
- **max 100 multimaps**

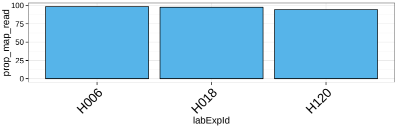
Number of mapped reads



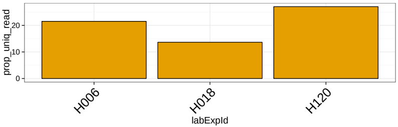
Number of uniquely mapped reads



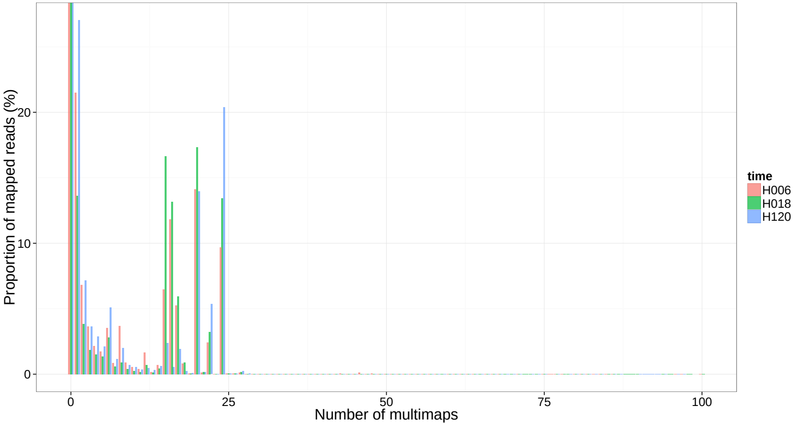
Proportion of mapped reads



Proportion of uniquely mapped reads

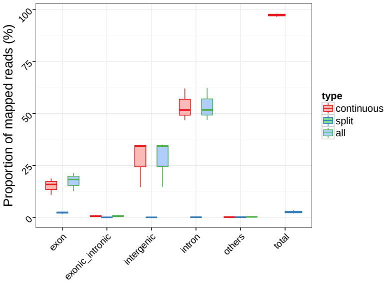


Distribution of multimaps

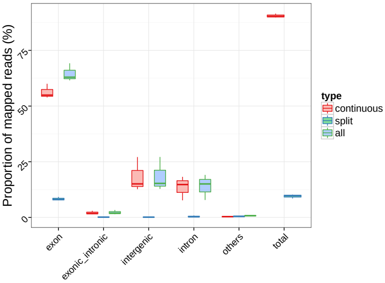


Genomic regions - max 100 multimaps

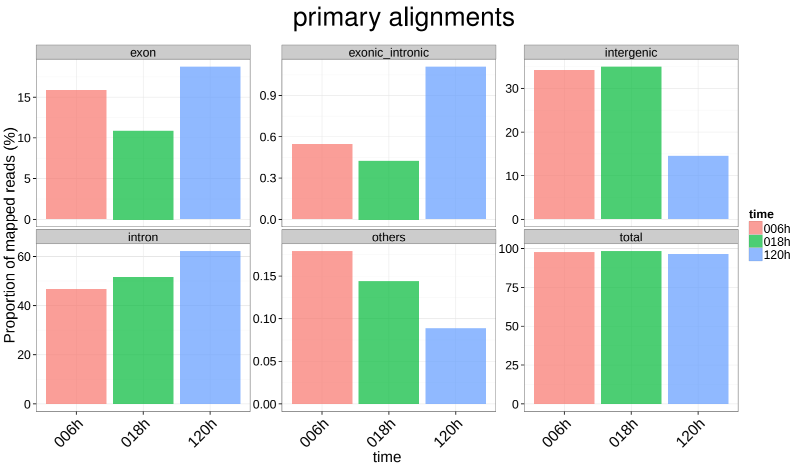
primary alignment



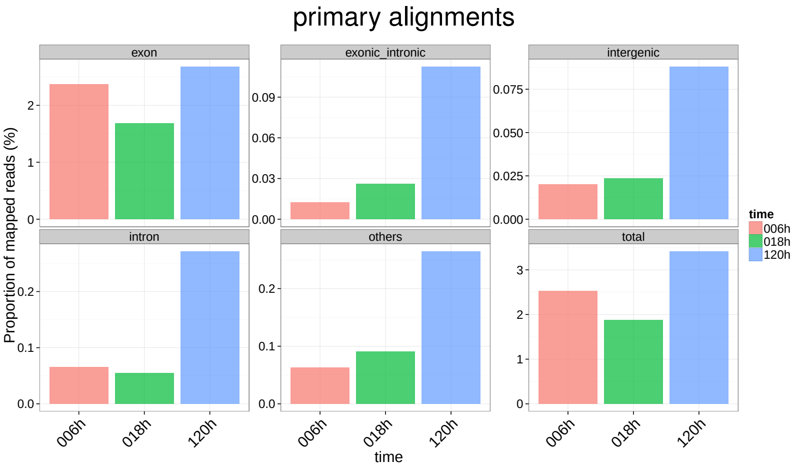
uniquely mapped reads



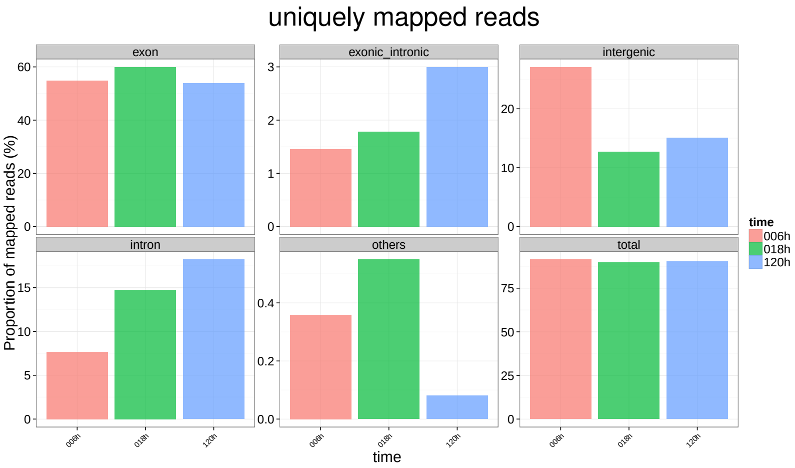
Genomic regions - continuous mapping - max 100 multimaps



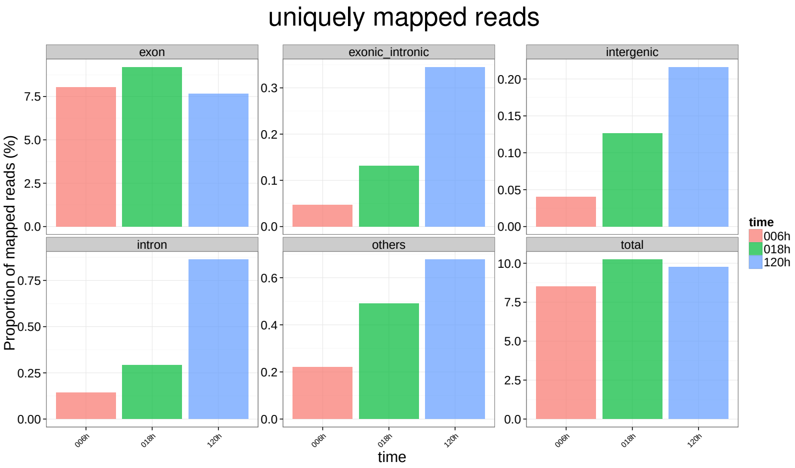
Genomic regions - split mapping - max 100 multimaps



Genomic regions - continuous mapping



Genomic regions - split mapping - max 100 multimaps



Comparing stats with references

Fatima’s lab

	RPF1 Mock	RPF1 KD	RPF2 Mock	RPF2 KD	RPF3 Mock	RPF3 KD
Total Reads	110,347,659	86,886,294	70,701,122	61,970,268	187,463,074	147,577,976
size-selected (22-36)	105,447,994	80,382,820	65,538,346	60,464,127	159,938,111	134,116,969
After rRNA,tRNA filtering	57,452,964	40,704,687	8,788,516	8,790,692	45,542,640	29,858,389
Aligned (-rRNA,tRNA)	22,252,759	15,717,018	4,376,899	4,232,701	23,323,320	12,759,154
In annotated CDSs	16,520,263	11,497,323	2,451,068	2,742,875	17,831,699	10,158,765

Current stats ERC

	6h	18h	120h
Reads	42,130,444	45,982,977	44,077,266
Size selected (25-44)	34,439,444	39,441,134	32,501,482
After rRNA filtering	5,340,484	27,464,231	19936640
Aligned (-rRNA, 10mm)	2,383,010	7,280,834	9,859,603
Exonic mapping primary alignment	945,086	3,315,618	3,986,174
Exonic uniquely mapped	712,202	2,531,362	3,130,052

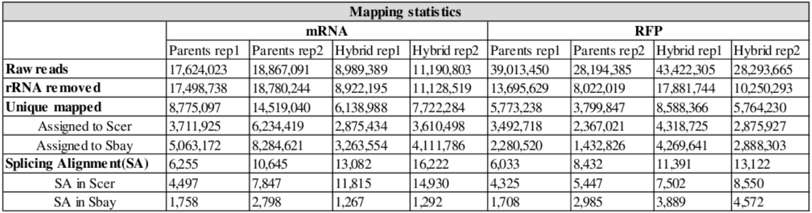
Comparing stats with references

GBE

Evolution of Gene Regulation during Transcription and Translation

Zhe Wang[1,†], Xuepeng Sun[1,2,†], Yi Zhao[1,3], Xiaoxian Guo[1], Huifeng Jiang[1,4], Hongye Li[2], and Zhenglong Gu[1,*]

Genome Biol. Evol. 7(4):1155–1167. doi:10.1093/gbe/evv059 Advance Access publication April 14, 2015 1155



Mapping statistics

	mRNA				**RFP**			
	Parents rep1	Parents rep2	Hybrid rep1	Hybrid rep2	Parents rep1	Parents rep2	Hybrid rep1	Hybrid rep2
Raw reads	17,624,023	18,867,091	8,989,389	11,190,803	39,013,450	28,194,385	43,422,305	28,293,665
rRNA removed	17,498,738	18,780,244	8,922,195	11,128,519	13,695,629	8,022,019	17,881,744	10,250,293
Unique mapped	8,775,097	14,519,040	6,138,988	7,722,284	5,773,238	3,799,847	8,588,366	5,764,230
Assigned to Scer	3,711,925	6,234,419	2,875,434	3,610,498	3,492,718	2,367,021	4,318,725	2,875,927
Assigned to Sbay	5,063,172	8,284,621	3,263,554	4,111,786	2,280,520	1,432,826	4,269,641	2,888,303
Splicing Alignment(SA)	6,255	10,645	13,082	16,222	6,033	8,432	11,391	13,122
SA in Scer	4,497	7,847	11,815	14,930	4,325	5,447	7,502	8,550
SA in Sbay	1,758	2,798	1,267	1,292	1,708	2,985	3,889	4,572

Cenik et al. Genome Res 2014

- To enable comparable analysis of high-throughput sequencing data sets, we used a uniform alignment and preprocessing pipeline.
- Reads were sequentially aligned using Bowtie 2 v.2.0.5 (Langmead and Salzberg 2012).
- All reads mapping to human rRNA and tRNA sequences were filtered out.

- **The remaining reads were aligned to APPRIS principal transcripts (release 12) (Rodriguez et al. 2013) from the GENCODE mRNA annotation v.15 (Harrow et al. 2012).**
- **For all transcript level analyses, reads that map only to coding regions were used.**

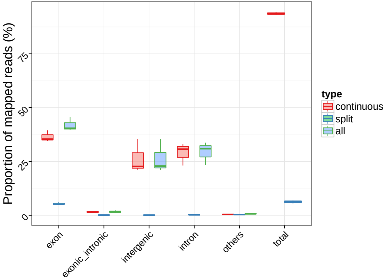
Cenik et al. Genome Res 2014

- **The remaining reads were aligned using parameters “-L 18 –norc” to APPRIS principal transcripts (release 12) (Rodriguez et al., 2013) from the GENCODE mRNA annotation v.15 (Harrow et al., 2012).**
- **This step was followed by alignment to all GENCODE transcripts and finally to the human genome (hg19).**
- **This strategy was preferred to avoid any differences in mappability of the exon-exon junction spanning reads due to read length differences between ribosome profiling and RNA-seq libraries.**
- We only retained alignments with a mapping quality greater than two for subsequent analyses.
- Reads mapping to coding regions, 5’UTRs, and 3’UTRs were counted separately using bedtools (Quinlan and Hall, 2010) and custom scripts.
- For all transcript level analyses, reads that map only to coding regions were used.

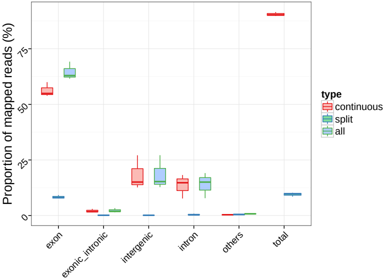
Extras

Genomic regions - max 10 multimaps

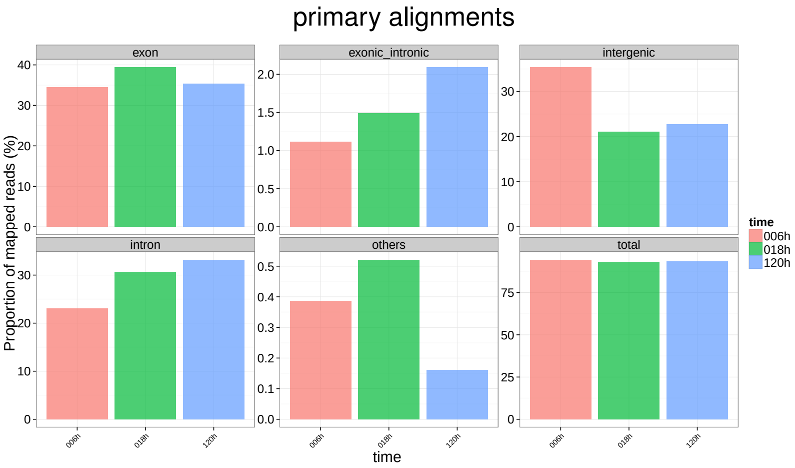
primary alignment



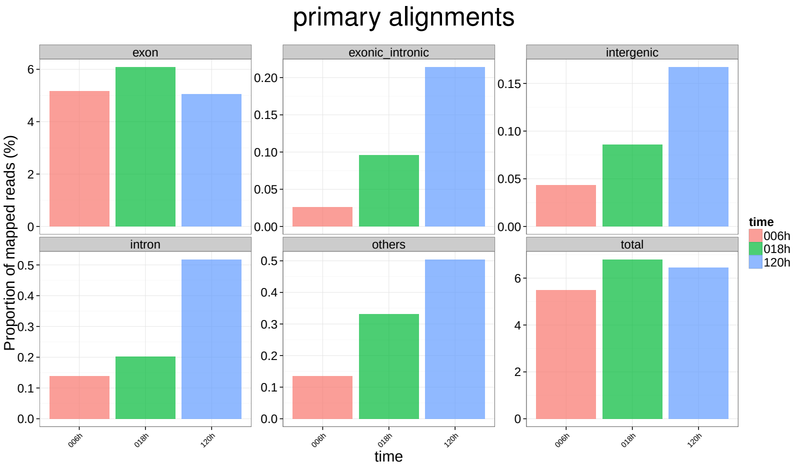
uniquely mapped reads



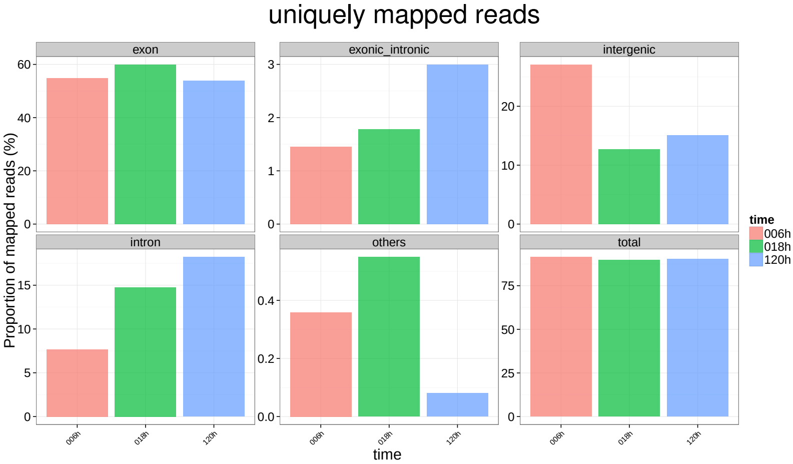
Genomic regions - continuous mapping - max 10 multimaps



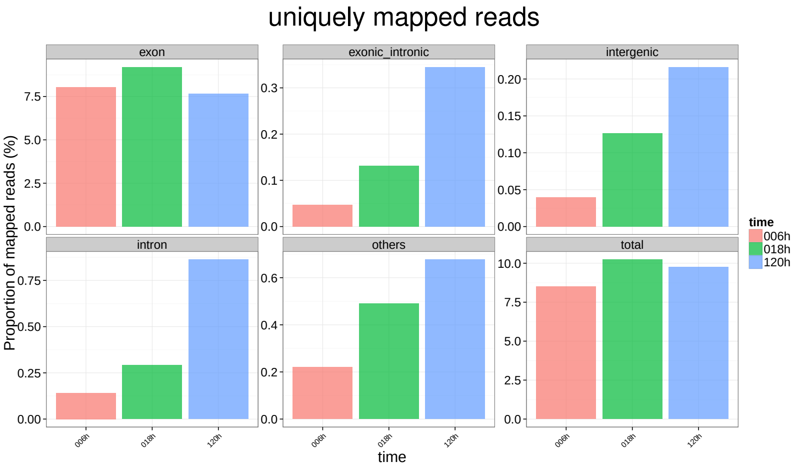
Genomic regions - split mapping - max 10 multimaps



Genomic regions - continuous mapping



Genomic regions - split mapping - max 10 multimaps


uniquely mapped reads
exon
exonic_intronic
intergenic
intron
others
total
Proportion of mapped reads (%)
time
006h
018h
120h
time
006h
018h
120h