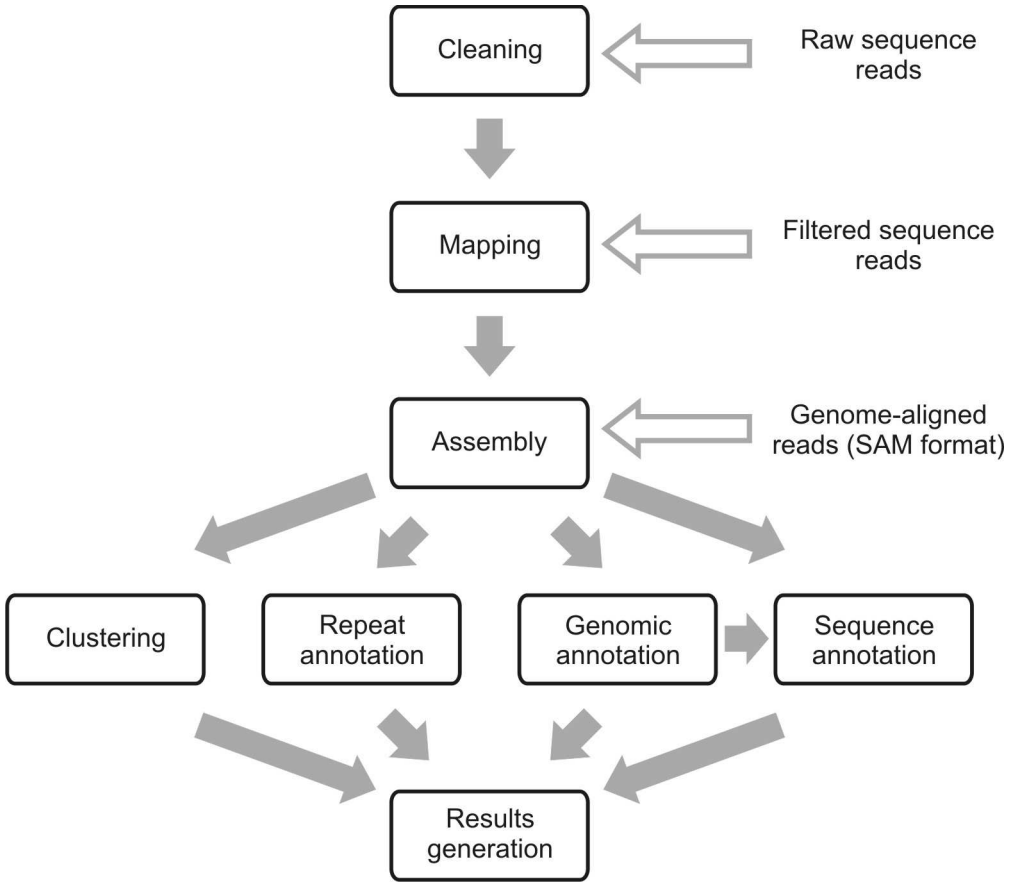
Supplementary Information

Revealing stable processing products from ribosome-associated small RNAs by deep-sequencing data analysis

Marek Zywicki[1]*, Kamilla Bakowska-Zywicka[1] and Norbert Polacek[1,2]*

[1]Innsbruck Biocenter, Medical University Innsbruck, Division of Genomics and RNomics, Fritz-Pregl-Strasse 3, 6020 Innsbruck, Austria, and [2]Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

*Correspondence should be addressed to M.Z. (marek.zywicki@i-med.ac.at) or N.P. (norbert.polacek@i-med.ac.at)



Supplementary Figure 1

A workflow of the APART pipeline. Subsequent steps of the analysis are shown and the information flow is indicated with the filled arrows. Three possible entry points for the analysis are indicated with open arrows.

A

Results sorted by: reads number maximum coverage loci length annotation position; general statistics; Genome Browser track file

Go to result page: 1 2

Contig name	Reads	Normalized reads (RPM)	Max coverage	Stable RNAs	Genomic uniqness	Length	Position	Repeat	Feature ID	Exon/Intron	Orientation	Feature status	Feature biotype	Feature alternative name	Feature description
loc.XII-454930_3867	3867	3.24e+05	3844	8..30_3726	2.00	250	chrXII:454937-455186	-	YLR154W-C	junction	antisense	Verified	gene	TAR1	Mitochondrial protein potentially involved in regulation of respiratory metabolism%3B interacts genetically with RPO41 and physically with Coq5p%3B encoded within the 25S rRNA gene on the opposite strand
									RDN25-1	exon	sense	-	rRNA	RDN25-1	25S ribosomal RNA%2C component of the large (60S) ribosomal subunit%3B encoded in the rDNA repeat (RDN1) as part of the 35S primary transcript
loc.XII+459675_1922	1922	1.61e+05	1028	3..34_923	4.82	125	chrXII:459676-459800	-	RDN5-1	exon	sense	-	rRNA	RDN5-1	5S ribosomal RNA%2C component of the large (60S) ribosomal subunit%3B localized to the nucleolus via interaction with Rpl5p%3B may play a role in translational frame fidelity%3B transcription is mediated by PolIII and activated by TFIIIA and TFIIIE
loc.XII-451783_913	913	7.66e+04	625	-	2.81	510	chrXII:451784-452293	-	RDN25-1	exon	sense	-	rRNA	RDN25-1	25S ribosomal RNA%2C component of the large (60S) ribosomal subunit%3B encoded in the rDNA repeat (RDN1) as part of the 35S primary transcript
loc.XII-455408_801	801	6.72e+04	679	10..166_454	2.00	172	chrXII:455409-455580	-	ITS1-1	junction	sense	-	rRNA	ITS1-1	Non-coding region between RDN58 and RDN18%2C that is transcribed as part of the 35S rRNA precursor transcript%3B excision during transcript maturation is coupled with processing of the 3' external transcribed spacer (3'ETS)
									ITS2-1	junction	sense	-	rRNA	ITS2-1	Non-coding region between RDN58 and RDN25%2C that is transcribed as part of the 35S rRNA precursor transcript%3B forms a stem-loop structure that is required for processing of the precursor transcript
									YLR154W-E	junction	antisense	Dubious	gene	YLR154W-E	Dubious open reading frame unlikely to encode a protein%3B encoded within the the 35S rRNA gene on the opposite strand
									RDN58-1	exon	sense	-	rRNA	RDN58-1	5.8S ribosomal RNA%2C component of the 60S ribosomal subunit%3B encoded in the rDNA repeat (RDN1) as part of the 35S primary transcript%3B 3' end formation involves processing by the exosome complex while 5' end is generated by Ker1p and Rat1p
loc.XII-455931_662	662	5.55e+04	513	-	2.00	340	chrXII:455932-456271	-	YLR154W-F	junction	antisense	Dubious	gene	YLR154W-F	Dubious open reading frame unlikely to encode a protein%3B encoded within the the 35S rRNA gene on the opposite strand
									RDN18-1	exon	sense	-	rRNA	RDN18-1	18S ribosomal RNA%2C component of the small (40S) ribosomal subunit%3B involved in codon recognition by tRNAs or the release factor eRF1 (Sup45p)%3B encoded in the rDNA repeat (RDN1) as part of the 35S primary transcript
loc.XII-454108_493	493	4.14e+04	284	295..312_130	2.00	585	chrXII:454109-454693	-	RDN25-1	exon	sense	-	rRNA	RDN25-1	25S ribosomal RNA%2C component of the large (60S) ribosomal subunit%3B encoded in the rDNA repeat (RDN1) as part of the 35S primary transcript
									YLR154W-B	exon	antisense	Dubious	gene	YLR154W-B	Dubious open reading frame unlikely to encode a protein%3B encoded within the the 25S rRNA gene on the opposite strand
loc.XII-453464_291	291	2.44e+04	288	-	2.00	94	chrXII:453465-453558	-	RDN25-1	exon	sense	-	rRNA	RDN25-1	25S ribosomal RNA%2C component of the large (60S) ribosomal subunit%3B encoded in the rDNA repeat (RDN1) as part of the 35S primary transcript
loc.XII-452580_226	226	1.90e+04	53	-	2.00	725	chrXII:452581-453305	-	YLR154W-A	junction	antisense	Dubious	gene	YLR154W-A	Dubious open reading frame unlikely to encode a protein%3B encoded within the the 25S rRNA gene on the opposite strand
									RDN25-1	exon	sense	-	rRNA	RDN25-1	25S ribosomal RNA%2C component of the large (60S) ribosomal subunit%3B encoded in the rDNA repeat (RDN1) as part of the 35S primary transcript

B

loc.VII+1004253_46

Length: 40

Number of reads aligned: 46

Normalized expression level (RPM): 3.86e+03

Maximum coverage: 46

Genomic uniqness: 3.00

Genomic position: chrVII:1004254-1004293

Genome mapped repeats overlapping the loci: -

Alignment of read-derived sequence consensus to genomic sequence:

```
Genomic:   ....................A.............
Consensus: TGTAATGCAAAGGTCGTTAGTTCATTTCTGACAGGTGGCA
Cons.qual  99999999999999999999999979999999999999
```

Predicted stable RNA species:

```
Length: 34
Read number: 31
Genomic:         ................A.............
Consensus:     7 GCAAAGGTCGTTAGTTCATTTCTGACAGGTGGCA 40
Cons.qual        99999999999999999979999999999999
```

Annotation features overlapping the contig:

Feature ID	Exon/Intron	Orientation	Feature status	Feature biotype	Feature alternative name	Feature description
tT(UGU)G2	exon	sense	-	tRNA	tT(UGU)G2	tRNA-Thr

Other genomic loci with partial sequence match:

Contig name	Reads	Normalized reads (RPM)	Max coverage	Stable RNAs	Genomic uniqness	Length	Position	Repeat	Feature ID	Exon/Intron	Orientation	Feature status	Feature biotype	Feature alternative name	Feature description
loc.VII+661784_46	46	3.86e+03	46	7..40_31	3.00	40	chrVII:661785-661824	-	tT(UGU)G1	exon	sense	-	tRNA	tT(UGU)G1	tRNA-Thr
loc.XVI+744313_46	46	3.86e+03	46	7..40_31	3.00	40	chrXVI:744314-744353	-	YPR108W-A	exon	sense	Uncharacterized	gene	YPR108W-A	Putative protein of unknown function%3B identified by fungal homology and RT-PCR
									tT(UGU)P	exon	sense	-	tRNA	tT(UGU)P	tRNA-Thr

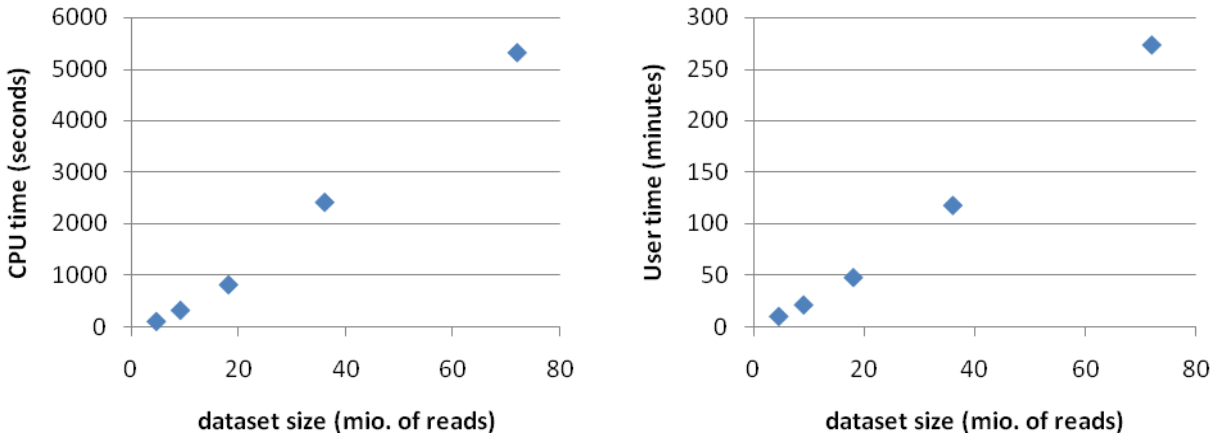
Download sequences of the reads

Download original alignment of the reads in SAM format

Supplementary Figure 2

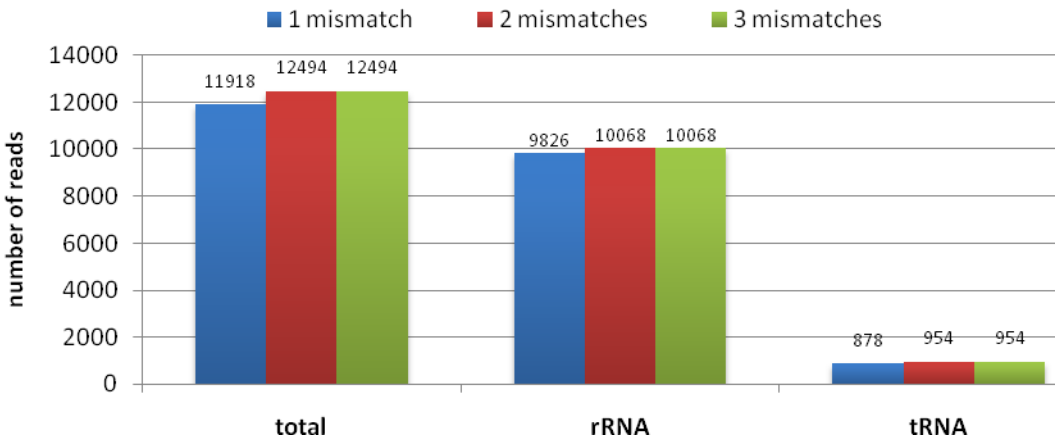
Example output from the APART pipeline. **A)** The main table presenting the representative contigs and information gathered during the analysis and annotation. For every contig three hyperlinks are provided. First points to the detail page of the contig, second to corresponding position in genome browser, and the last to detailed information about the annotation feature assigned to the contig. **B)** Detail page for the ribosomal RNA contig presented in Supplementary Figure 3. Besides of the information from the main table, the sequences of the contig and predicted processing products are provided, including the alignments with the reference genome and consensus quality values. Additionally all the clustered contigs are listed together with the annotation and analysis results.

A **B**



Supplementary Figure 3

Dependence of the APART running times on input read numbers. **A)** CPU time course **B)** User time course. In both cases the time increase is linear.



Supplementary Figure 4

Comparison of number of reads aligning to the reference genome by using different number of mismatches allowed. Two known hyper-modified types of ncRNA transcripts (rRNA, tRNA) are shown. The gain of reads in case of rRNA and tRNA by allowing more mismatches is comparable to total gain of alignments (total), suggesting that in yeast RNA modifications do not influence the mapping procedure.