Two methods for full-length RNA sequencing for low quantities of cells and single cells

Pan X, Durret R et al.

Proceedings of the National Academy of Sciencies,
published January 8, 2013

Triinu Kõressaar
Seminar in Bioinformatics

TARTU 2013

Introduction

To solve some biological problems, low quantities of cells or single cells expression pattern is required

Current methods preserve either 3' biased or variable transcript representation

The method introduced here is for the amplification and high-throughput sequencing of very small amounts of RNA

The method produces :
– uniformly distributed sequences
- which cover the full length of almost all transcripts independent of their sizes
- from 1,000 to 10 cells and even with single cells

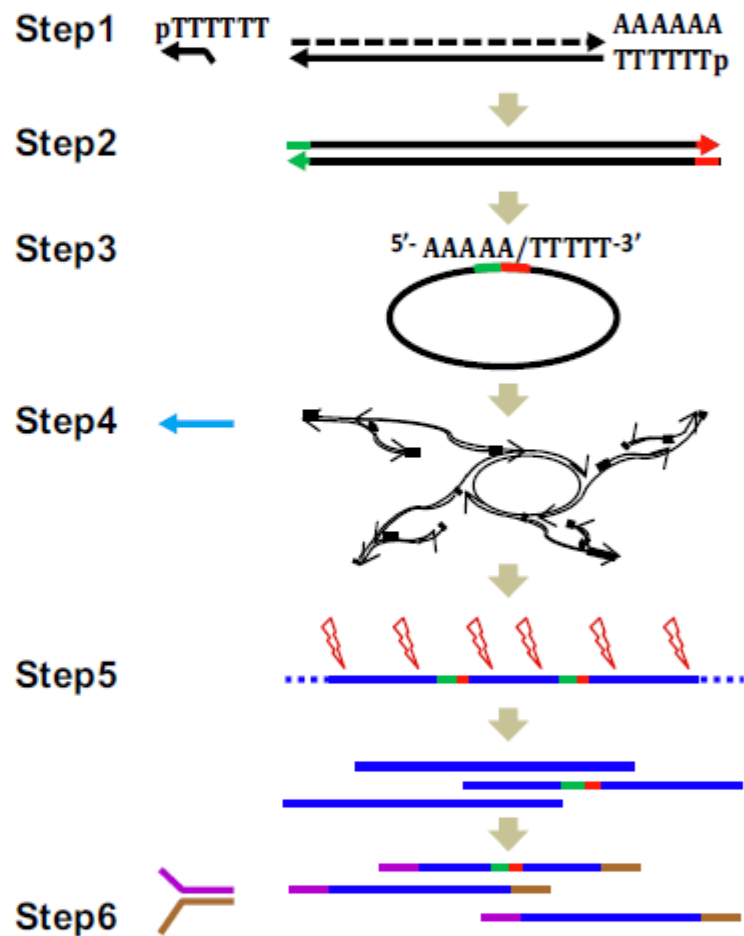
Previous approaches

1. RT and oligo-dT primers are applied with a T7 phage RNA polymerase promoter sequence.

   This process often truncates the cDNA molecule, losing 5′ sequences of the original mRNA and requires multiple rounds of processing when starting with low quantities cells

2. Several methods, based on PCR amplification of cDNA

   Yield biased representation of sequences
   Fail to give complete sequences for long mRNAs

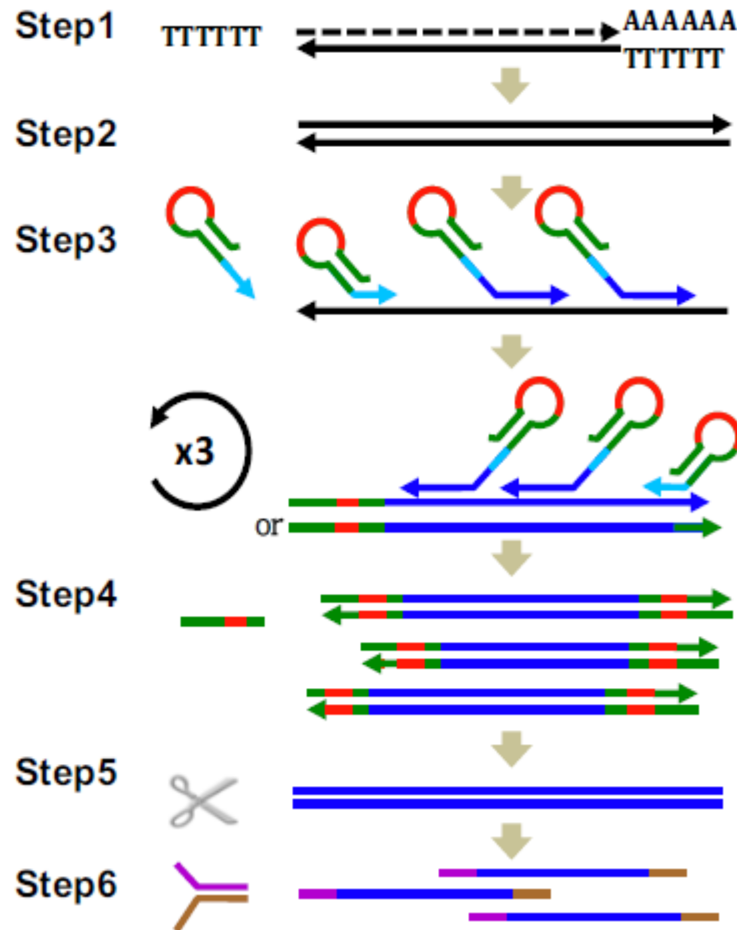


PMA - Phi29 DNA polymerase-based mRNA transcriptome amplification

Full length mRNA-derived cDNA was circularized by intramolecular ligation before amplification

This method captures potentially all end sequences

PTA – Phi29 DNA polymerase-based transcriptome amplification



SMA – semirandom primed PCR-based mRNA transcriptome amplification

After cDNA generation, semirandom primed PCR is used to amplify the overlapping segments along the entire length of cDNAs for mRNA sequencing

STA – semirandom primed PCR-based whole transcriptome amplification

Comparision of PMA and SMA

K562 erythroleukemic cell RNA cells
Diluted aliquots – 1 000 (k), 100 (h) and 10 (t) cells of

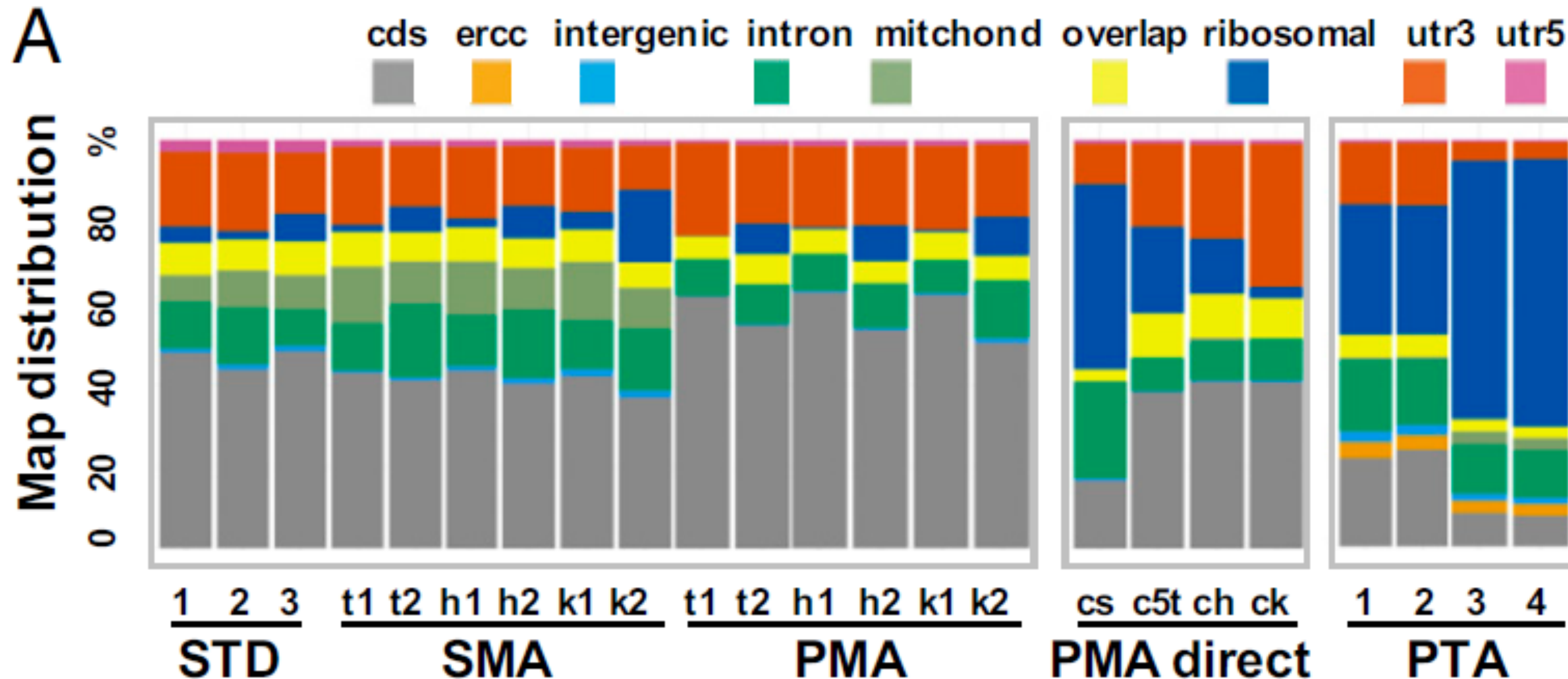


Fig. Mapping distribution of cDNA sequence reads from SMA and PMA amplicons and unamplified controls (**STD**)
ercc- seeded synthetic oligonucleotides
overlap – sequences overlapping more than one type of target
PMA direct – cDNA prepared from cell lysates
cs –single cell, ***c5t*** – 50 cells, PTA – 3,000 cells

Comparision of PMA and SMA

The majority of sequences are sequenced only once

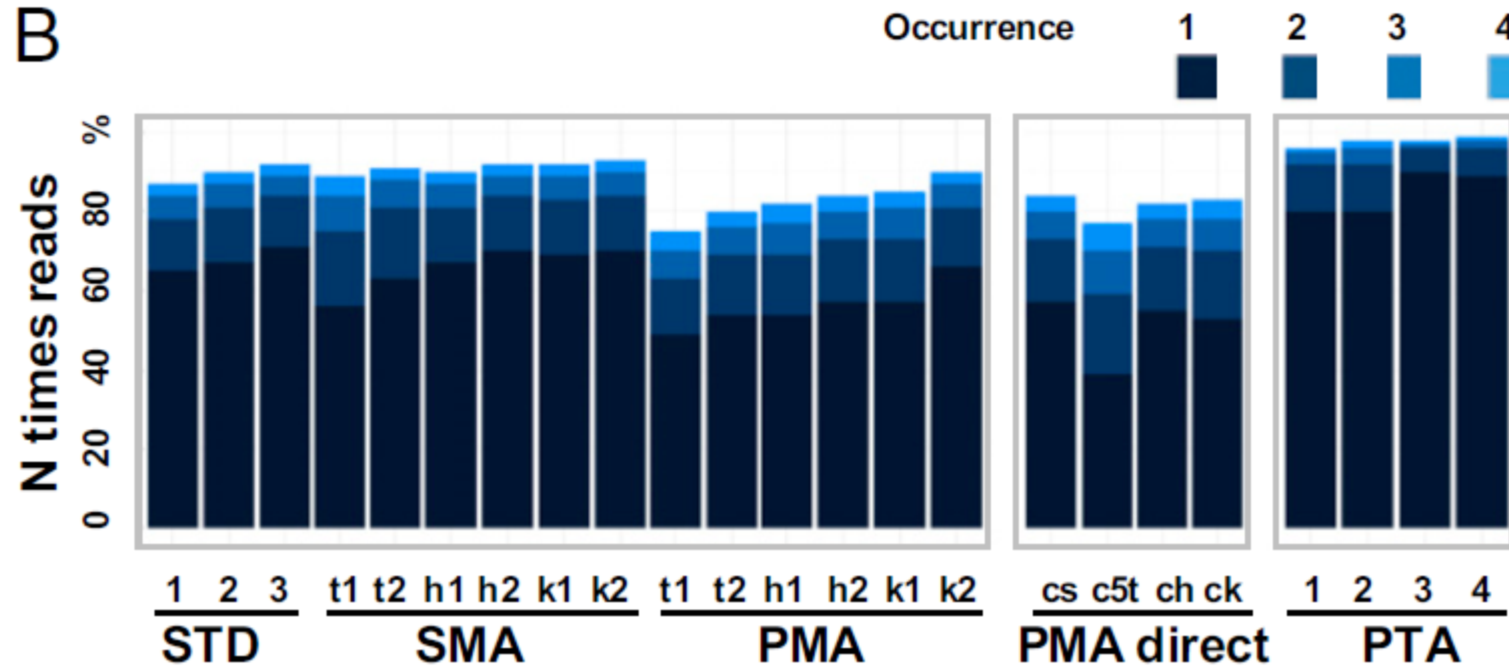


Fig. Frequency of multiply sequenced cDNA fragments

Transcript full length coverage

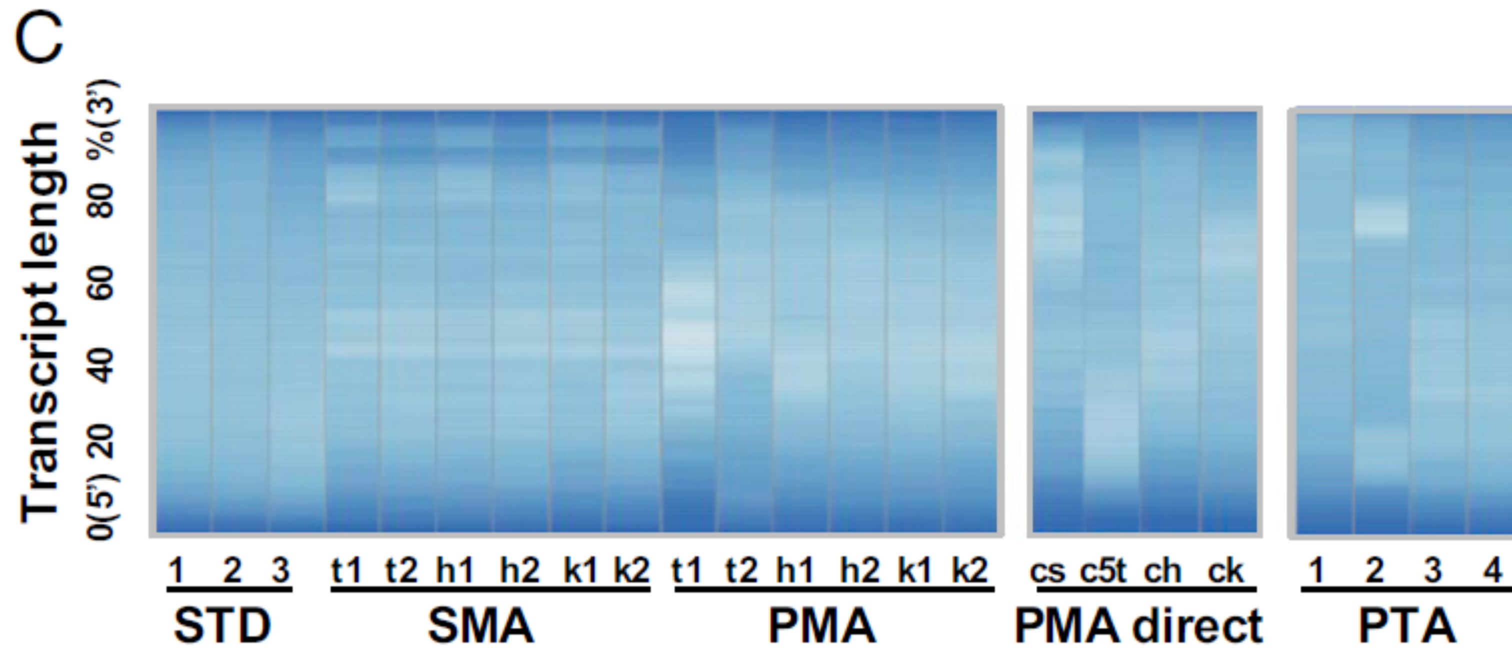


Fig. Average coverage of transcripts from 5' to 3' ends. The transcripts were divided into 100 equal segments and the relative intensity of each segment was displayed. Darker blue represents less coverage.

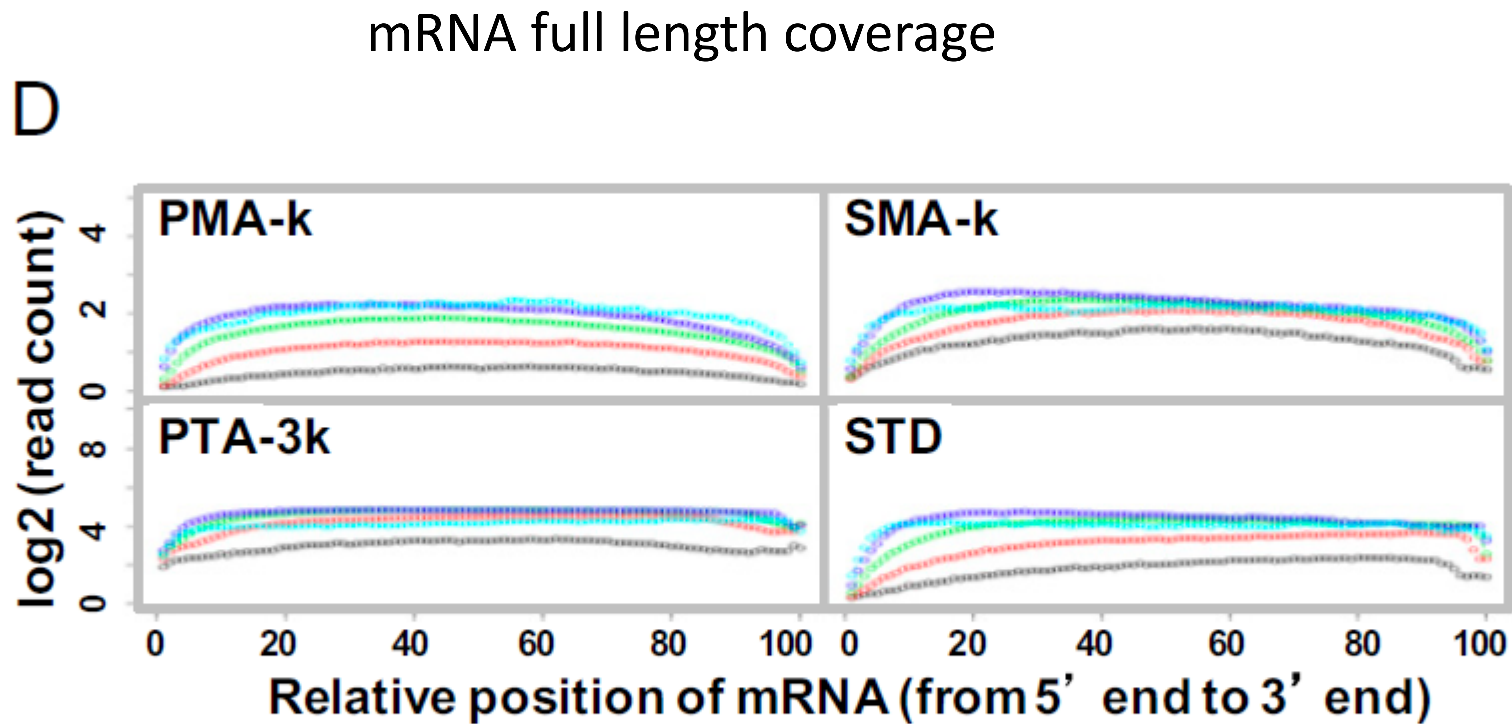


Fig. Distribution of reads across of cDNAs of various lengths; 1,000 K562 cells

RefSeq genes are divided into five classes according to mRNA length. **Black**: 45–1,000 bp, 4,081 genes (transcripts); **red**: 1,001–2,000 bp, 5,483 genes; **green**: 2,001–5,000 bp, 9,634 genes; **blue**: 5,001–10,000 bp, 2,264 genes; **cyan**: 10,001–101,674 bp, 259 genes.

Gene identification

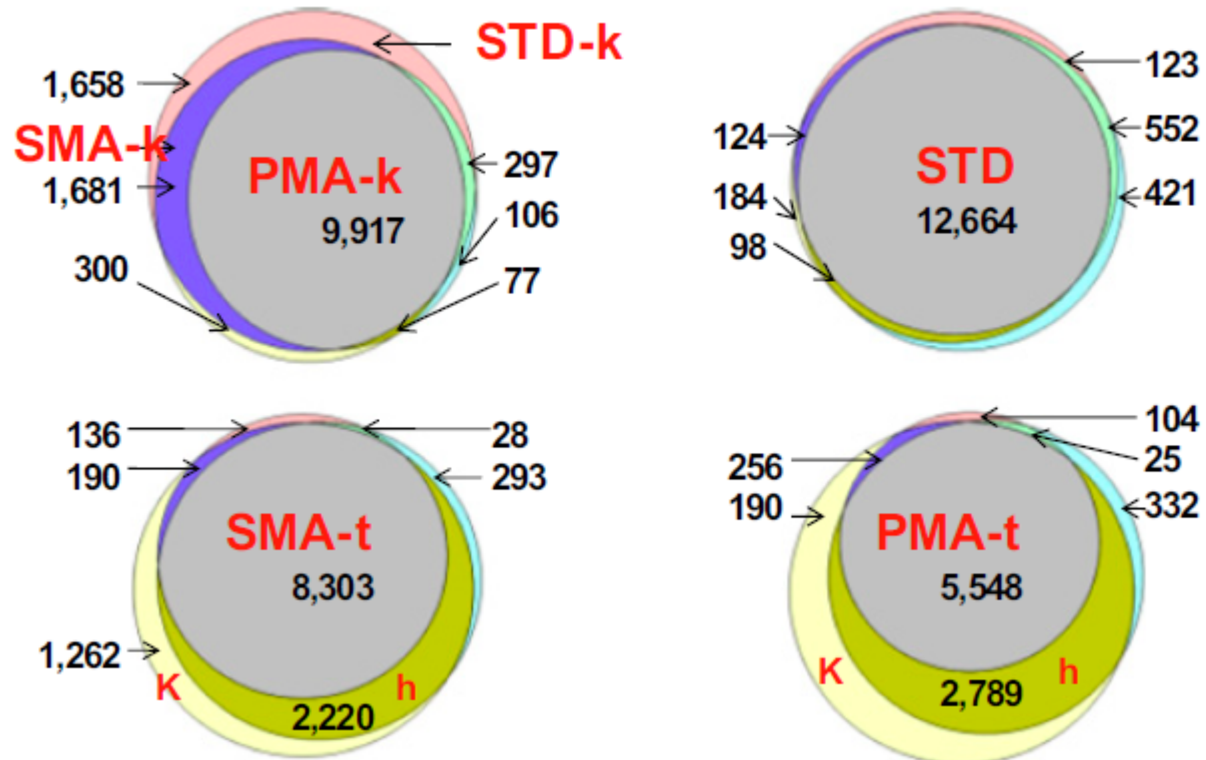


Fig. Number of genes detected from various numbers of cells (>0.1RPKM – reads per kilobase of mRNA per million total reads)

Single cell mRNA transcriptome sequencing

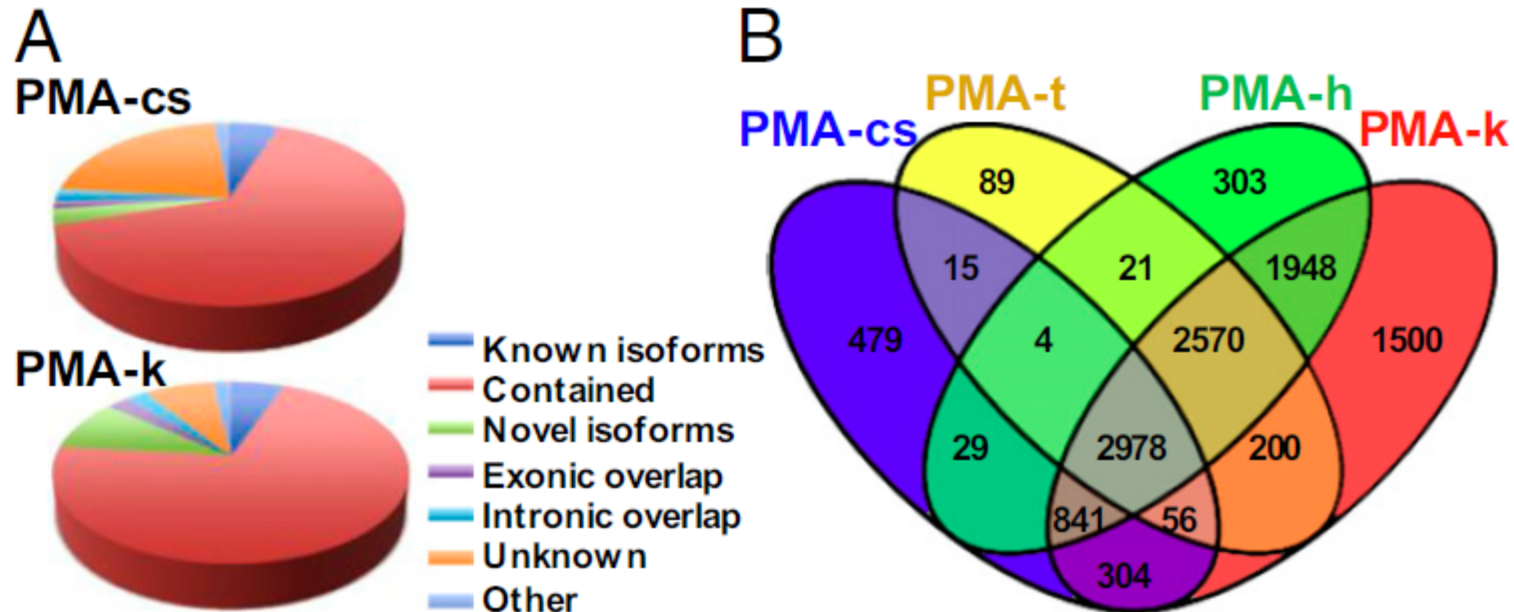


Fig. A. PMA-cs totally covers 5,277 genes (each at least w one transcript). **Novel isoforms** – multiexonic transcripts, which which share at least one known exon. The "**exonic overlap**" and "**intron overlap**" are those signals that are not included in known or novel isoforms, whereas "**exons**" indicate single-exonic transcripts, which are overlapped with known exons. "**Contained**" means truncated isoforms where, for example, one exon is not detected, but the other exons completely match with known exons, so this category mostly contains known isoforms.

Conclusions

Two methods introduced

1. Both of them produce full-length coverage of the RNA sequences

2. Both of them are independent of the length of the transcript (even 23kb cDNA could be amplified)

3. Both of them can cover 5’ UTR and 3’ UTR

SMA detects more genes than PMA, is more similar to STD and is more sensitive (less starting material required, more suitable for sc RNA amplification)

When suitable technology arises, PMA might turn useful as it generates intact full length copies of cDNA (e.g. splice isoforms)