Hidden layers of complexity

How small RNA distribution and sample handling affect sequencing results

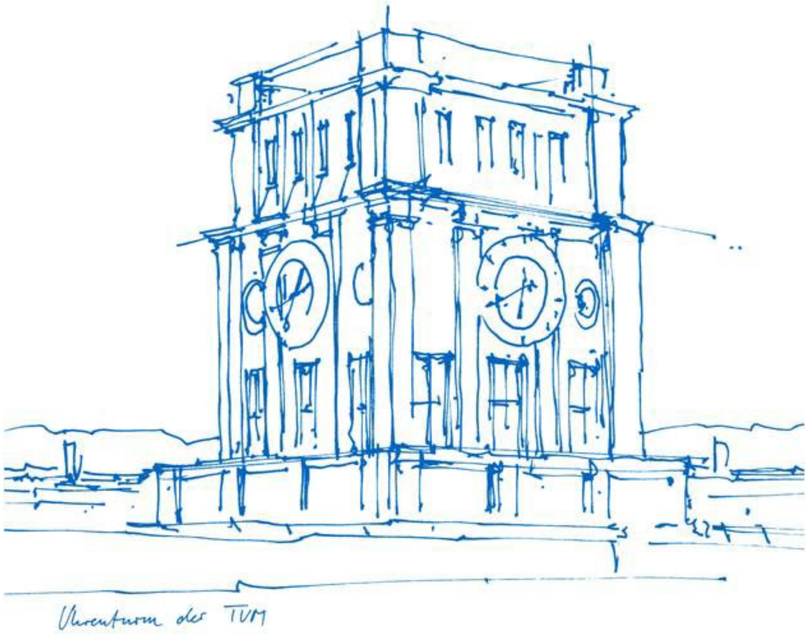
Dr. Benedikt Kirchner

Technische Universität München

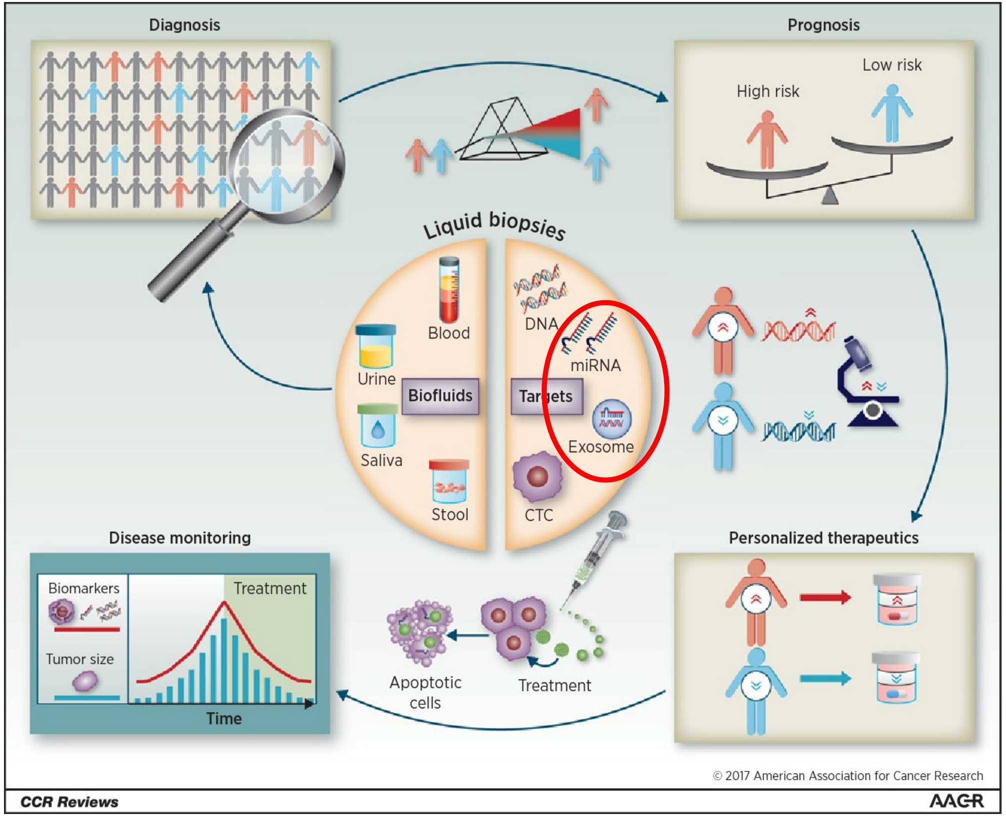
School of Life Sciences

Division of Animal Physiology and Immunology

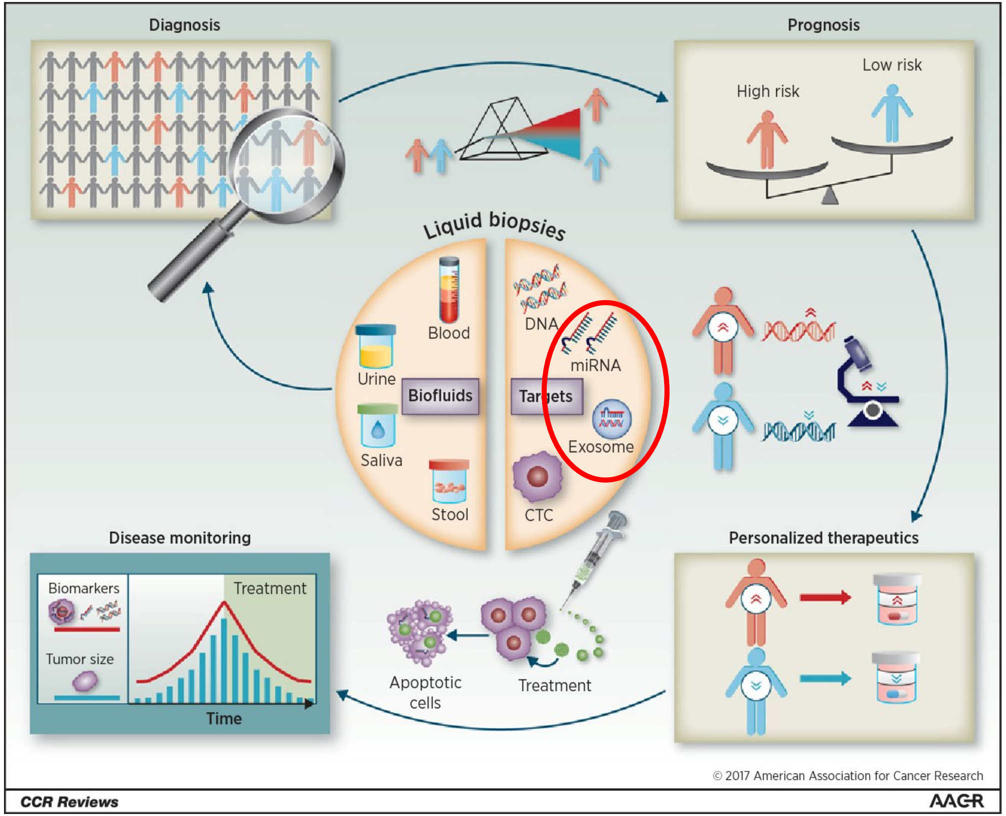
June 30th, 2023



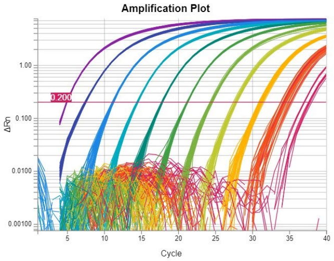
Biomarker development in clinical settings



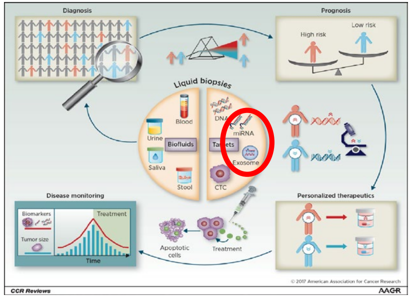
Biomarker development in clinical settings

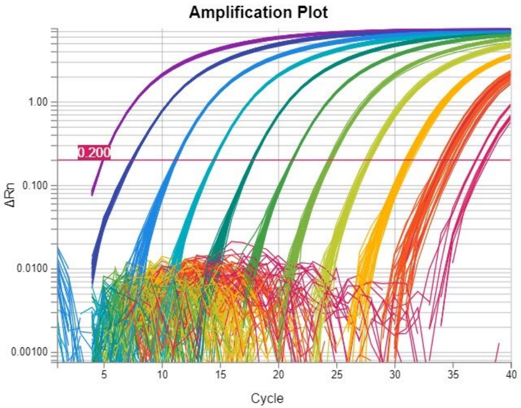


© 2017 American Association for Cancer Research



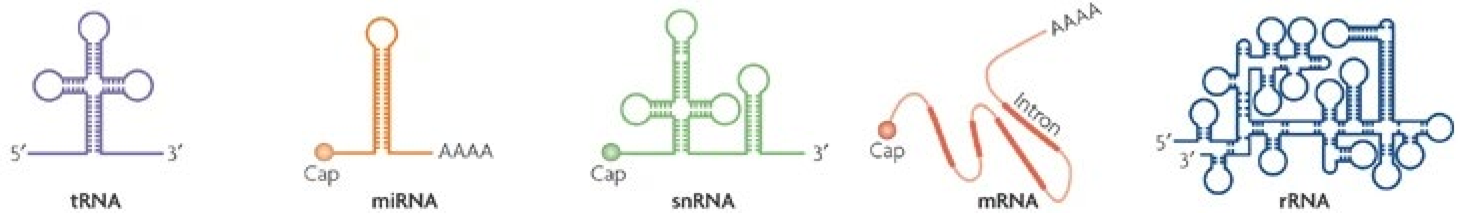
Disagreement in methods





≠

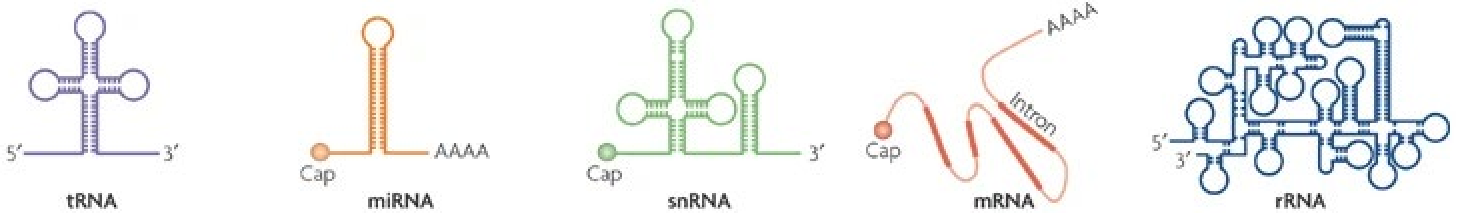
Size selection bias in small RNA NGS

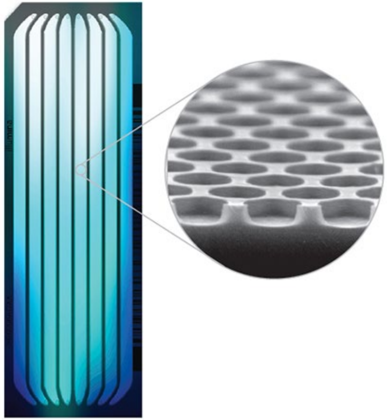


Enrichment of RNAs smaller than ~50 nt

Köhler et al, Nature Reviews Molecular Cell Biology 2007

Size selection bias in small RNA NGS

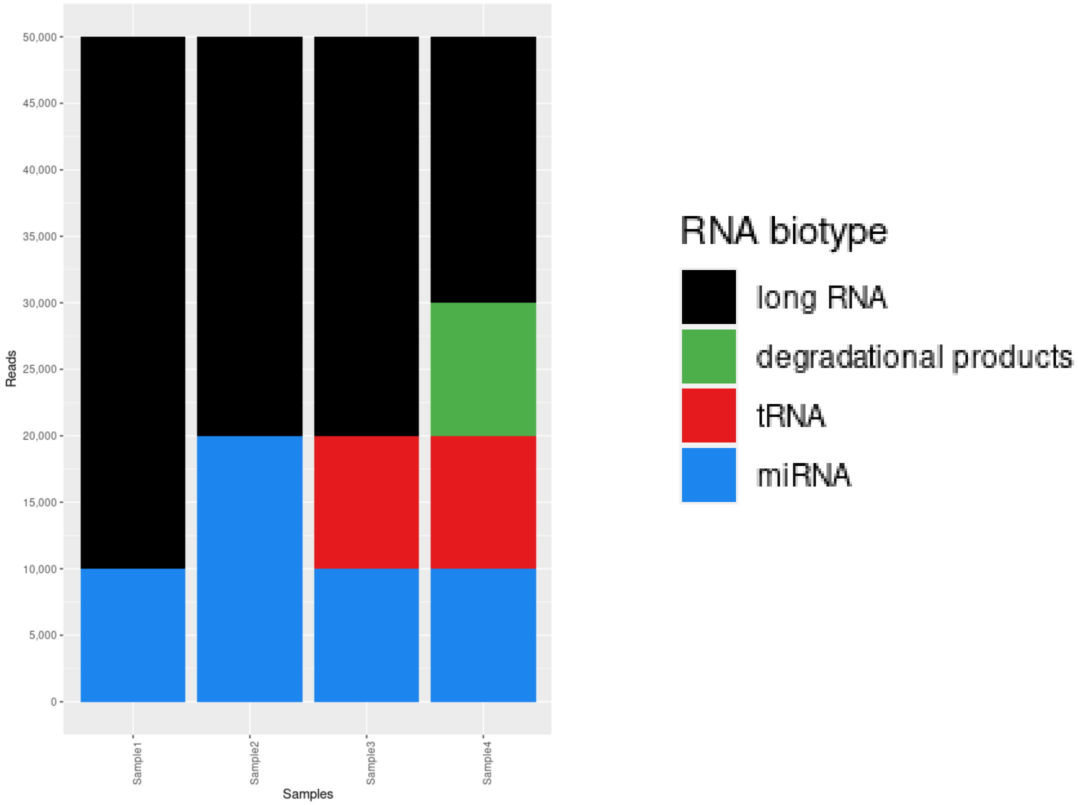




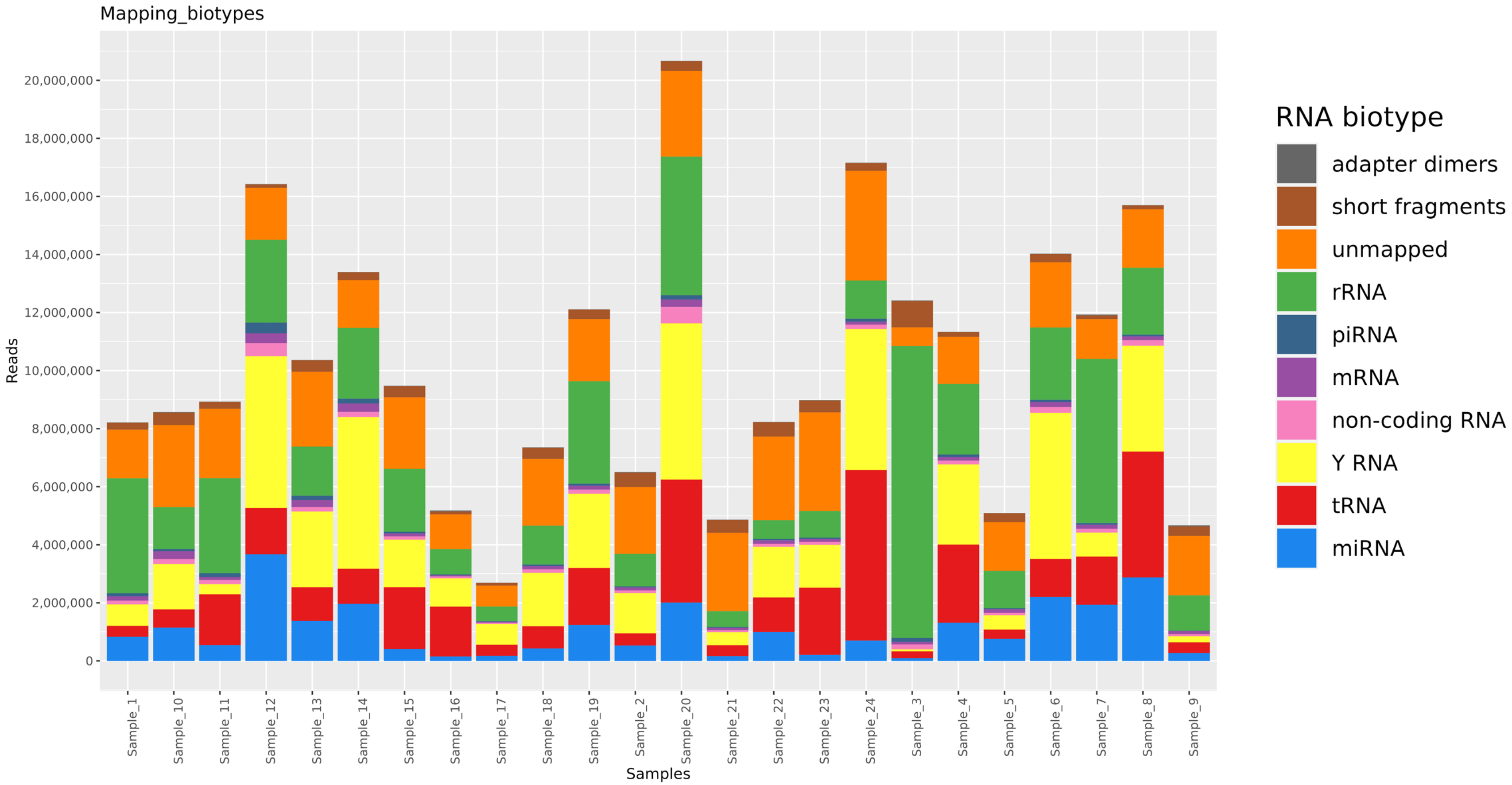
- Finite number of reads per flowcell
- Quantification always in relation to other small RNA transcripts
- Spike-in controls still rarely used

Köhler et al, Nature Reviews Molecular Cell Biology 2007

Importance of normalization strategies

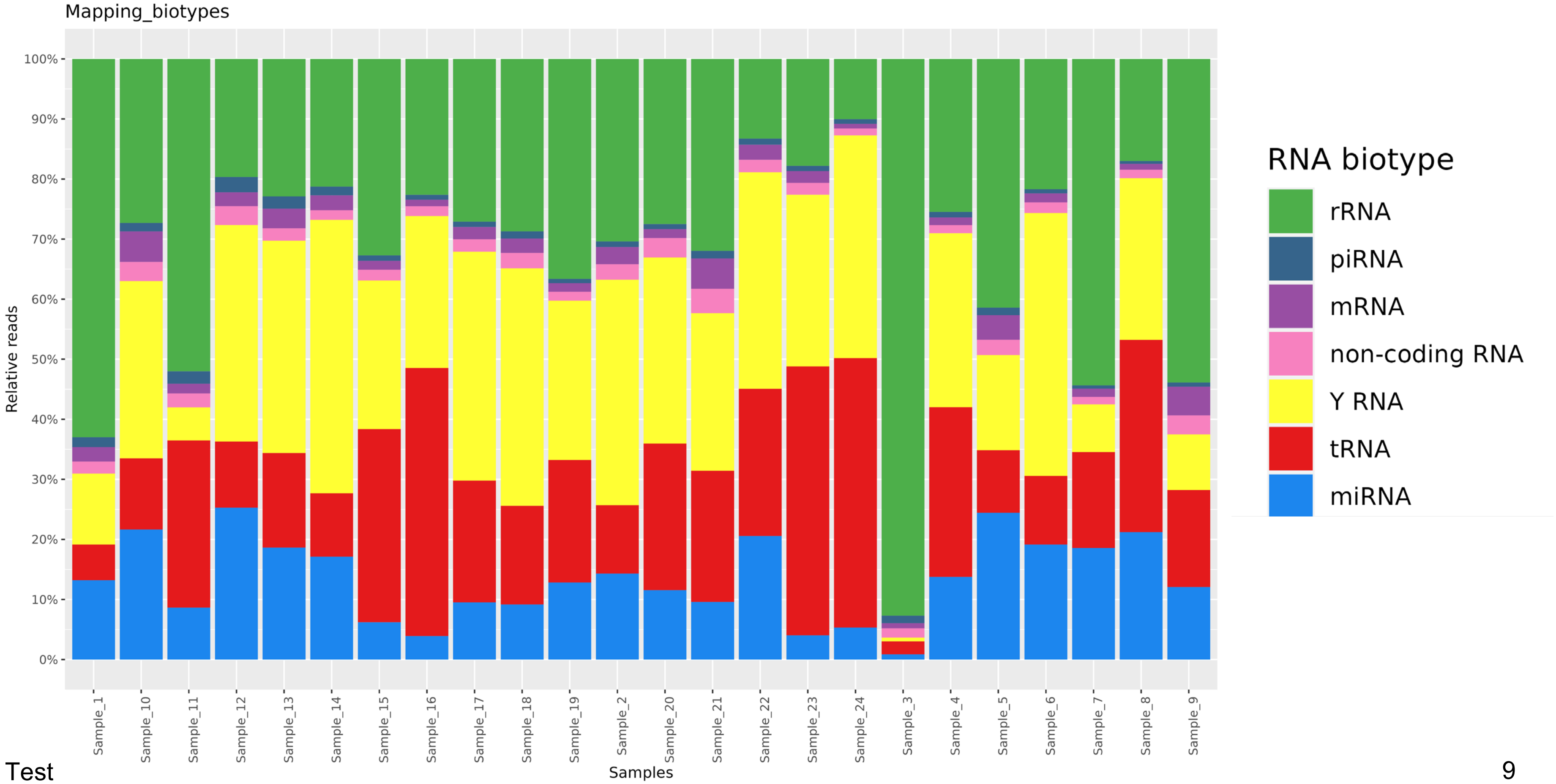


Heterogenous biotype distribution

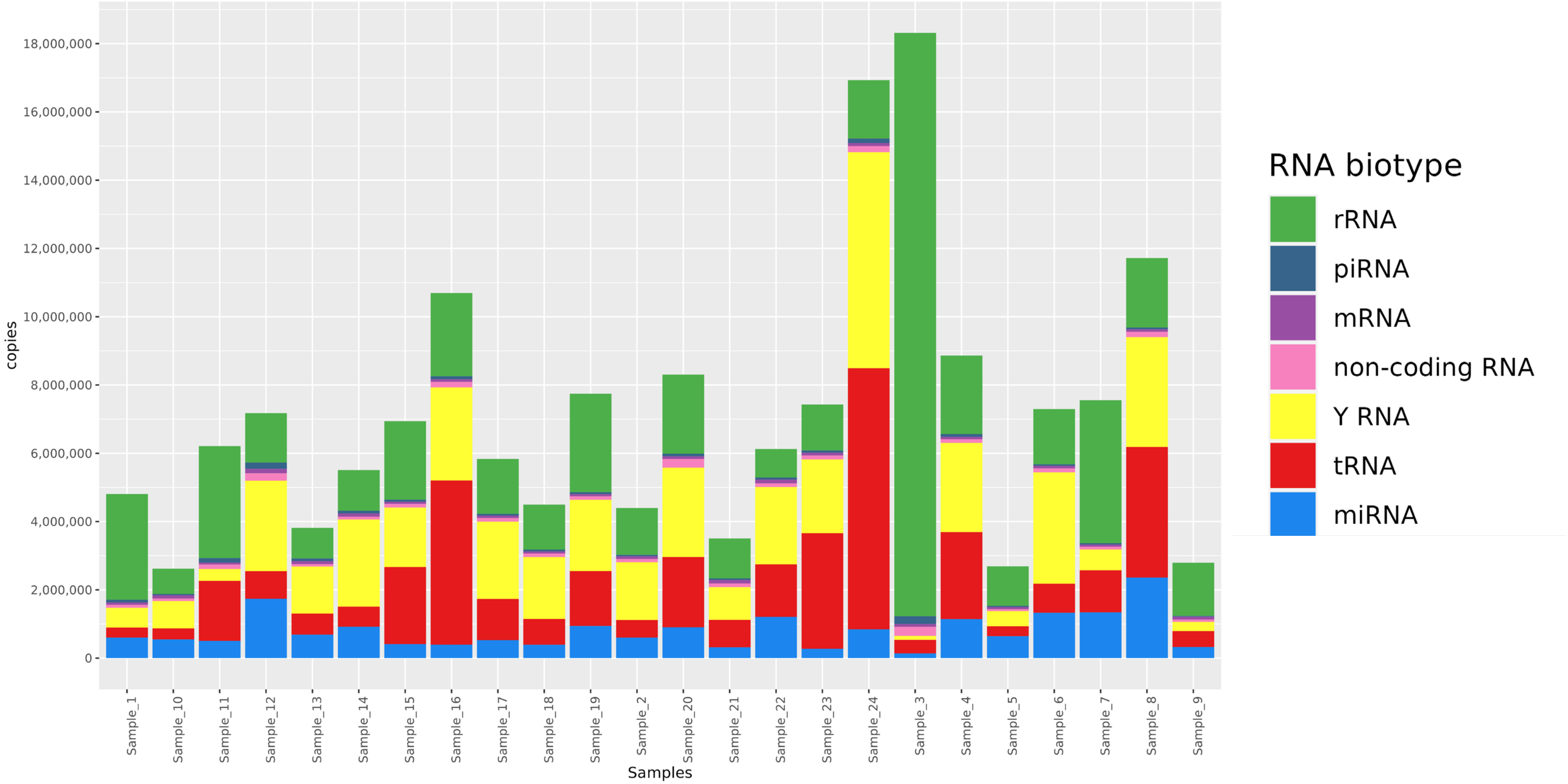


Test

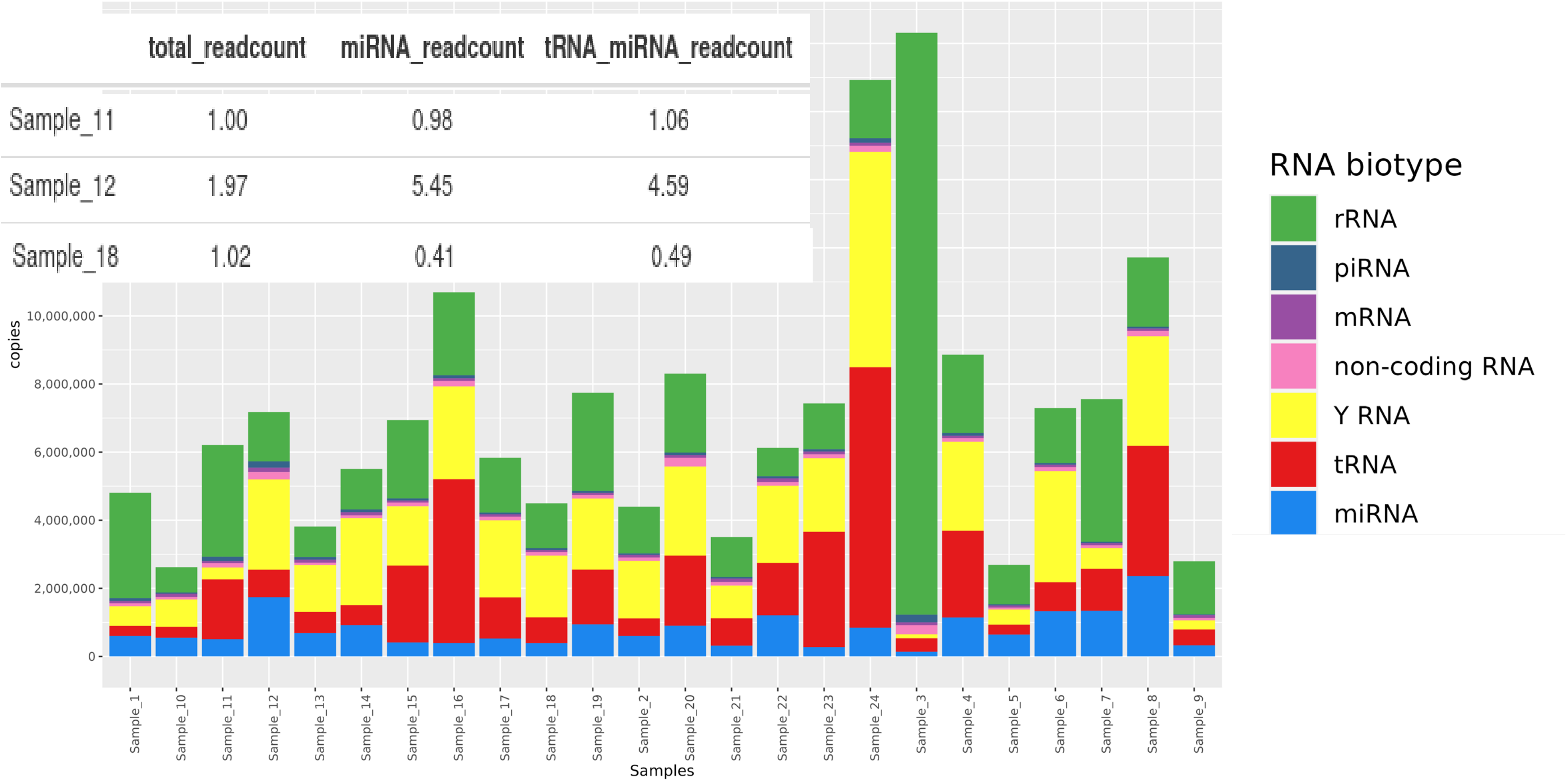
Library size normalization is skewed by transcript abundance



Median gene expression normalization is skewed by observations

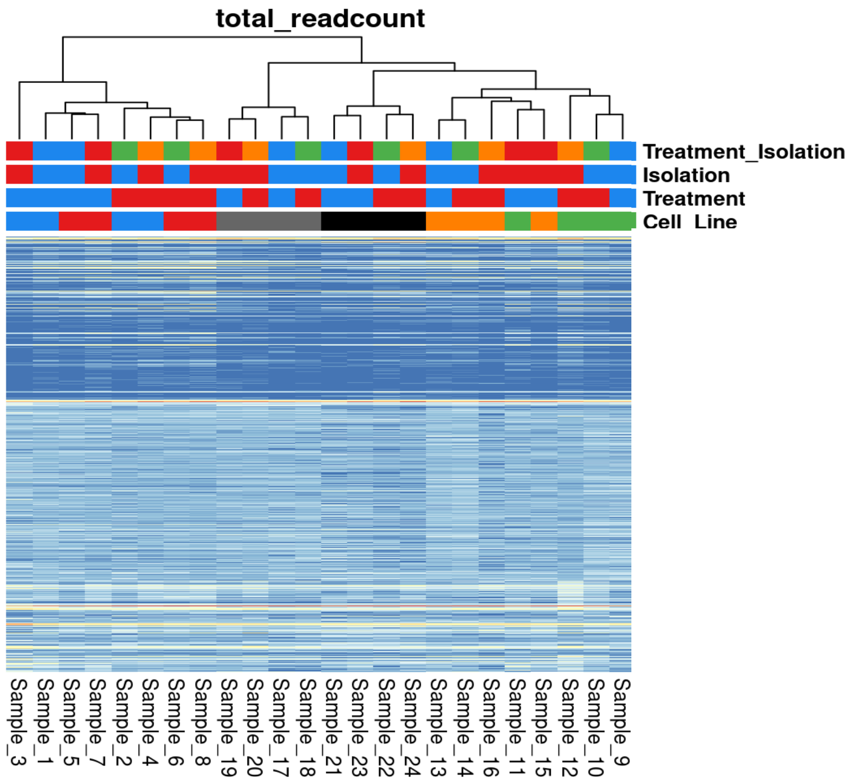


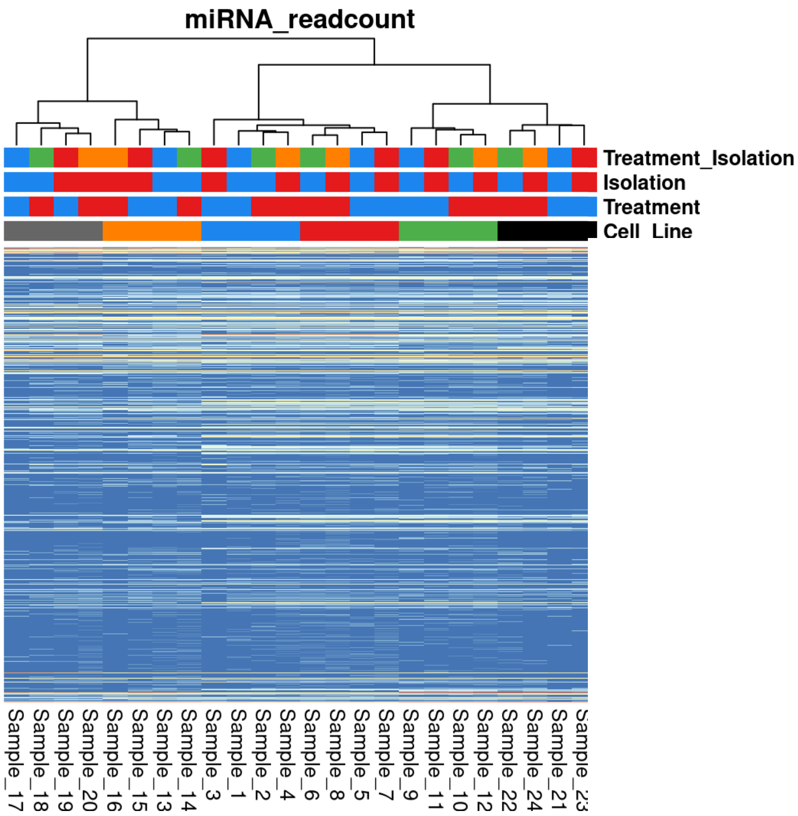
Median gene expression normalization is skewed by observations



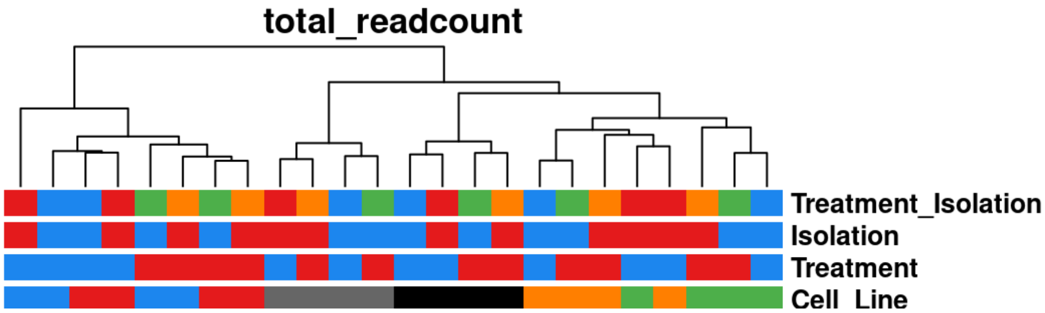
	total_readcount	miRNA_readcount	tRNA_miRNA_readcount
Sample_11	1.00	0.98	1.06
Sample_12	1.97	5.45	4.59
Sample_18	1.02	0.41	0.49

Normalization can be a major source of bias

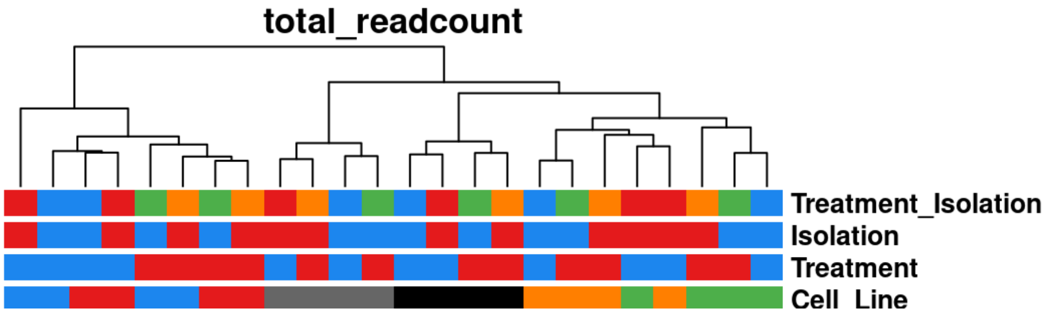


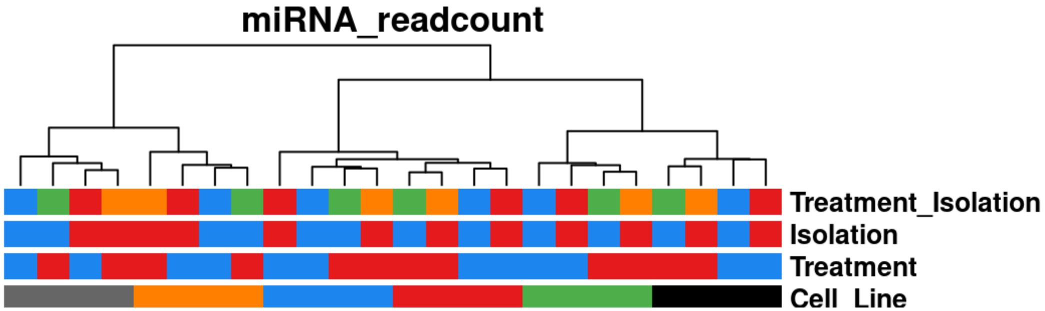


Normalization can be a major source of bias

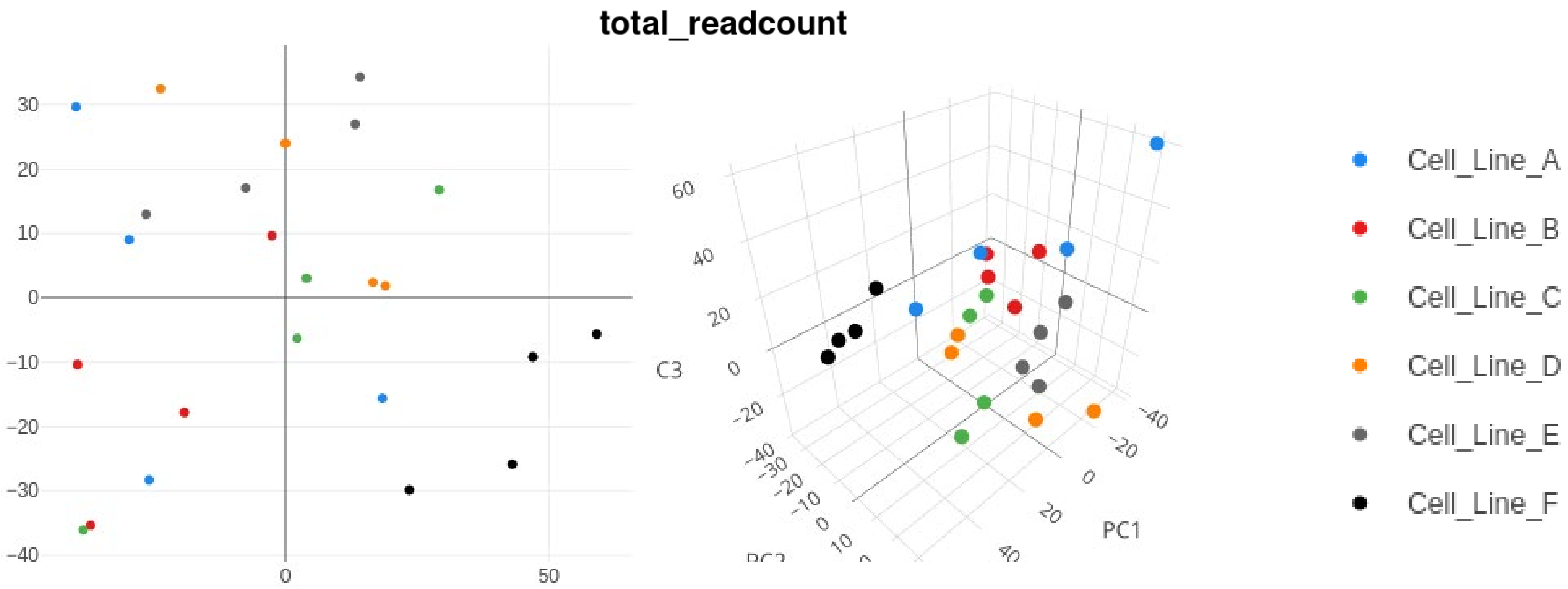


Normalization can be a major source of bias

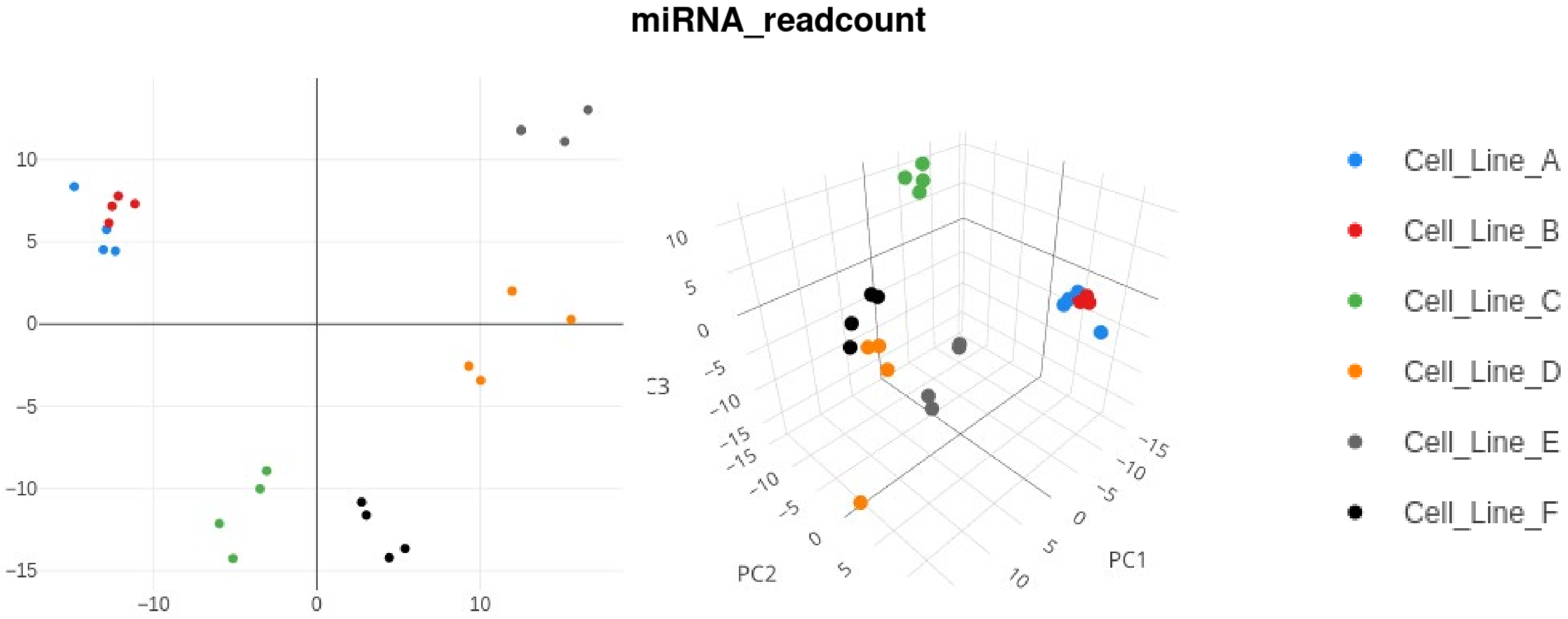




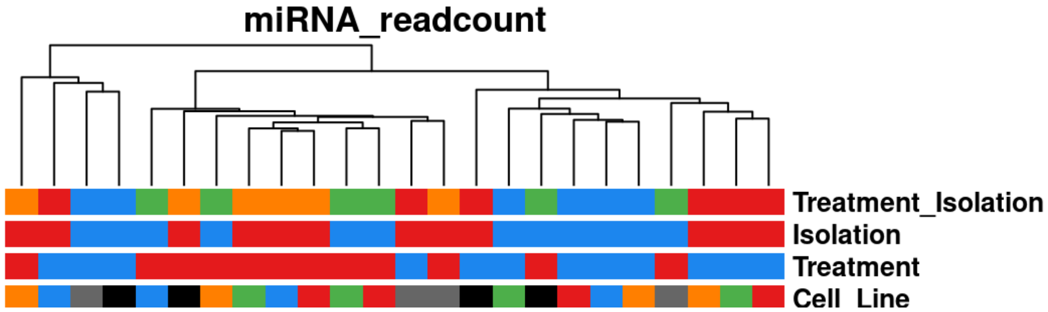
Correct normalization enables batch effect detection



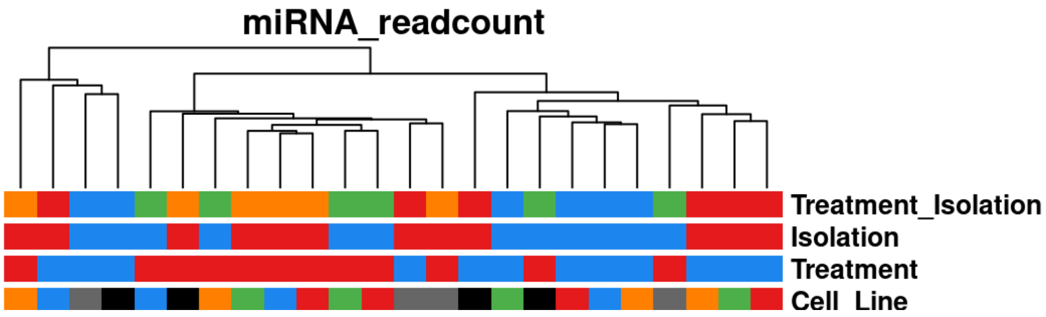
Correct normalization enables batch effect detection

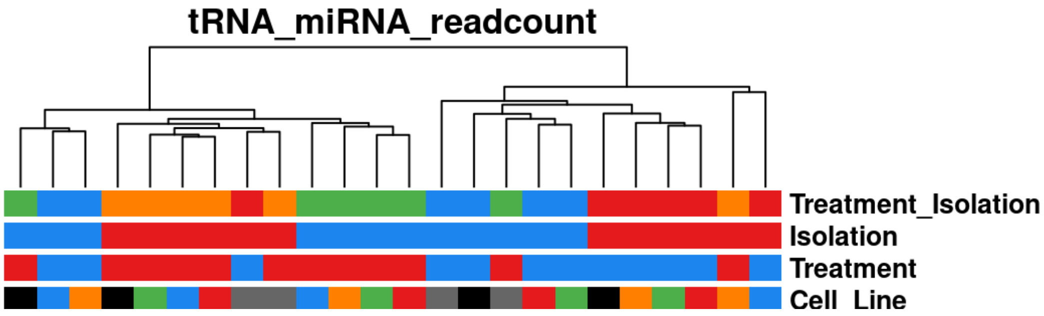


Exploit small RNA NGS beyond miRNAs

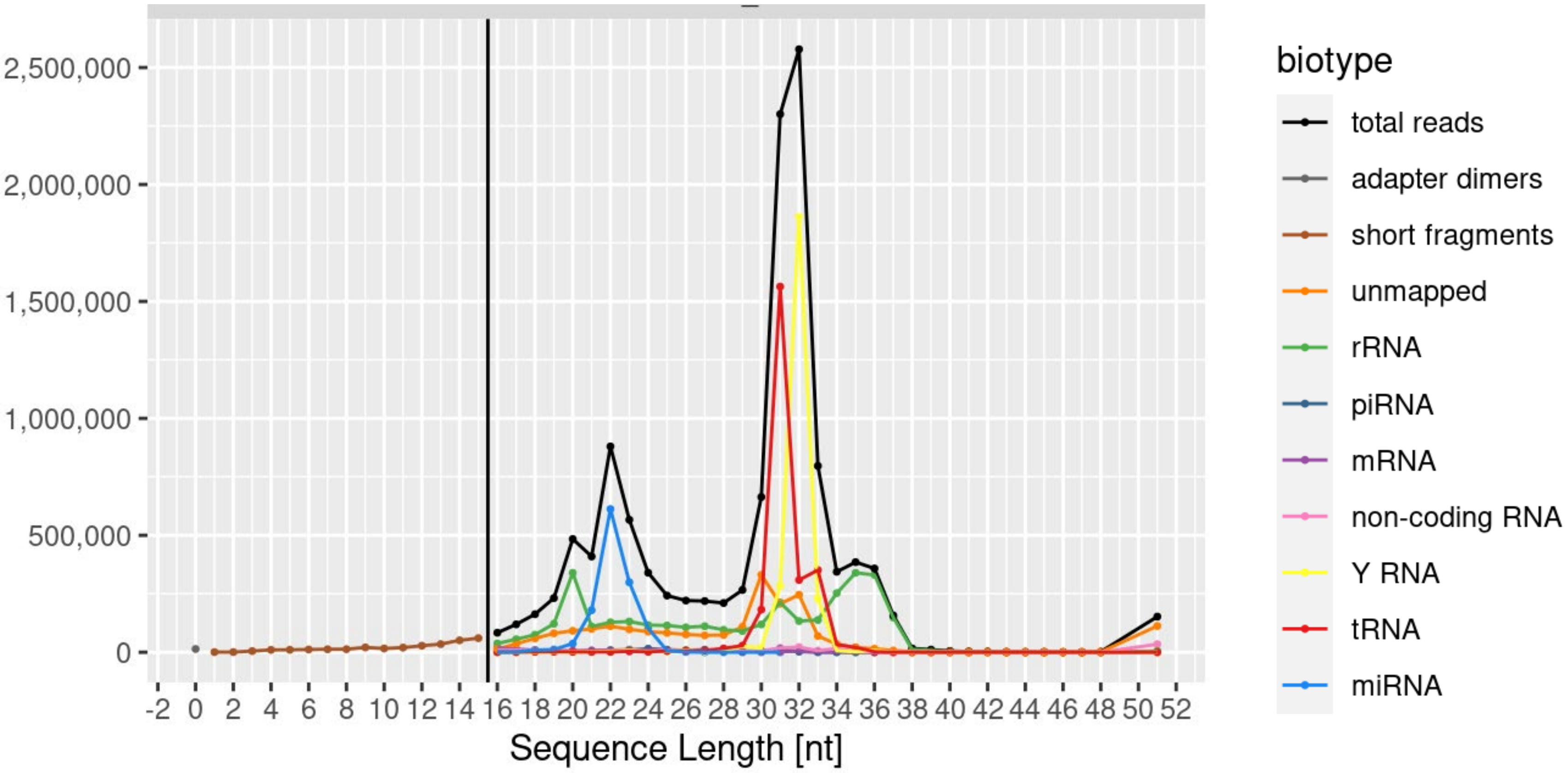


Improvement of normalization in small RNA NGS

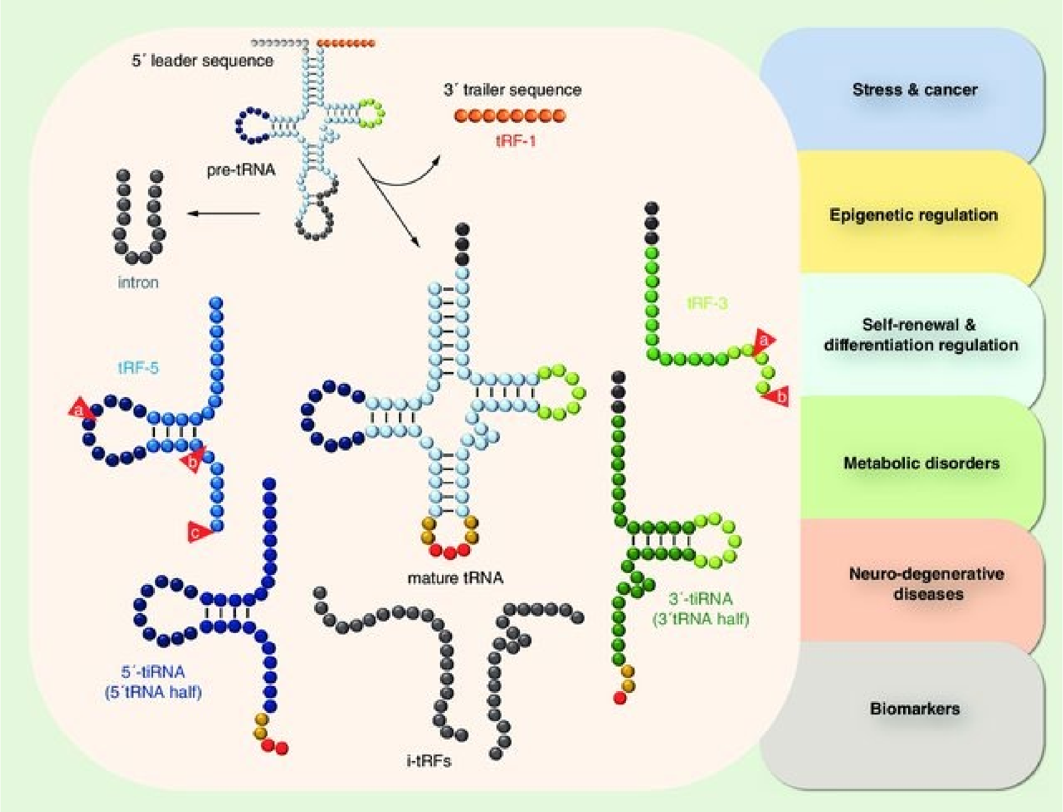




Degradation vs Specificity in small RNAs

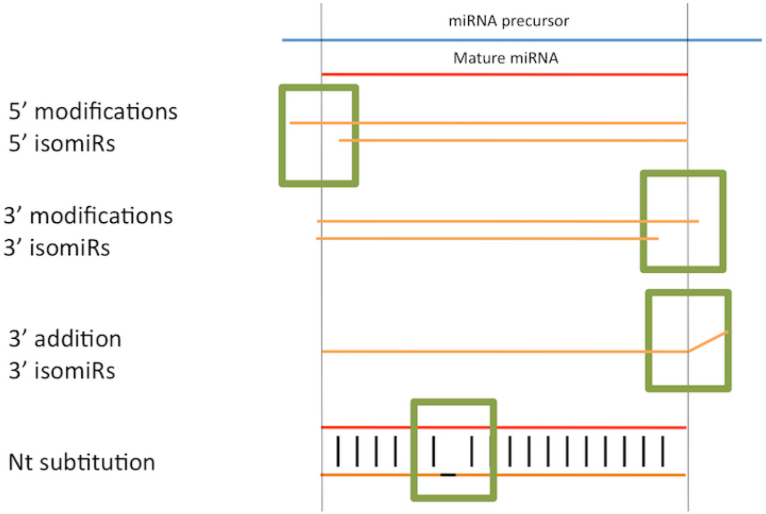


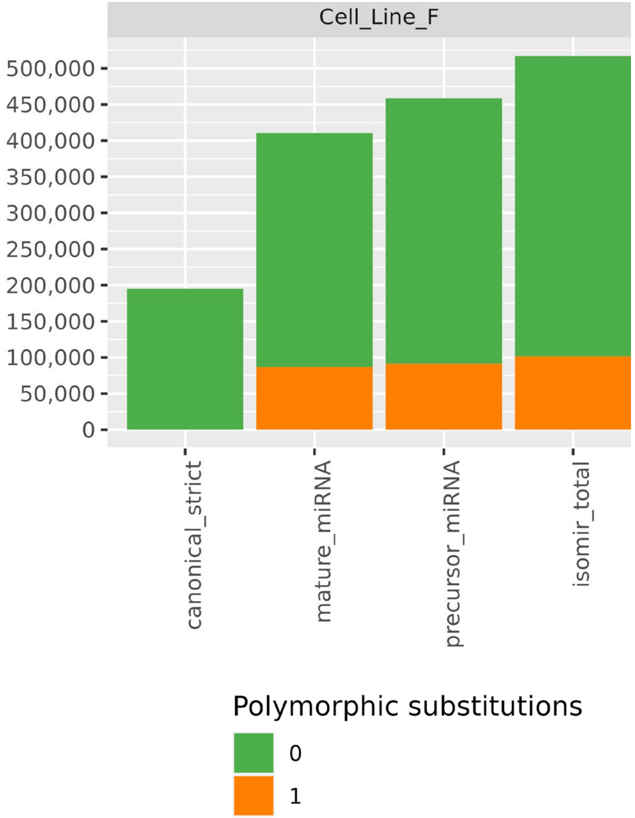
Hidden complexity in small RNA NGS - tRFs



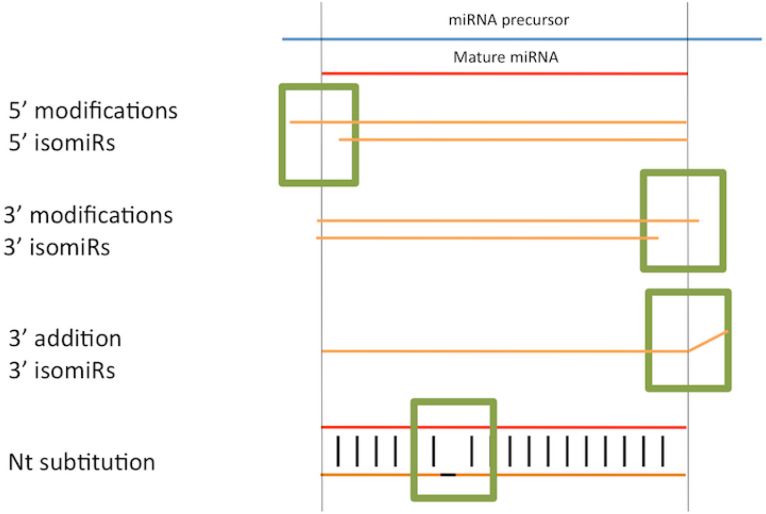
Grafanaki et al, Epigenomics 2018

Hidden complexity in small RNA NGS - isomiRs





Hidden complexity in small RNA NGS - isomiRs


miRNA precursor

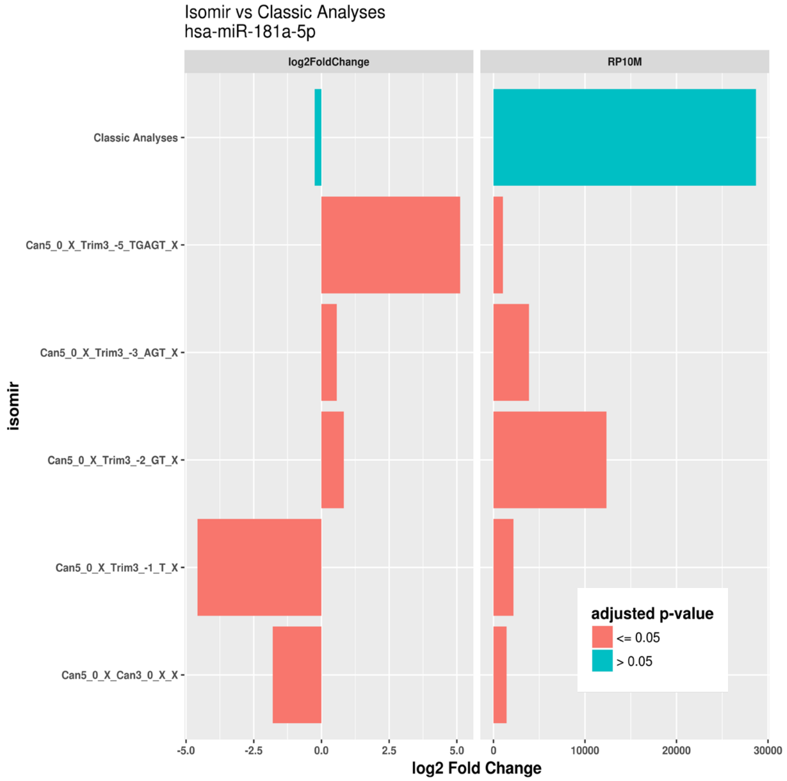
Mature miRNA

5’ modifications
5’ isomiRs

3’ modifications
3’ isomiRs

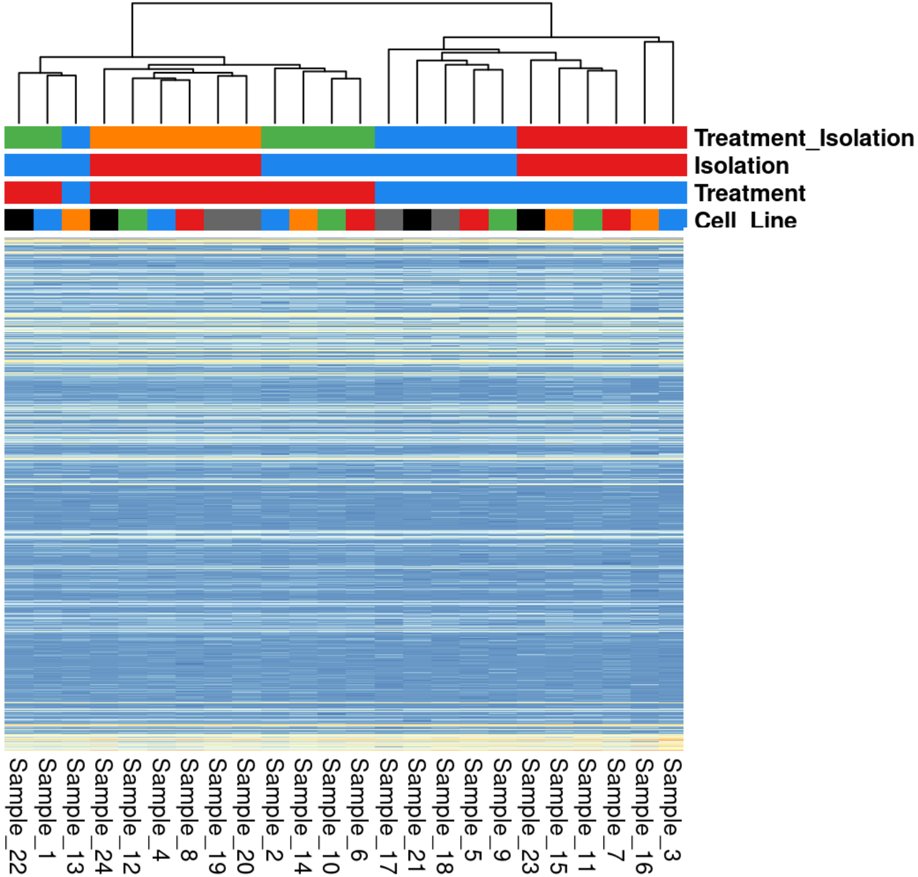
3’ addition
3’ isomiRs

Nt subtitution


Isomir vs Classic Analyses
hsa-miR-181a-5p

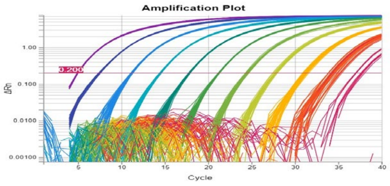
Comprehensive analysis of small RNA transcriptome increases efficiency of biomarker detection

tRFs + isomiRs



Conclusions

- Importance of spike-ins in small RNA NGS
- Comprehensive characterization of full transcriptome necessary for correct normalization and batch effect detection
- Detection of functional fragments and isoforms increases specificity in classification and biomarker profiles

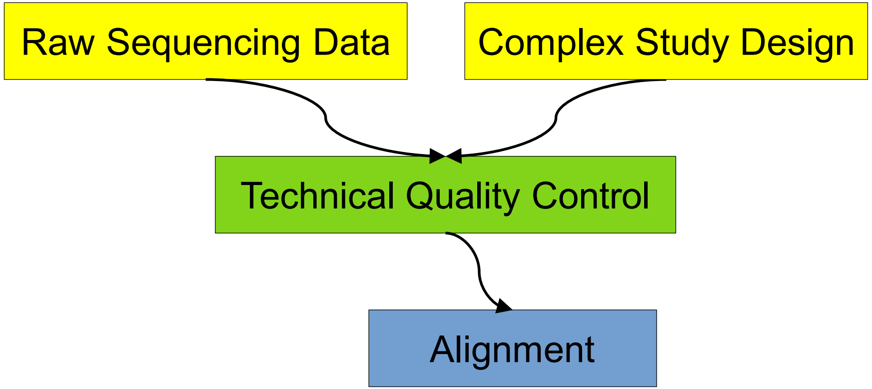


≠

- Detection of stable reference transcripts in small RNA NGS for biomarker validation

caRNAge

Comprehensive **A**nalysis of small **RNA** **G**ene **E**xpression



Raw Sequencing Data

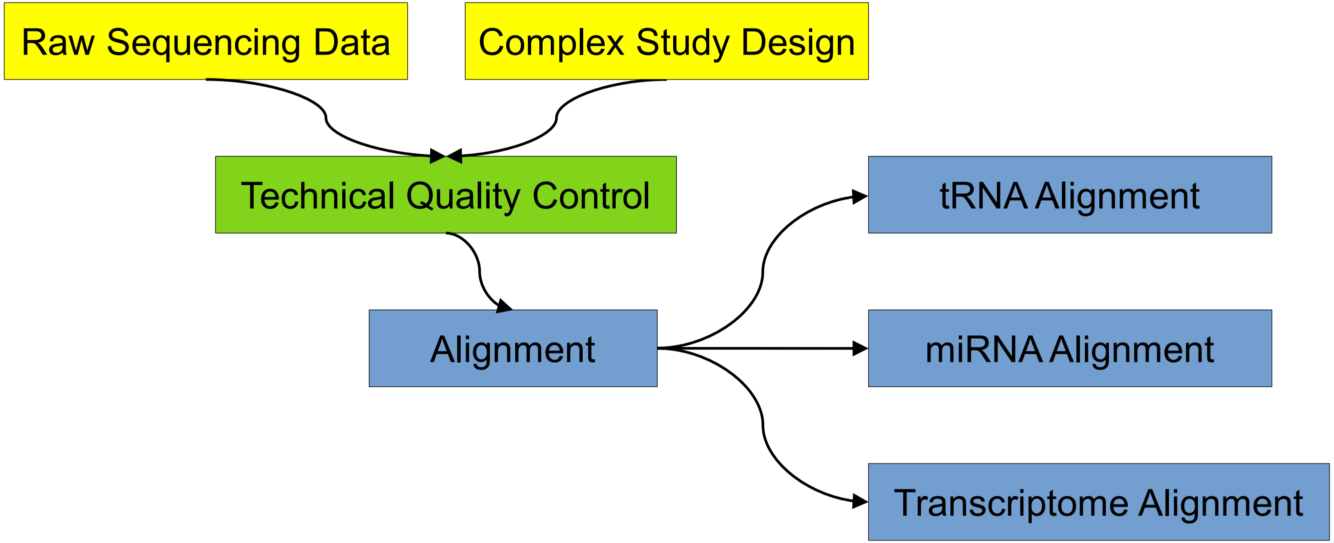
Complex Study Design

Technical Quality Control

Alignment

caRNAge

Comprehensive **A**nalysis of small **RNA** **G**ene **E**xpression


Raw Sequencing Data

Complex Study Design

Technical Quality Control

Alignment

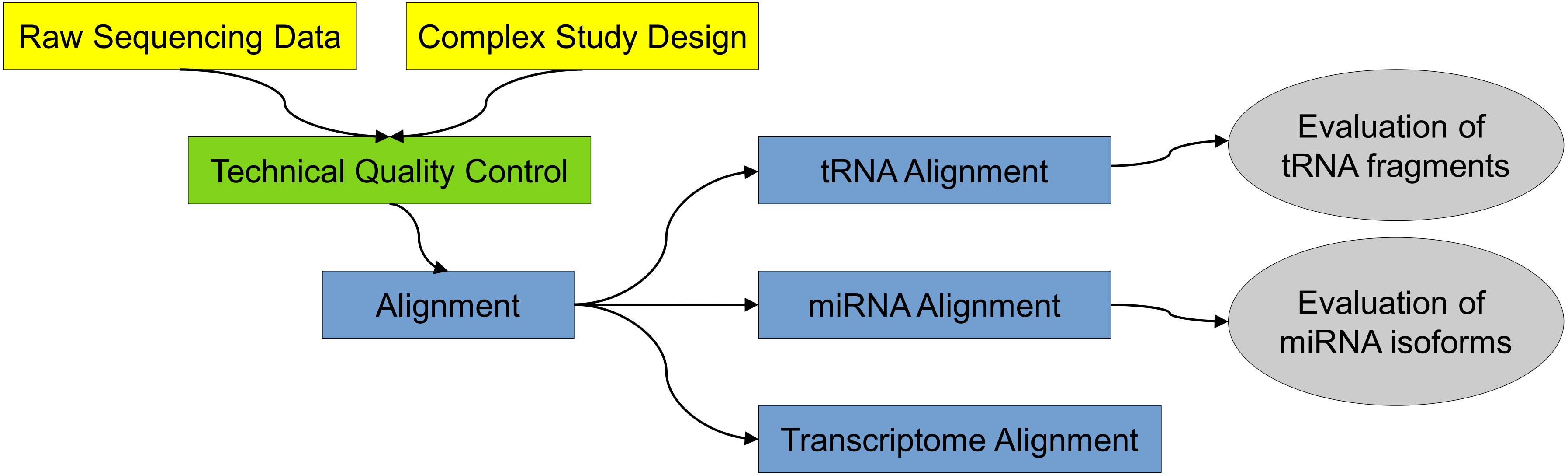
tRNA Alignment

miRNA Alignment

Transcriptome Alignment

caRNAge

Comprehensive **A**nalysis of small **RNA** **G**ene **E**xpression



Raw Sequencing Data

Complex Study Design

Technical Quality Control

Alignment

tRNA Alignment

miRNA Alignment

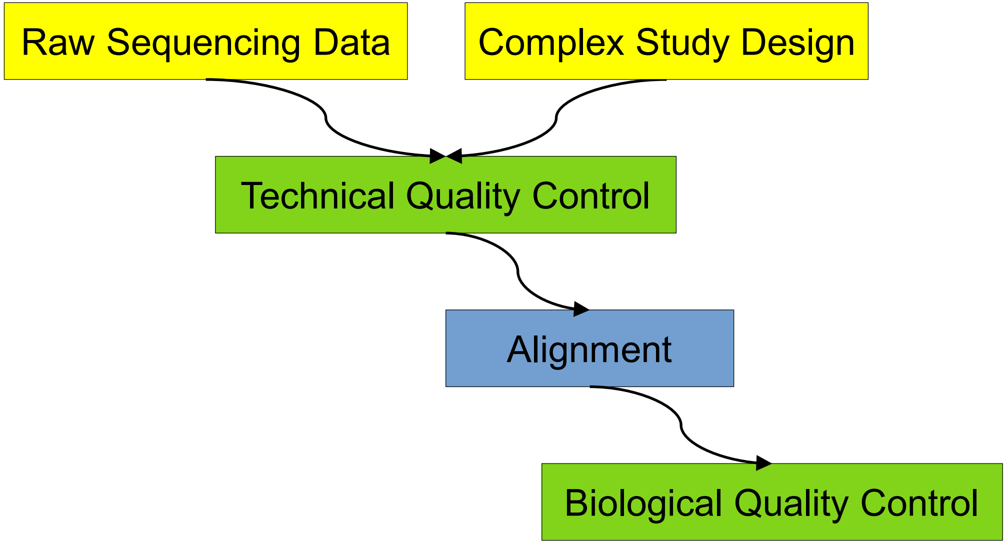
Transcriptome Alignment

Evaluation of tRNA fragments

Evaluation of miRNA isoforms

caRNAge

Comprehensive **A**nalysis of small **RNA** **G**ene **E**xpression



Raw Sequencing Data

Complex Study Design

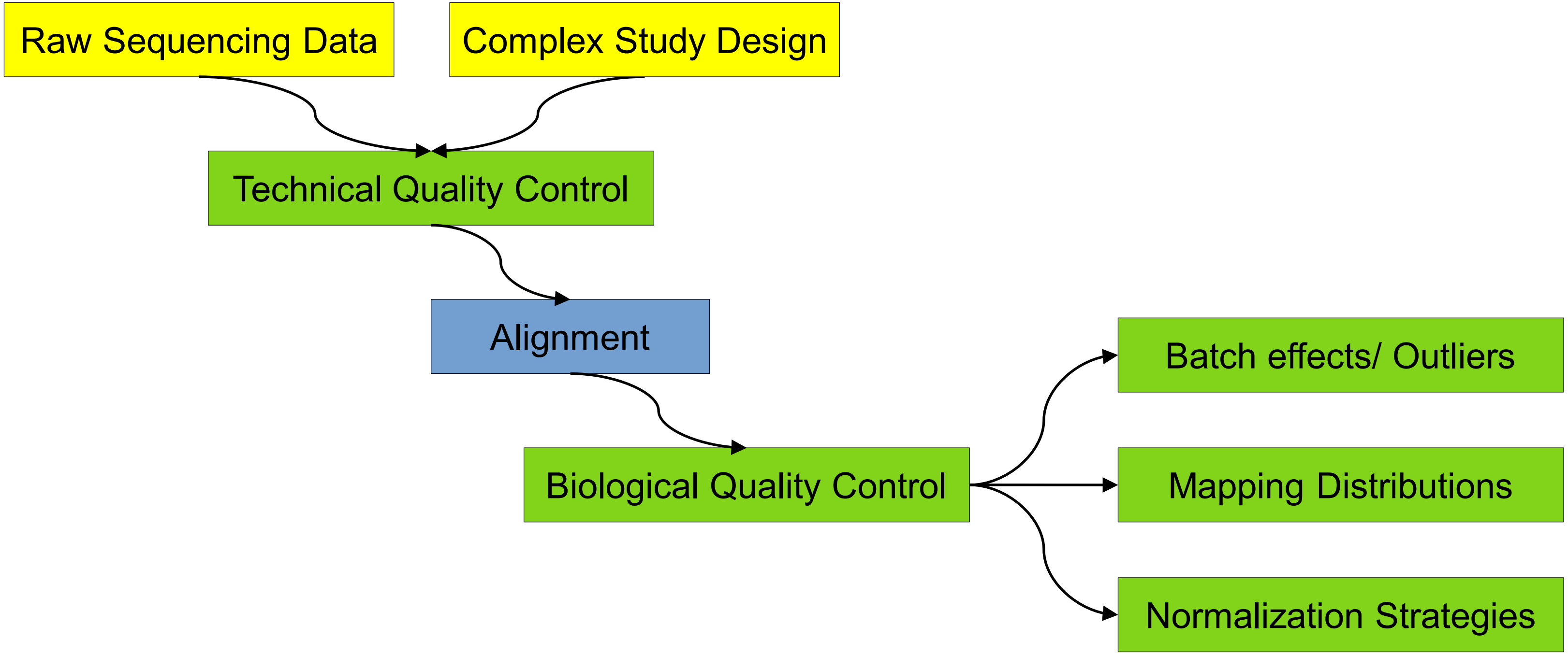
Technical Quality Control

Alignment

Biological Quality Control

caRNAge

Comprehensive **A**nalysis of small **RNA** **G**ene **E**xpression



Raw Sequencing Data

Complex Study Design

Technical Quality Control

Alignment

Biological Quality Control

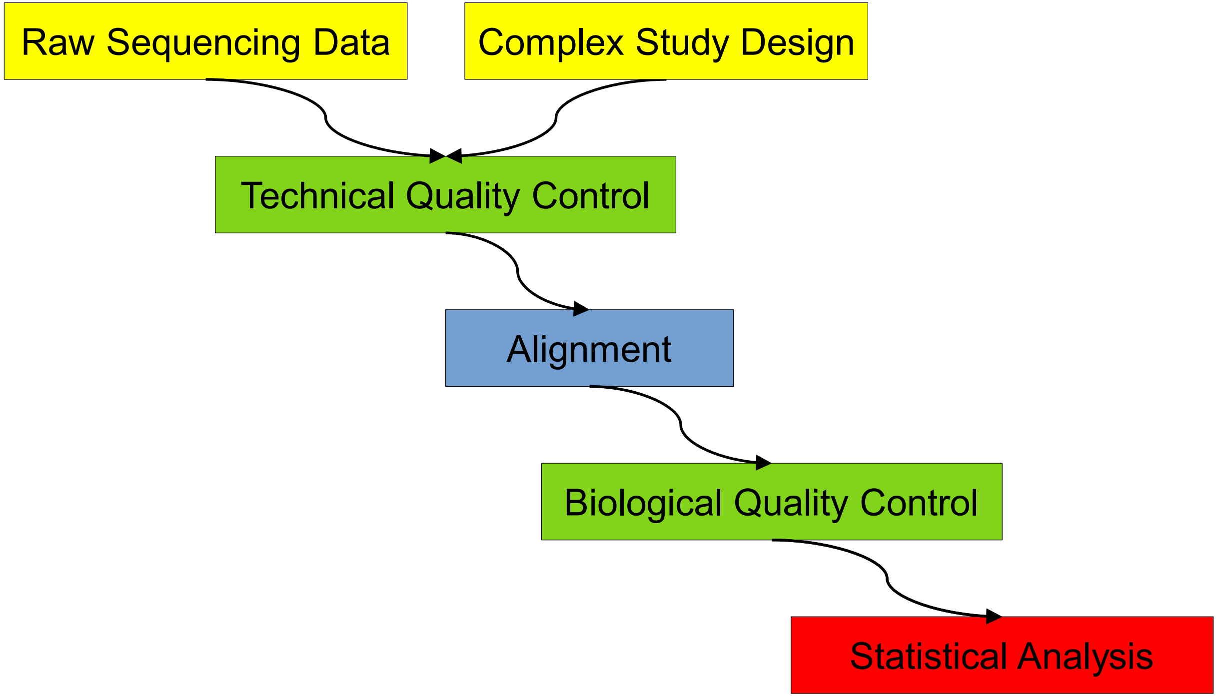
Batch effects/ Outliers

Mapping Distributions

Normalization Strategies

caRNAge

Comprehensive **A**nalysis of small **RNA** **G**ene **E**xpression



Raw Sequencing Data

Complex Study Design

Technical Quality Control

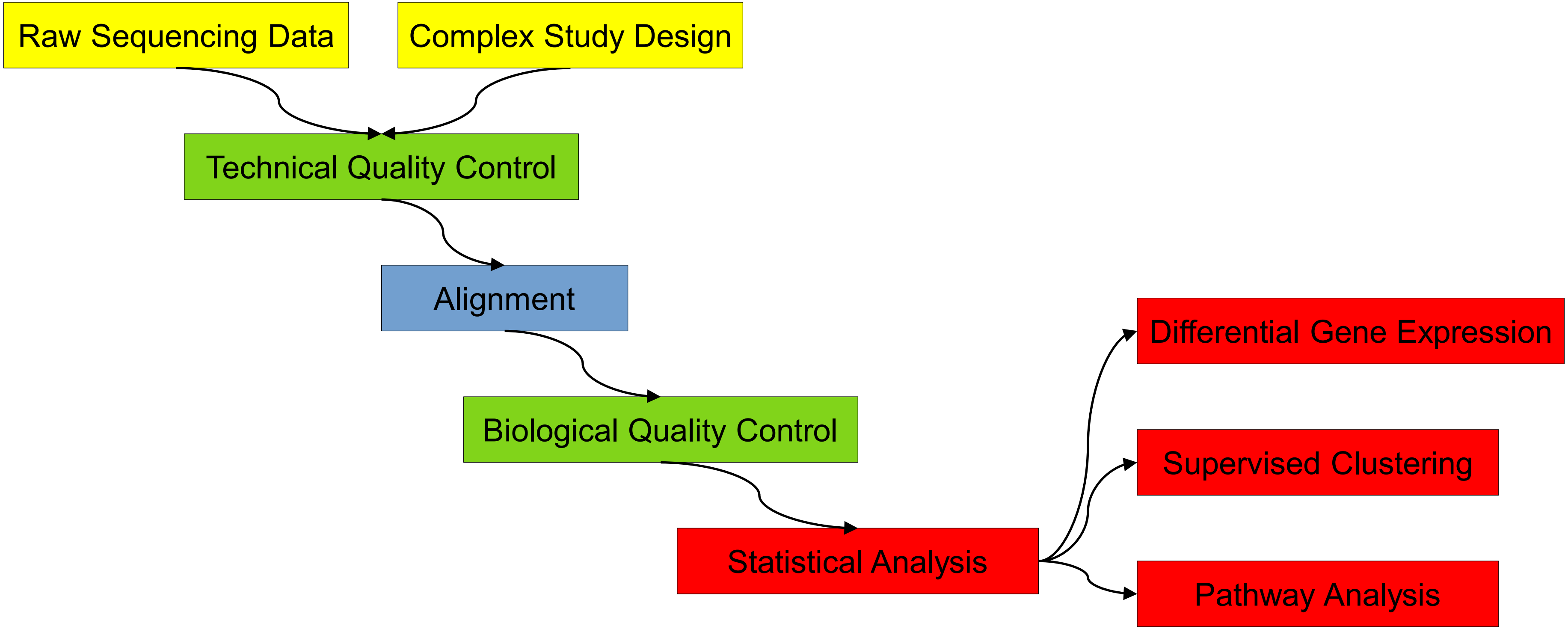
Alignment

Biological Quality Control

Statistical Analysis

caRNAge

Comprehensive **A**nalysis of small **RNA** **G**ene **E**xpression



Raw Sequencing Data

Complex Study Design

Technical Quality Control

Alignment

Biological Quality Control

Statistical Analysis

Differential Gene Expression

Supervised Clustering

Pathway Analysis

Output

Quality Control

- Detection of degradational status
- Detection of sequencing errors
- Detection of batch effects
- Detection of outliers
- Relative distribution of all RNA species
- Length distribution of alignments

Differential gene expression

- Normalization according to specified RNA species
- Statistical evaluation of gene expression changes
- Pairwise comparisons of all specified experimental groups
- Evaluation of stable reference transcripts for qPCR validation

microRNA and isomiR analysis

- Integrative differential transcript analysis between predominant isomiRs and parent miRNAs
- Analysis of miRNA clusters
- Detection of miRNA localization motifs
- Evaluation of miRNA stability by isomiR modifications

Pathway and gene list analysis

- Identification of predicted and experimentally validated targets of mature miRNAs
- Overrepresentation analysis of significantly regulated transcripts
  - Gene Ontology
  - KEGG
  - Reactome
  - WikiPathways

Supervised clustering

- Detection of optimal error rates
- Sparse projection on latent squares (sPLS)
- Classification based on biomarker profiles

Acknowledgements

- Division of Animal Physiology and Immunology, TUM
  - Prof. Michael Pfaffl
- Research Group Big Data in Biomedicine, TUM
  - Johannes Kersting
- Institute of Human Genetics, LMU
  - Marlene Reithmair
  - Martina Schuster

Test for yourself!

https://www.physio.wzw.tum.de/caRNAge/