Experimental design in RNA-seq

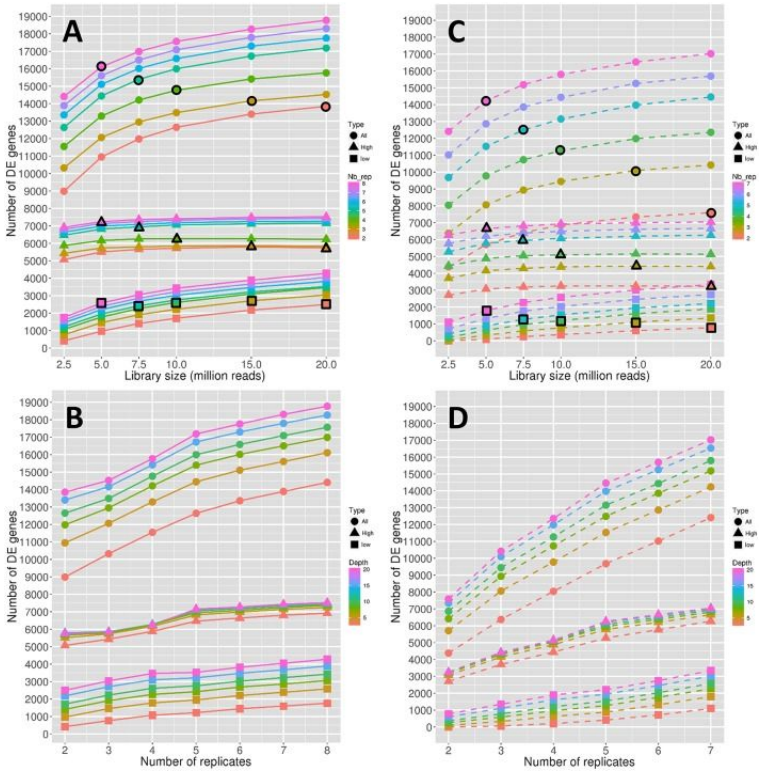
05.12.20

Parameters

- Type of library
- Number of replicates
- Coverage
- Money

Effect of number of replicates and sequencing depth

Both number of replicates and sequencing depth have an impact on the number of found DE genes

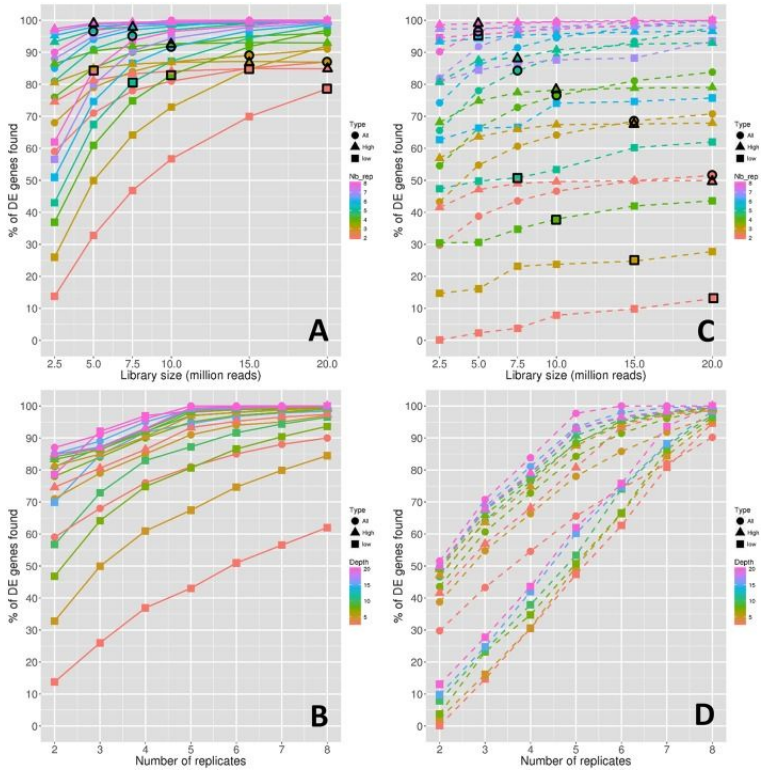


10.3389/fpls.2018.00108

Power curves for replicates and depth

More DE genes are found for highly expressed genes

Percentage of found DE genes grows slower after 5-10M reads per library

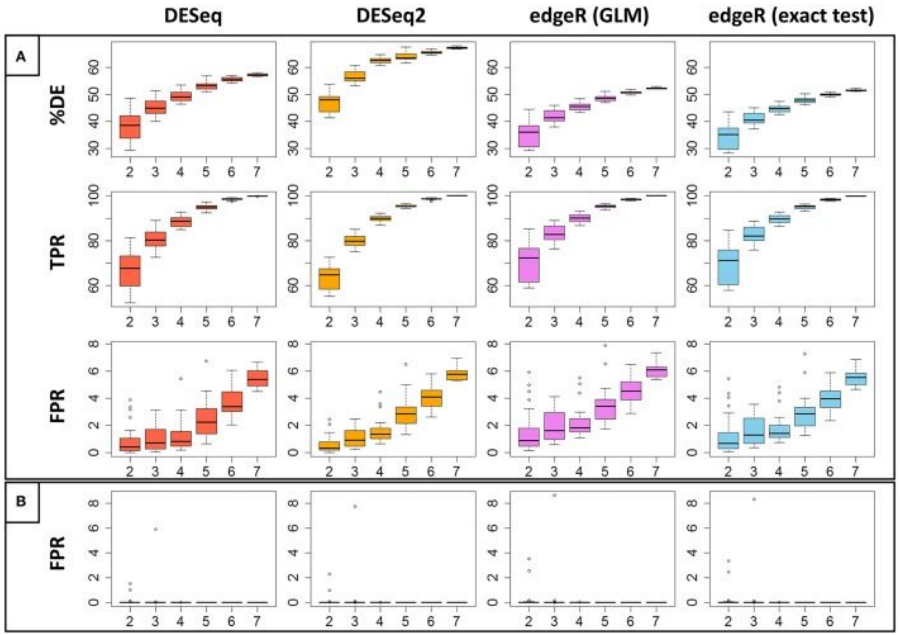


10.3389/fpls.2018.00108

TPR and FPR for different tools

TPR increases with number of replicates

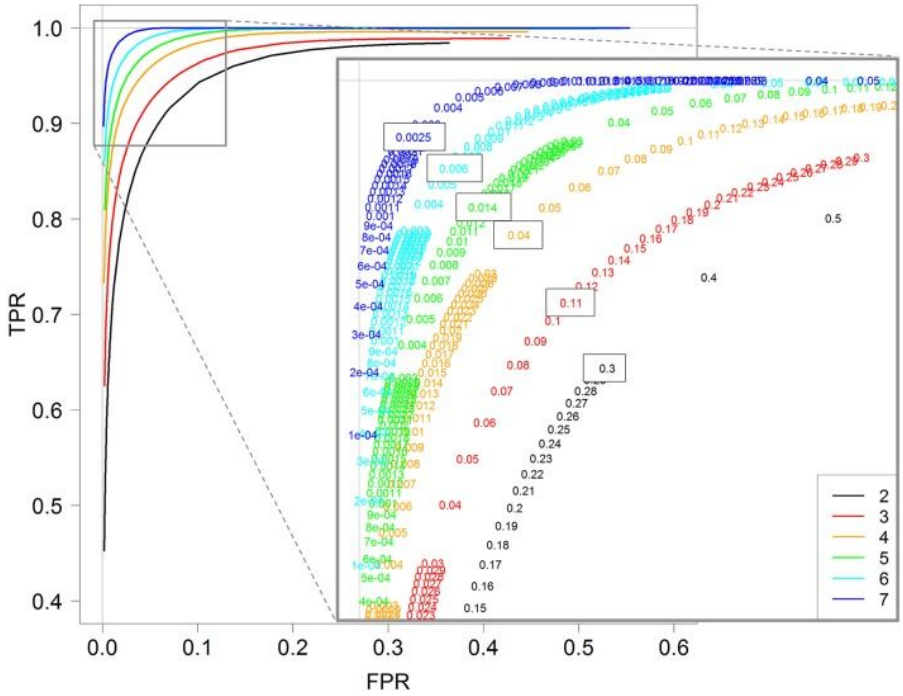
FPR also increases



10.3389/fpls.2018.00108

ROC curves for different number of replicates

Optimal threshold depends on the number of replicates

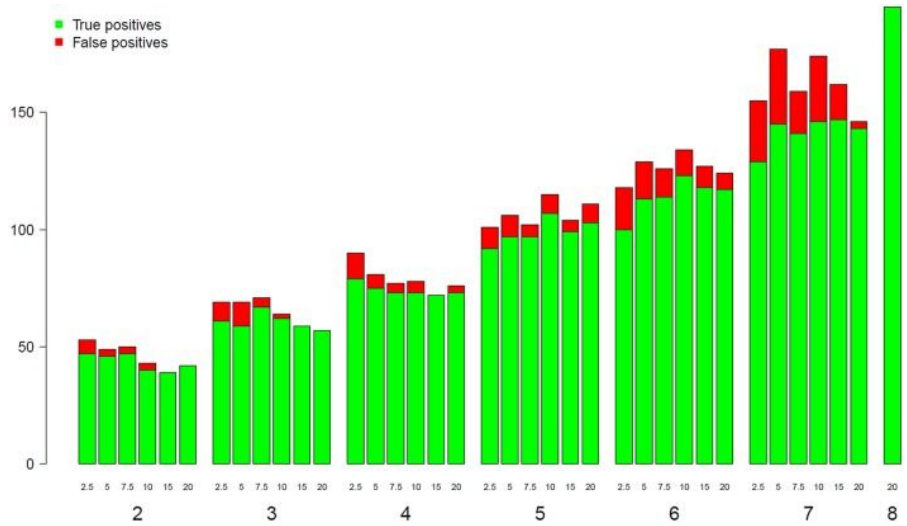


10.3389/fpls.2018.00108

GO enrichment and replicates

Number of replicates increases true positive GO enrichments

Sequencing depth doesn't affect the results



10.3389/fpls.2018.00108

Recommendations

At least four replicates per condition with 20M reads (depends on the organism) to get 1000 DE genes if they exist at all

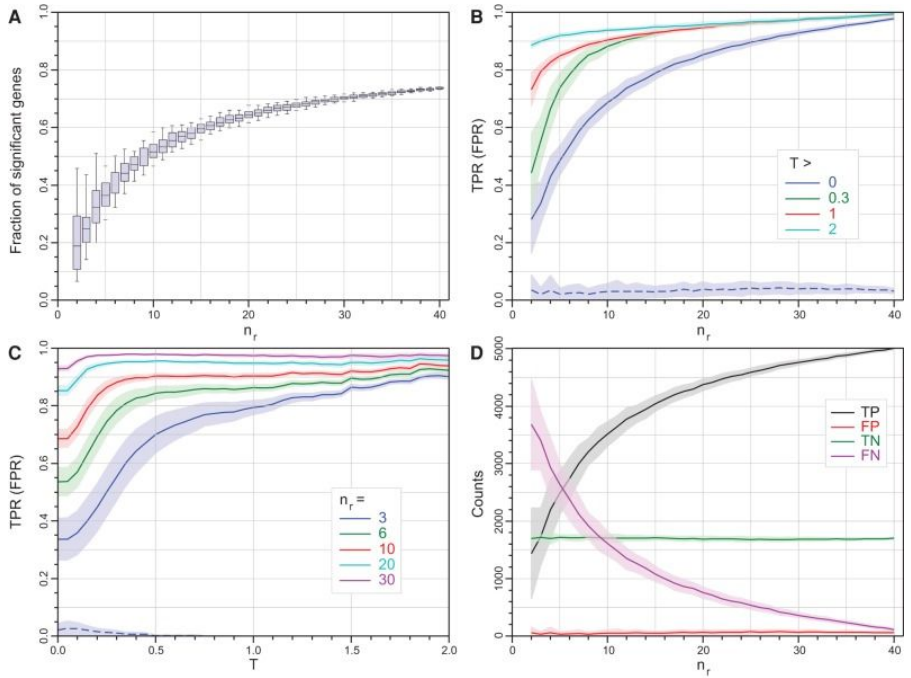
FDR threshold for DE ~ 2^{-r} where r - number of replicates

DESeq, DESeq2, edgeR give roughly the same results

10.3389/fpls.2018.00108

Effect of number of replicates

Three replicates are enough to find 20-40% of true DE genes



10.1261/rna.053959.115

Recommendations

At least six replicates per condition for all experiments

At least 12 replicates per condition for experiments where identifying the majority of all DE genes is important

For experiments with <12 replicates per condition; use edgeR (exact) or DESeq2

For experiments with >12 replicates per condition; use DESeq

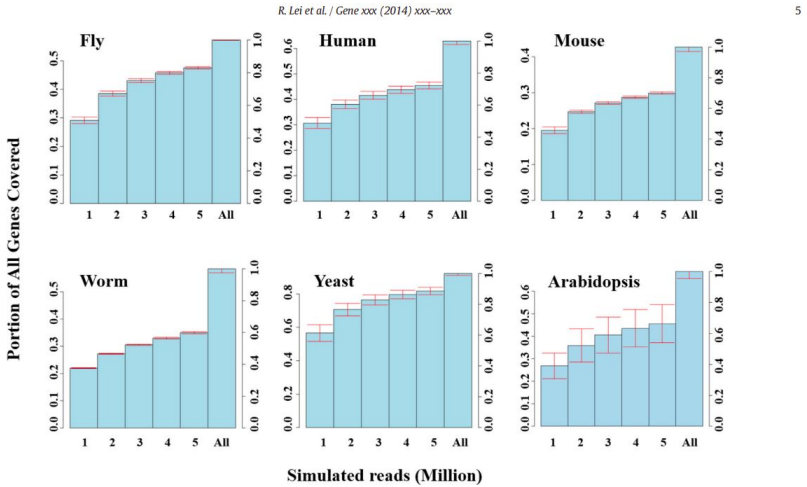
Apply a fold-change threshold appropriate to the number of replicates per condition between $0.1 \leq T \leq 0.5$

10.1261/rna.053959.115

Sequencing depth impact on found DE genes

Even 1M reads are enough to pick up highly expressed genes

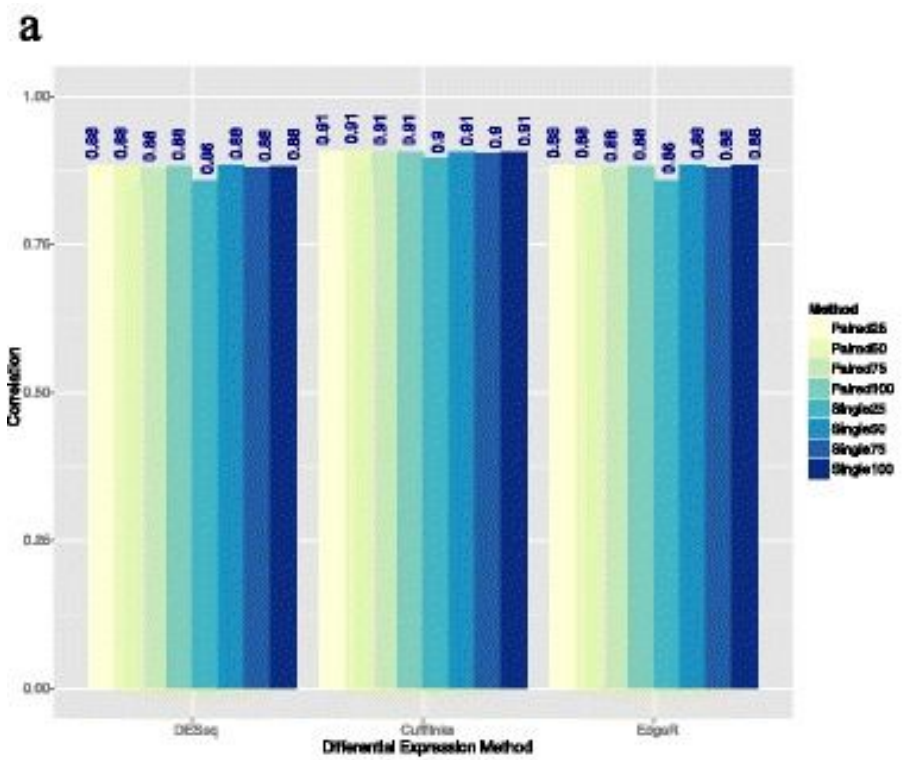
Genes with low expression are mostly uncovered even with extreme (>100M) library sizes



10.1016/j.gene.2014.12.013

Read length effect

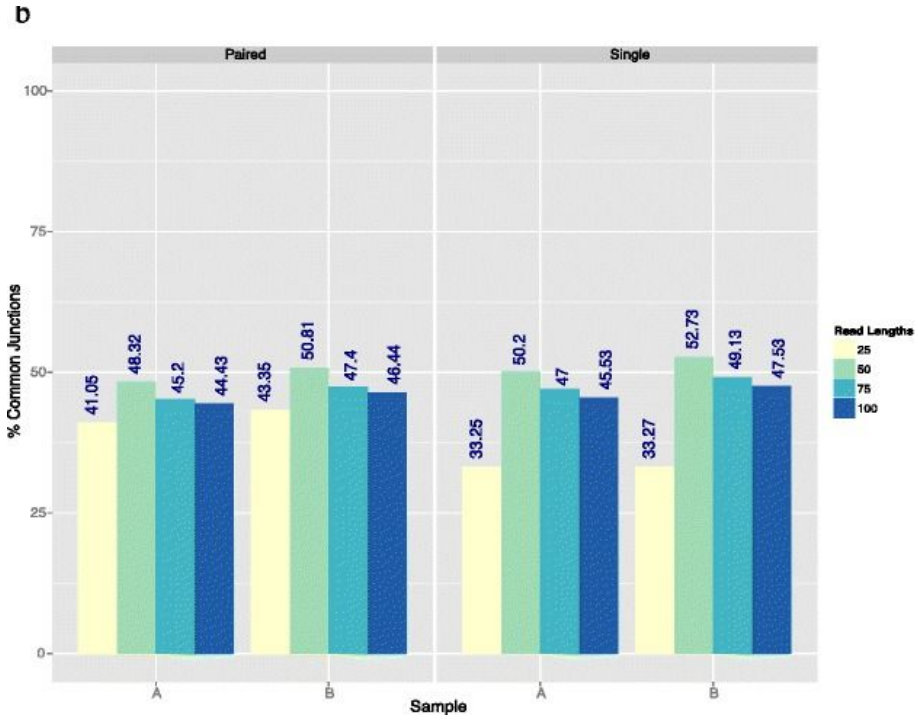
Any length above 25 doesn’t add much to the analysis of DE genes



10.1186/s13059-015-0697-y

Read length effect

Any read length above 50 can be used to detect splice junctions



10.1186/s13059-015-0697-y

Paired-end vs single-end

Single

Differential gene expression

Paired

Splice-site and isoform detection

Differential exon usage

De novo transcript annotation

Fusion transcripts

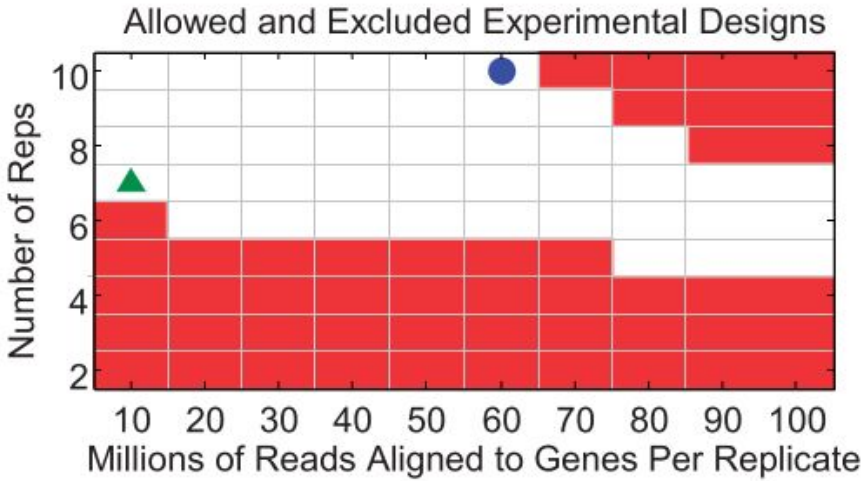
Number of replicates > Library size > Read length > Pairedness

Automated experiment design

Calculates the depth and number of reps

Can also compute cost

Works better with a small-scale pilot study



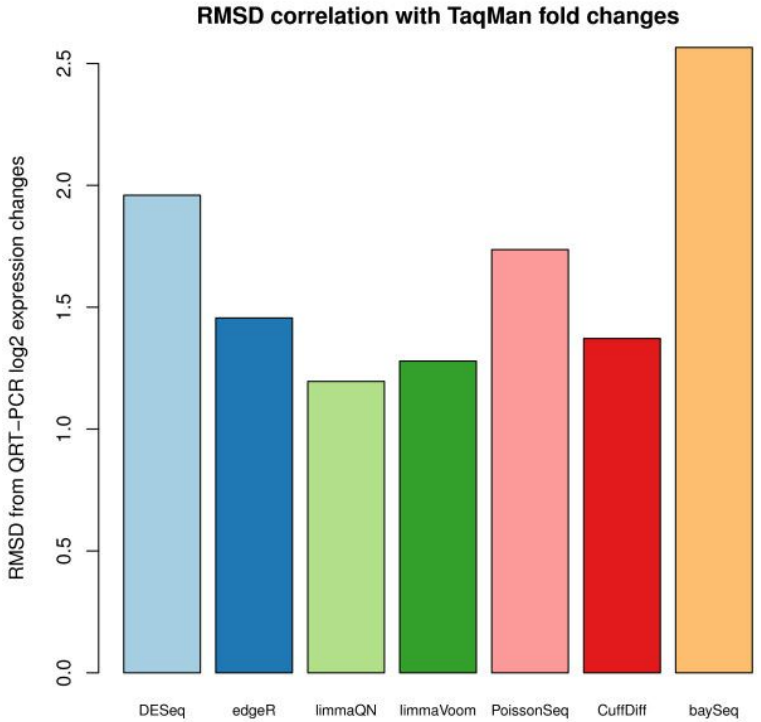
http://scotty.genetics.utah.edu/

DE tools comparison

All tools are reasonably good

Microarray tools adapted for RNA-seq work fine

Number of replicates is the most important parameter for DE



10.1186/gb-2013-14-9-r95

Sequence costs

~1800 euro per one Illumina lane = ~400M reads

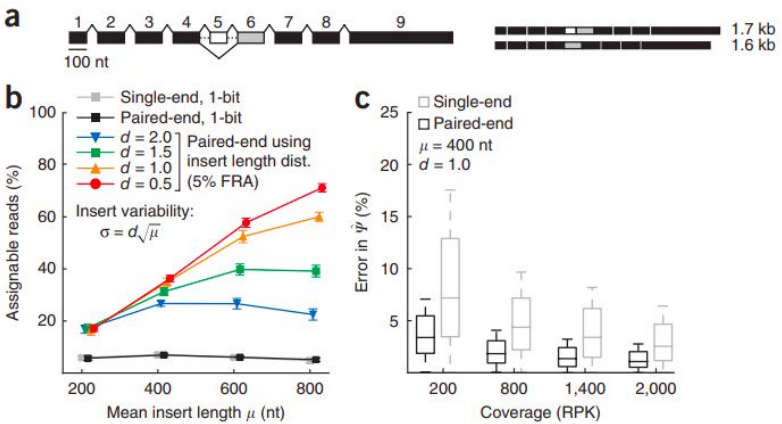
Transportation and sample preparation may add something

Cost calculator for single-cell RNA-seq

https://satijalab.org/costpercell

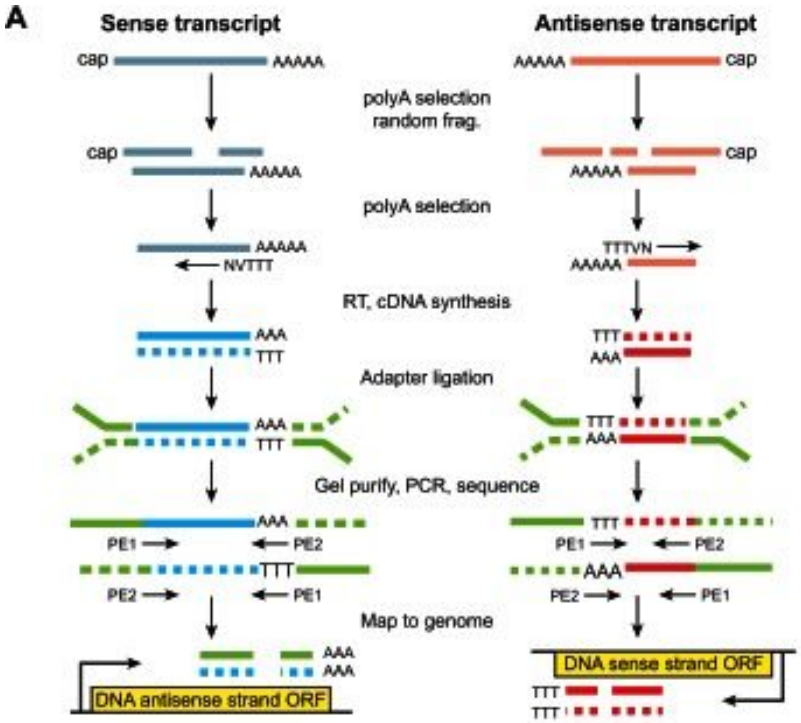
Paired-end for finding isoforms

Paired-end reads give improve assigning reads to different isoforms



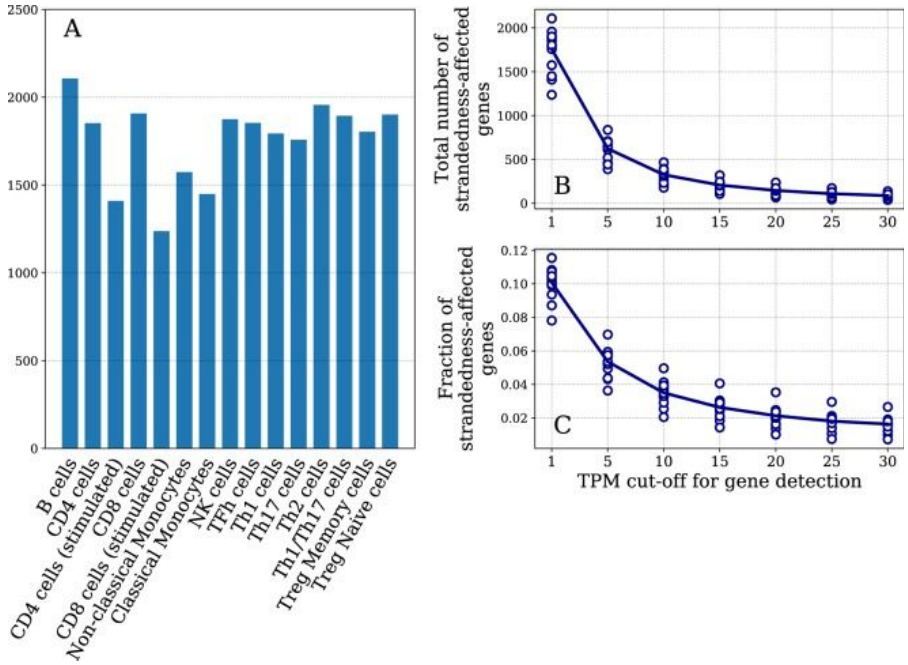
10.1038/nmeth.1528

Stranded libraries



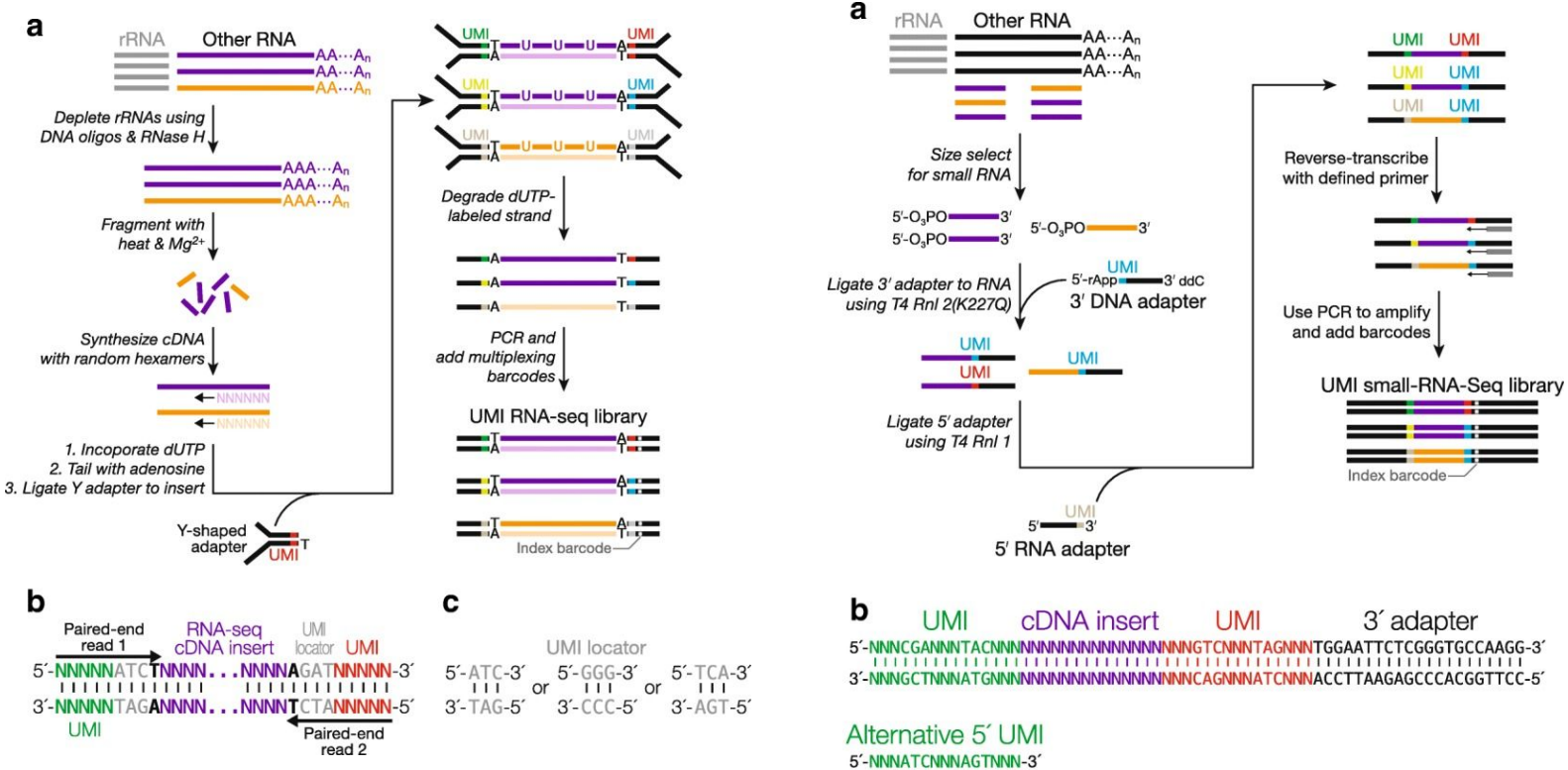
Unstranded libraries may bias results of DE

10% of all genes and 2.5% of protein coding genes have a two-fold or higher difference in estimated expression when strand information of the reads was ignored



10.1038/s41598-019-52584-w

Unique molecular identifiers (UMI)



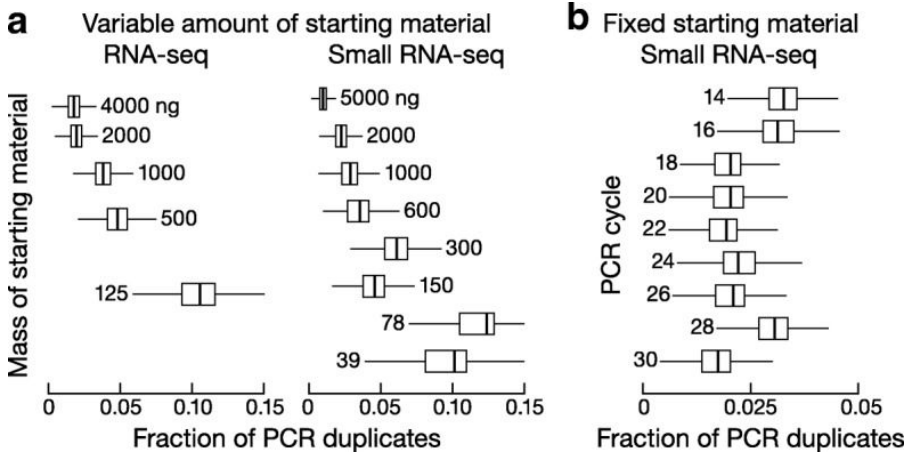
10.1186/s12864-018-4933-1

UMI against PCR duplicates

PCR duplicates do not only depend on PCR cycle number but also on the initial amount of sequencing material

Using UMI improves finding of PCR duplicates

Removing PCR duplicates based on coordinates biases results against highly abundant genes

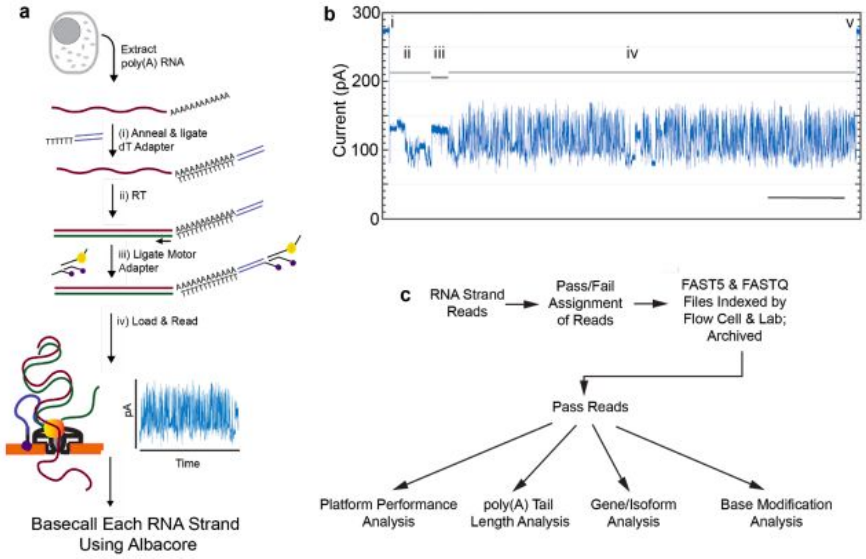


10.1186/s12864-018-4933-1

Nanopore RNA sequencing

Allows some unusual analyses

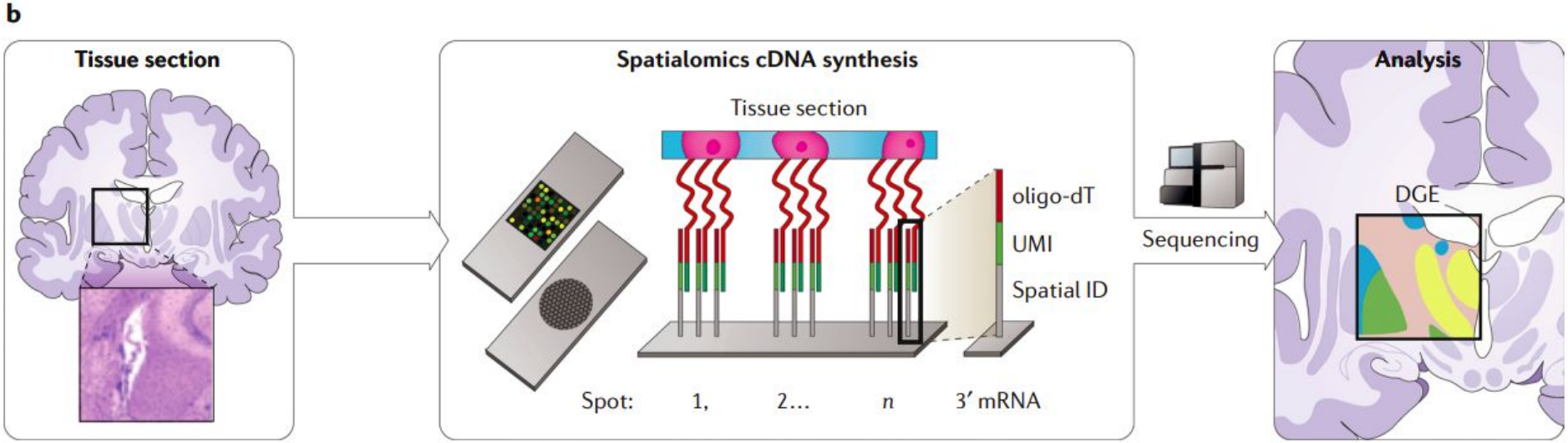
Does not require reverse transcription



10.1038/s41592-019-0617-2

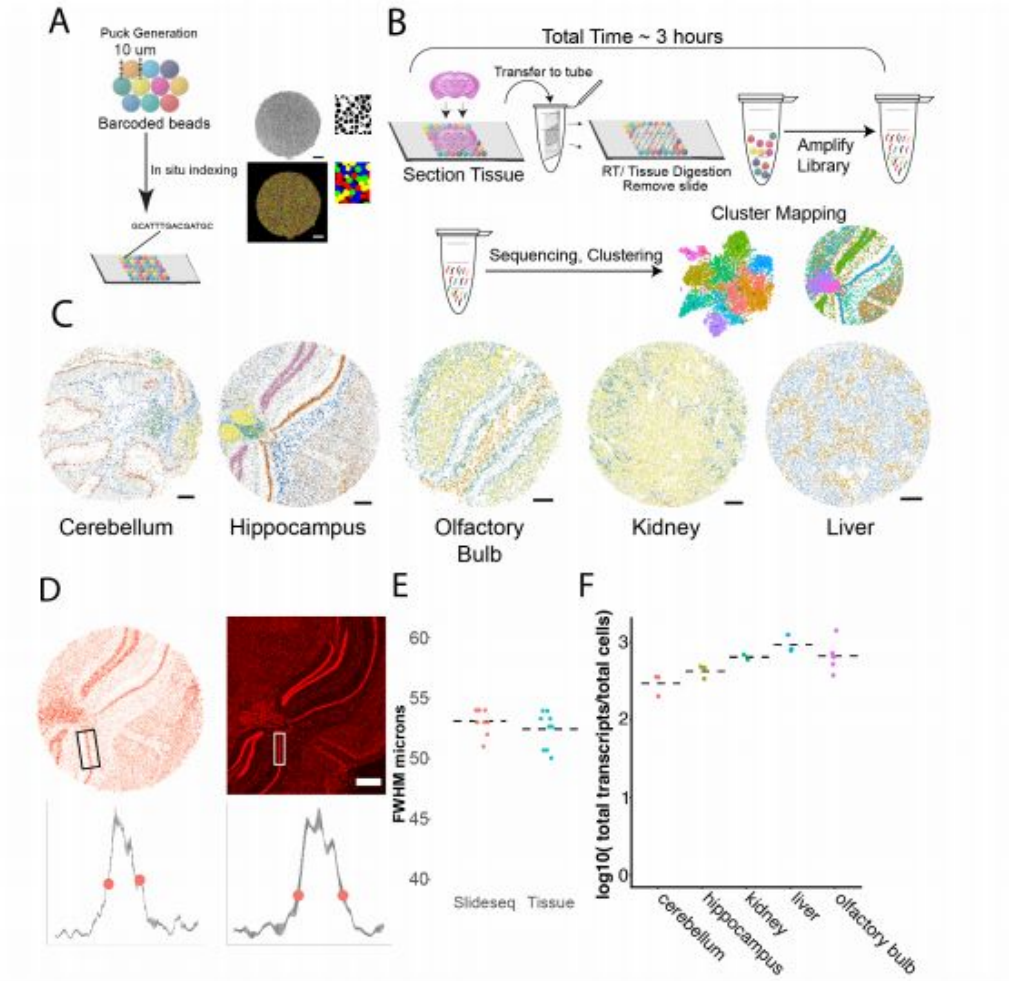
Unusual types of RNA-seq

Spatial RNAseq

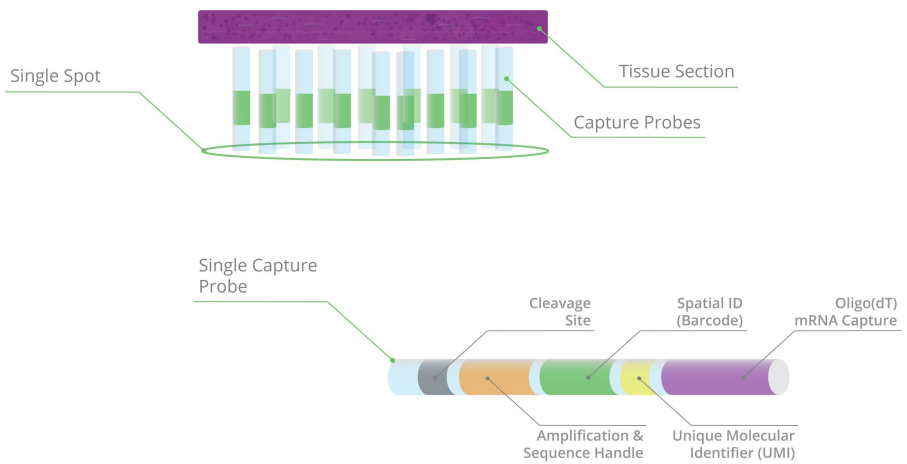


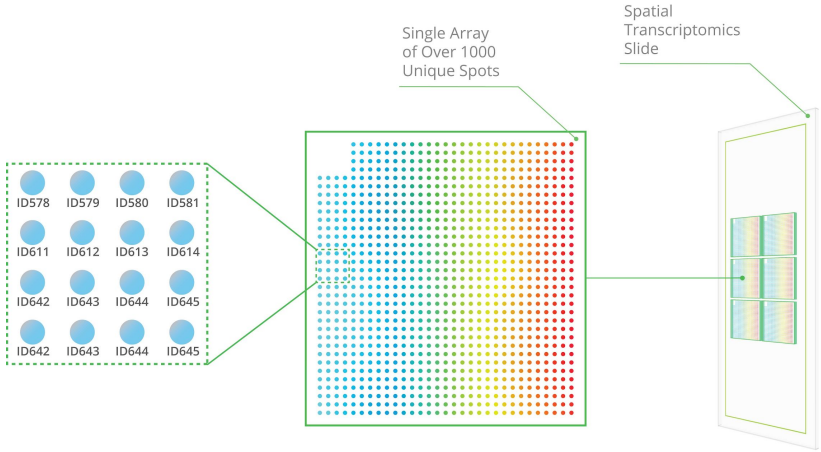
10.1038/s41576-019-0150-2

Slide-seq


A
Puck Generation
10 um
Barcoded beads
In situ indexing
B
Total Time ~ 3 hours
Transfer to tube
Section Tissue
RT/ Tissue Digestion
Remove slide
Amplify
Library
Cluster Mapping
Sequencing, Clustering
C
Cerebellum
Hippocampus
Olfactory
Bulb
Kidney
Liver
D
E
FWHM microns
60
55
50
45
40
Slideseq
Tissue
F
log10(total transcripts/total cells)
3
2
1
0
cerebellum
hippocampus
kidney
liver
olfactory bulb

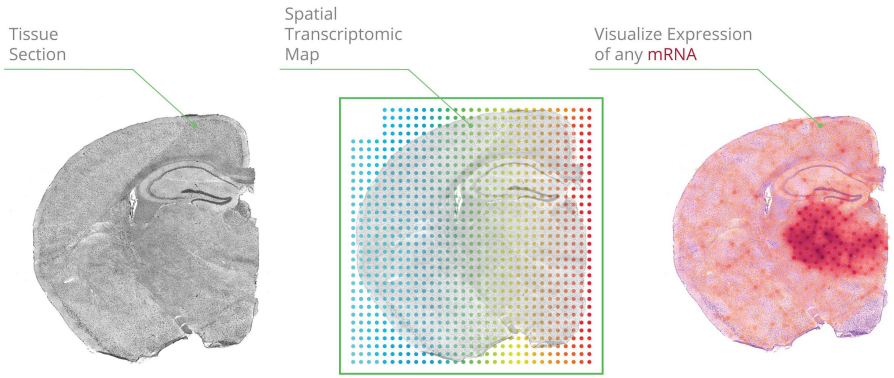
Spatial Transcriptomics





Spatial Transcriptomics

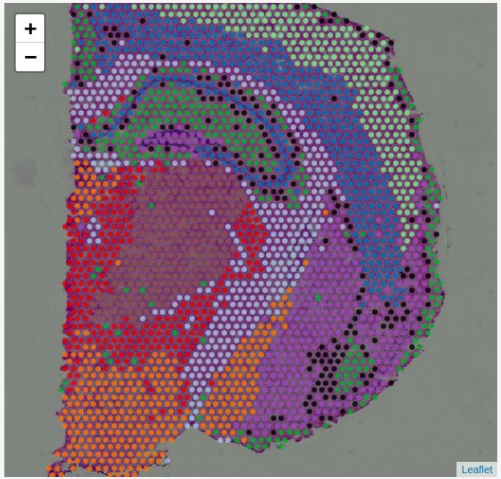
Analysis can reveal spatial distribution of transcripts in the sample



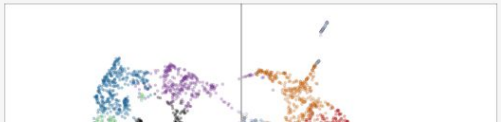
Spatial Transcriptomics

Visualization Controls

Use the sliders under the tissue image to adjust how you visualize and combine the tissue image and the gene expression data. Colors represent clusters identified by differentially expressed genes.



Change map spot opacity:

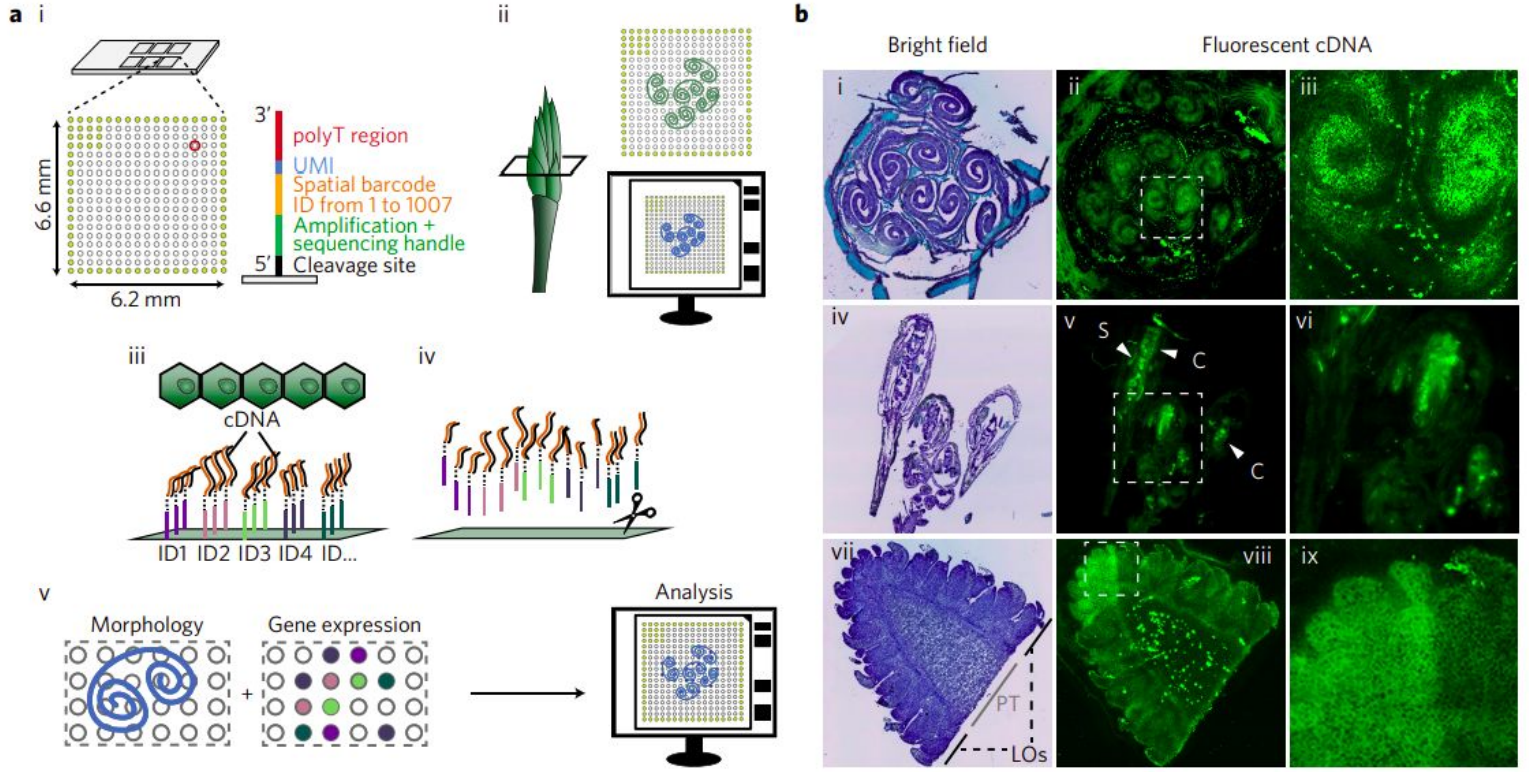


Gene Identification

By placing the pointer above a gene name within the table, spots in the tissue image will be colored based on the expression of that gene. Alternatively, by placing the pointer above a value within the table, you can observe the expression of a specific gene with the spots from an individual cluster highlighted.

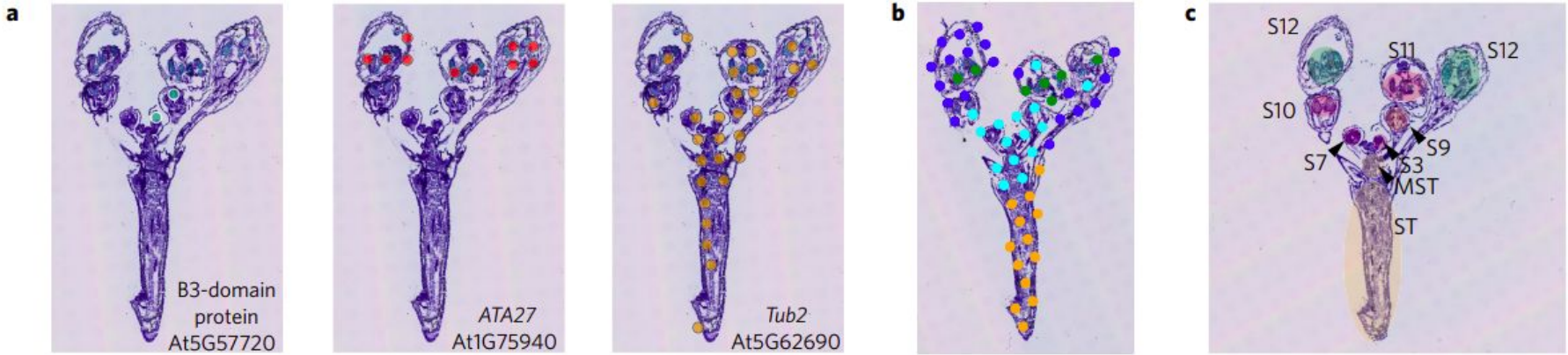
Cluster	1	2	3	4	5	6	7	8	9
Nptxr	3.5	1.1	2.2	4.7	1.8	4.1	1.6	1.2	3.1
Agt	0.59	2.0	3.6	1.2	0.78	0.42	3.2	2.6	0.61
Ttr	3.5	4.9	3.0	4.8	3.1	2.6	2.7	2.8	3.8
Pmch	0.77	1.8	4.5	1.5	0.56	0.79	1.3	0.63	0.66
Camk2n1	5.7	3.7	3.9	5.8	3.9	6.8	4.3	4.2	5.1
Olfm1	5.6	2.8	3.5	5.9	3.1	5.5	3.2	3.6	4.5
Pcp4	4.9	2.8	4.2	4.4	2.0	2.5	3.3	6.2	3.0
Prkcd	0.50	0.81	0.72	0.51	0.48	0.31	1.6	4.7	0.43
Cck	5.5	2.4	2.3	4.8	3.0	5.2	2.3	4.4	4.2
Nnat	2.2	2.7	5.2	3.8	1.6	1.2	3.0	1.8	2.4
Plp1	4.8	7.9	4.8	3.9	3.6	3.1	6.0	6.2	3.7
6330403K07Rik	3.5	2.0	5.3	3.6	1.4	3.3	3.4	1.5	2.2
Ctxn1	4.5	1.7	3.2	4.7	1.6	3.9	1.3	1.1	3.0
Atp1a1	4.2	2.7	3.1	4.2	1.8	4.9	2.3	2.0	3.0
Nrgn	6.6	3.0	3.8	6.2	3.7	6.8	2.3	2.2	5.1
Calb2	0.28	0.52	3.1	1.0	0.31	0.53	2.2	1.1	0.37
Hpca	4.3	1.5	2.3	3.6	1.8	3.8	1.8	1.6	2.8
Fth1	7.3	8.3	7.1	7.1	5.7	6.5	7.1	6.8	6.4
Atp1a2	3.6	3.8	4.2	3.8	3.5	3.7	4.2	4.6	3.7
Sparc	2.4	3.2	5.3	3.1	2.1	2.1	4.7	4.3	2.3

Spatial Transcriptomics in plants



10.1038/nplants.2017.61

Spatial Transcriptomics in plants

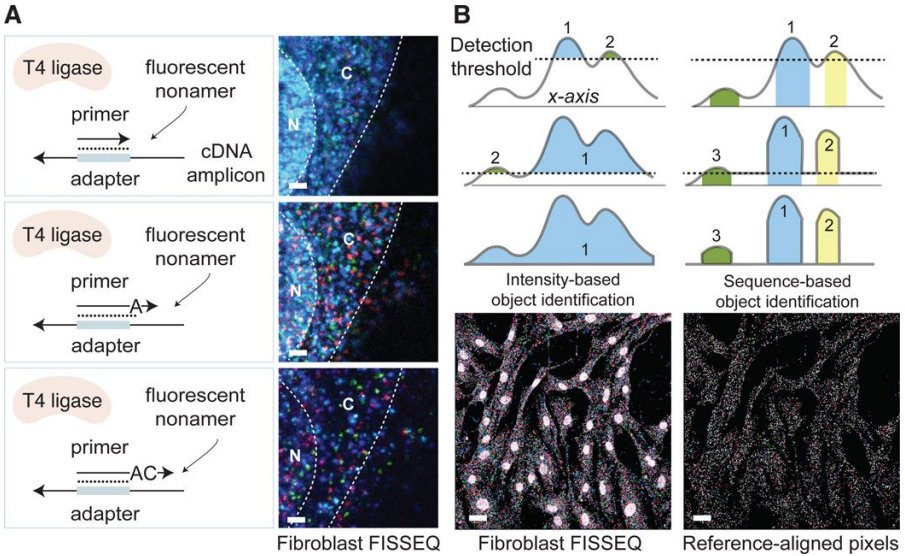


Clustering of genes reflects floral organs differentiation

10.1038/nplants.2017.61

Sequencing in situ

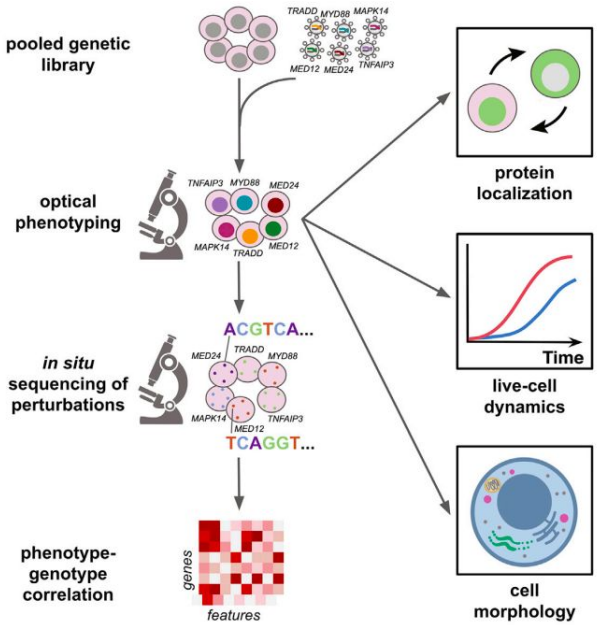
RNA molecules are reverse-transcribed in situ and amplified

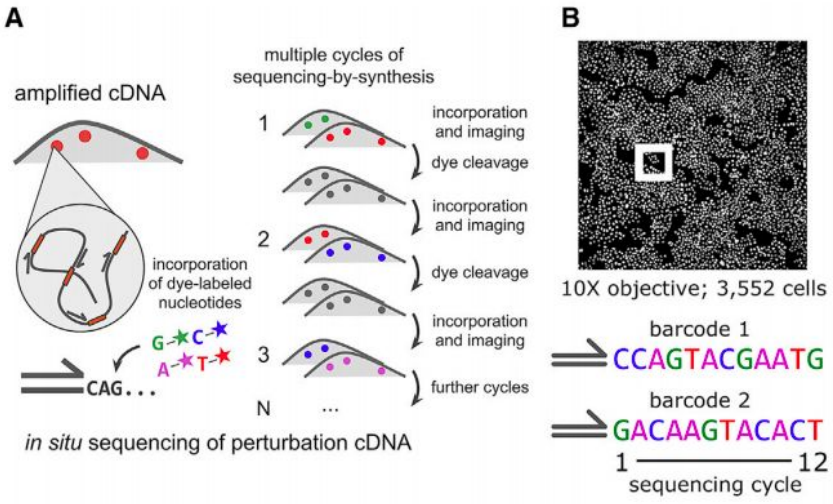


10.1126/science.1250212

In situ sequencing example

In situ sequencing is used to screen different genetic perturbations

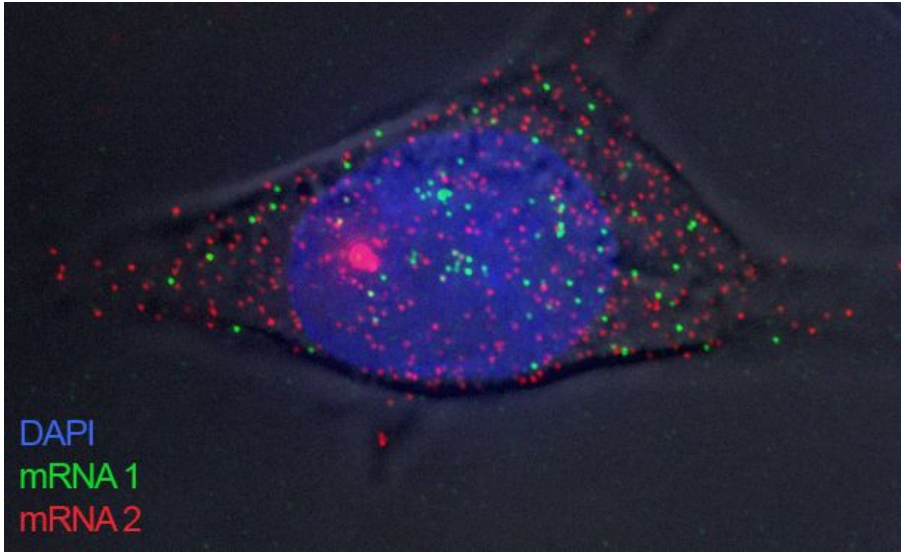




10.1016/j.cell.2019.09.016

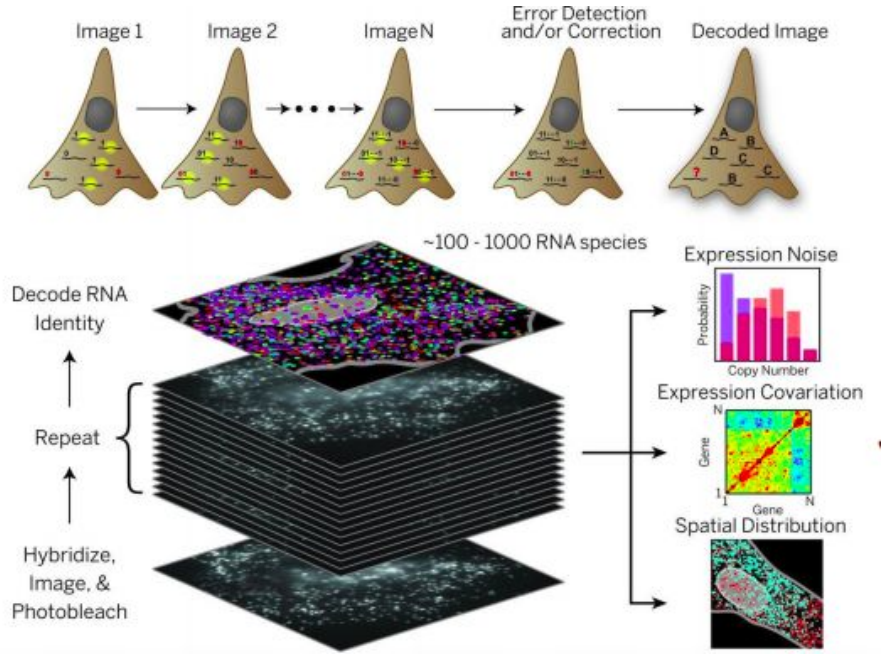
Single-molecule RNA FISH

High-performance FISH for detecting transcripts in the sample



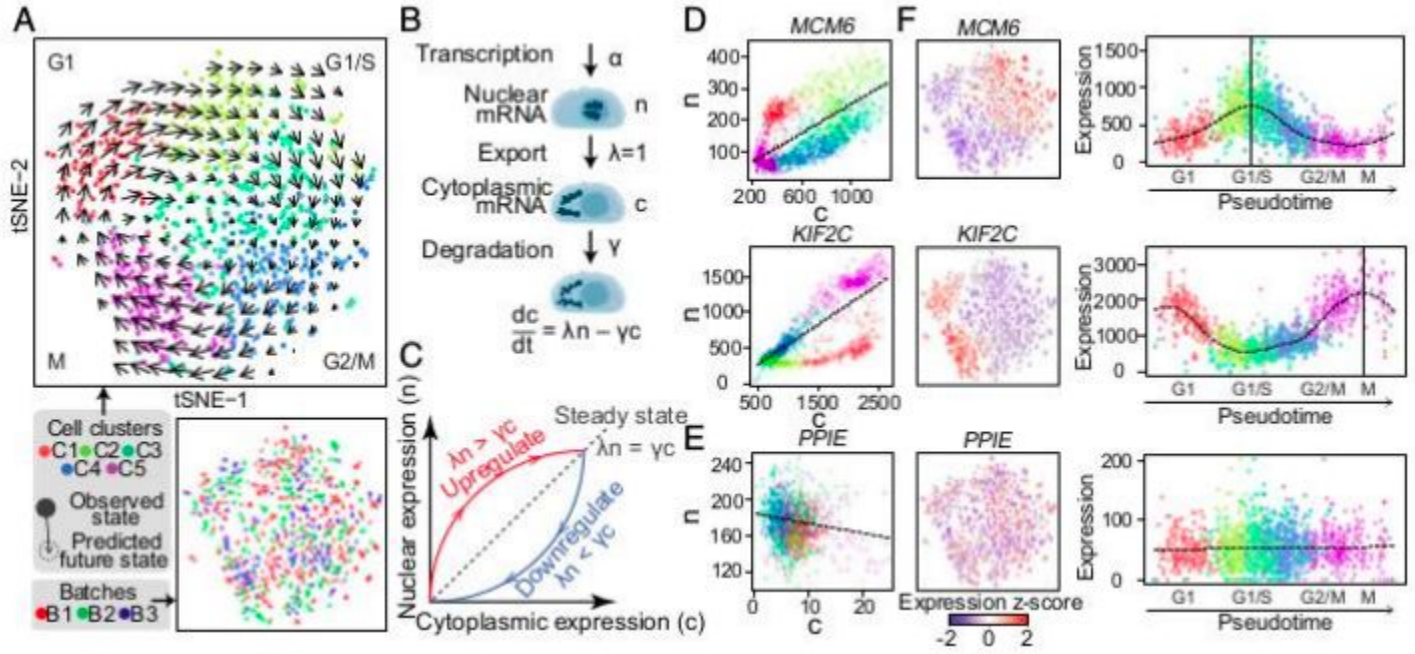
MERFISH

Enables analysis of subcellular localization and RNA velocity



10.1073/pnas.1912459116

MERFISH



10.1073/pnas.1912459116

Spatial RNA-seq downsides and shortcomings

Resolution is usually small (not single-cell level)

Small size of slides

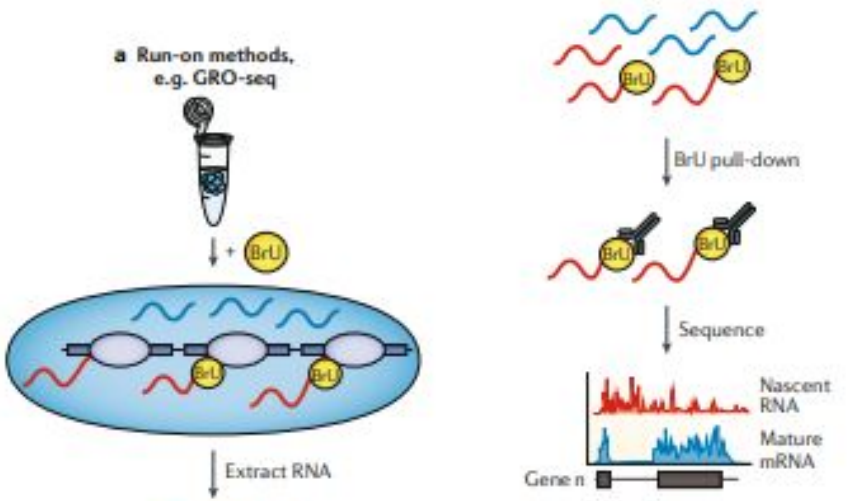
Run-on transcription RNA-seq

Normal RNA-seq actually measures a balance between RNA synthesis and degradation

Run-on methods

Global run-on and precision nucleus run-on (GRO-seq and PRO-seq)

Labeled nucleotides are delivered to the cell to detect nascent transcripts

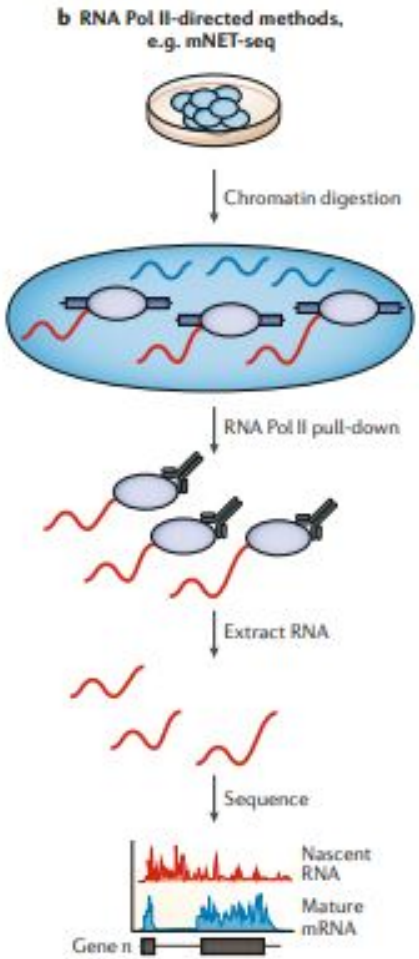


10.1038/s41576-019-0150-2

Polymerase immunoprecipitation

Native elongation transcription sequencing (NET-seq)

Antibodies to RNApol allow to capture transcriptions start sites



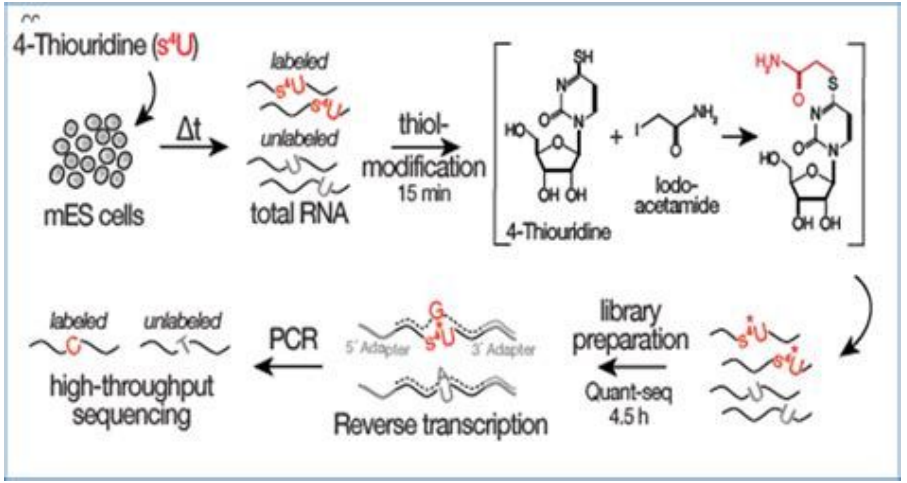
10.1038/s41576-019-0150-2

SLAM-seq

Thiol(SH)-linked alkylation for metabolic sequencing of RNA

Limited in time pulse of 4-tiouridine

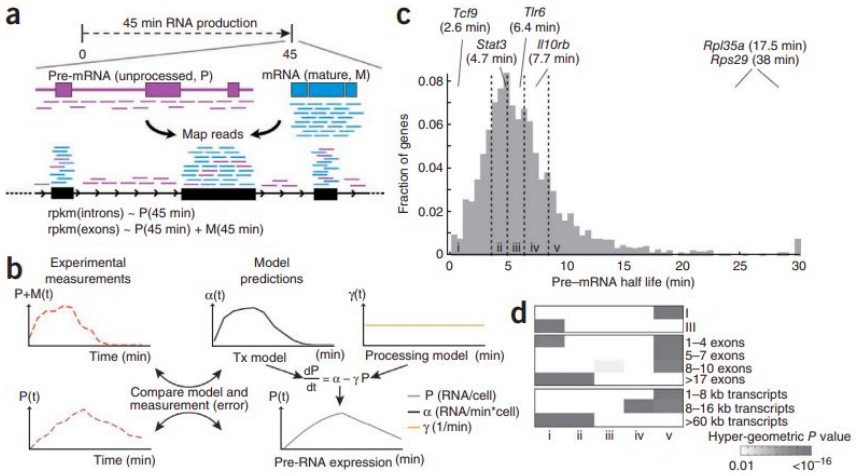
This nucleotide analog leads to substitutions during reverse transcription



Study of degradation and processing rates

Degradation and processing rates are not uniform within a genome

Genes that have an altered degradation rates are enriched in some functional categories (inflammation and immune signaling)



10.1038/nbt.1861

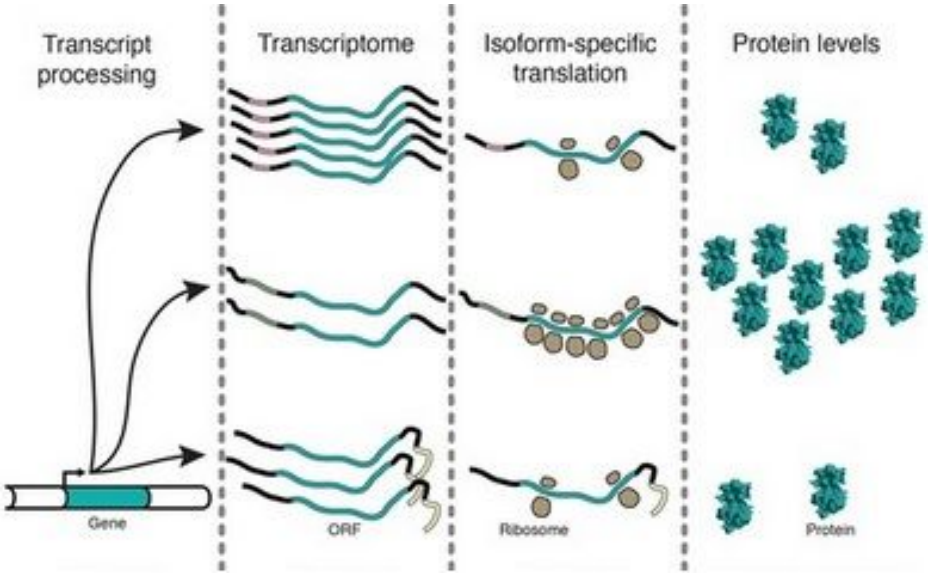
Run-on transcription shortcomings

All methods are vulnerable to degraded RNA and non-specific enrichment

Measuring active translation

mRNA and protein levels do not always correlate

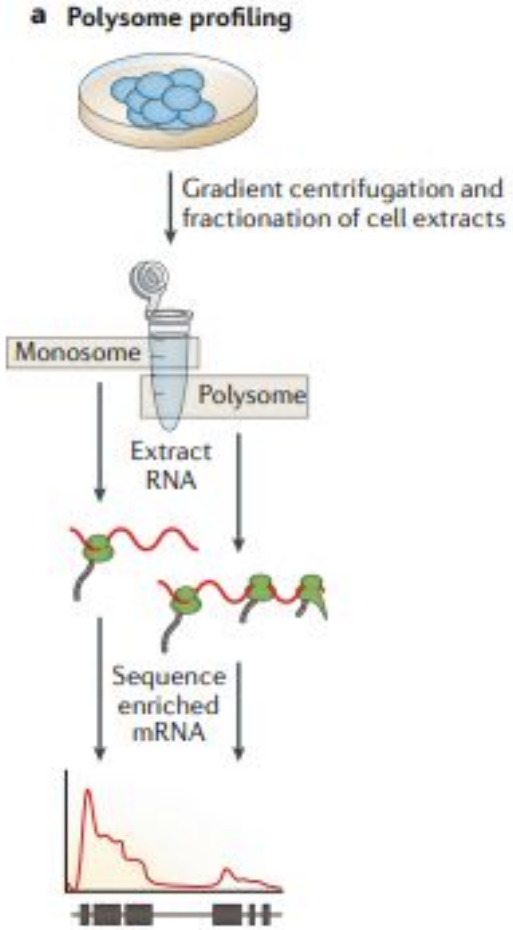
Translation level may better reflect the expression



Polysome profiling

Density gradient centrifugation to separate polysome and single-ribosome fractions

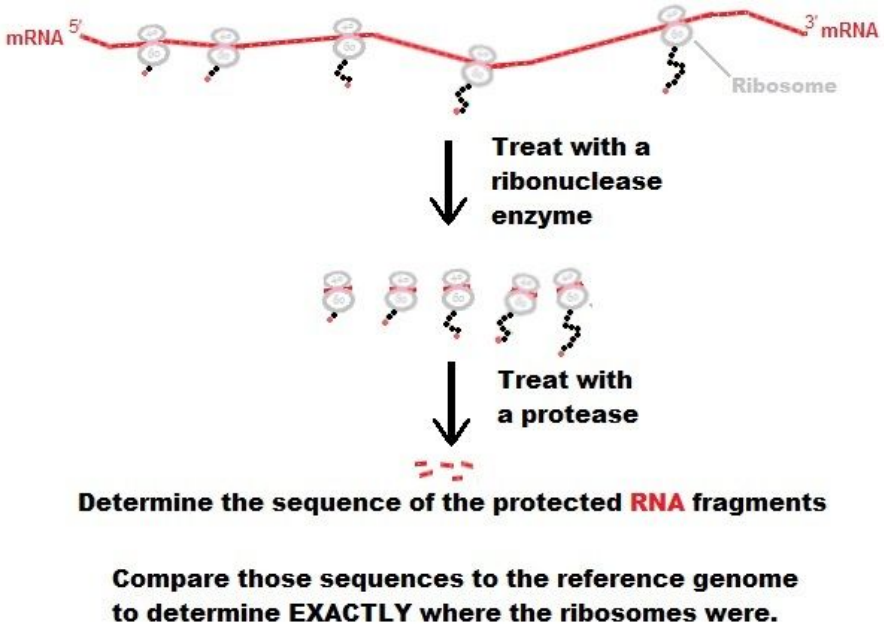
Polysome RNAs are translated more actively



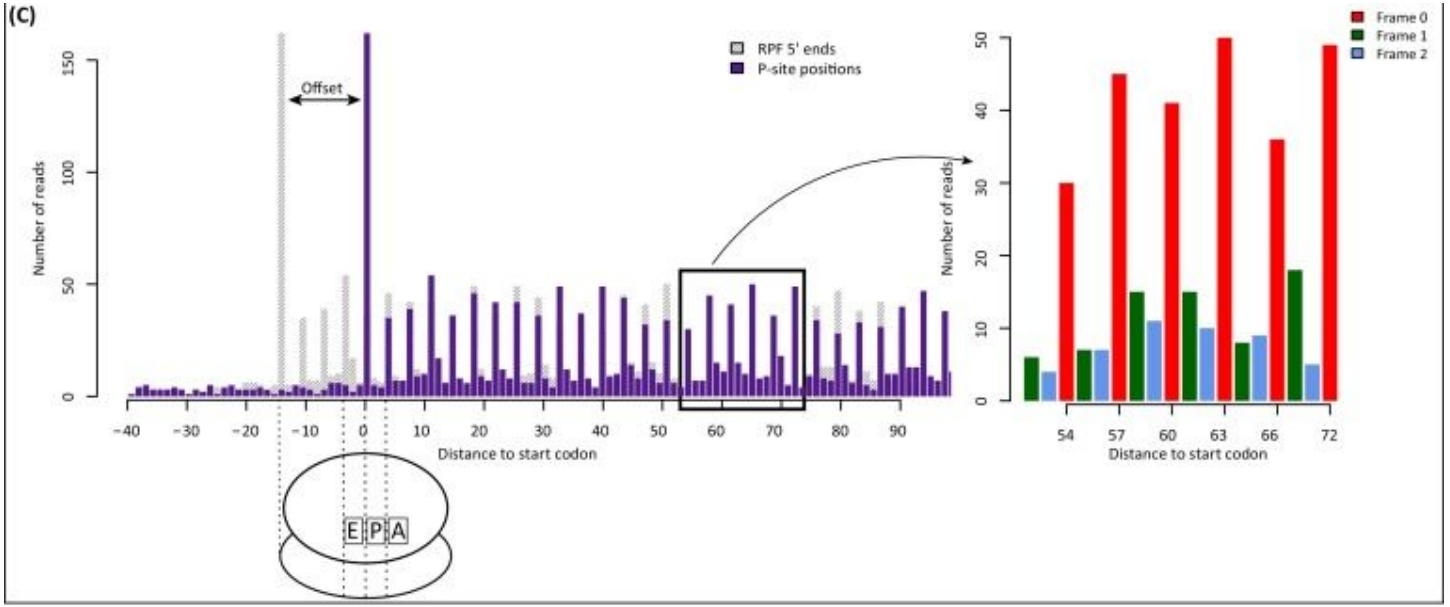
10.1038/s41576-019-0150-2

Ribo-seq

Analysis of RNA fragments protected by the ribosome

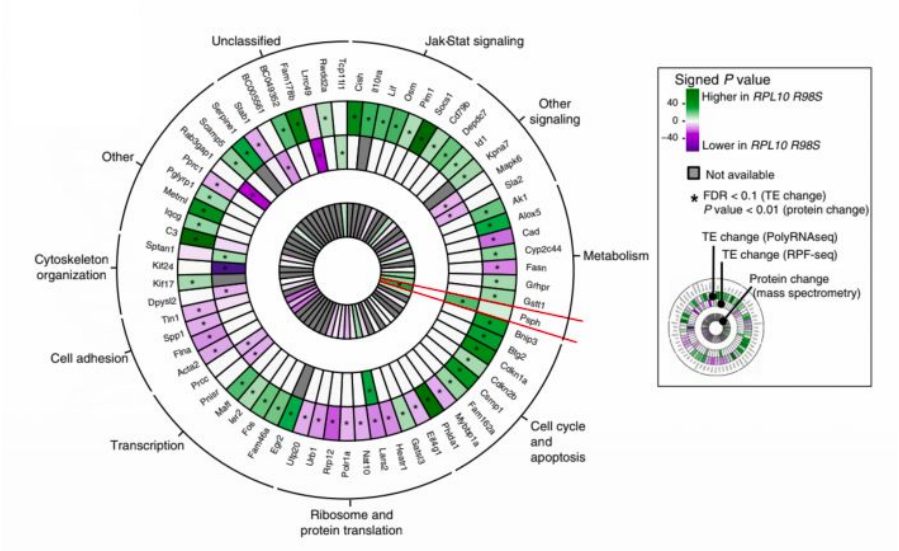


Ribo-seq


(C)
Number of reads
150
100
50
0
Offset
RPF 5' ends
P-site positions
−40 −30 −20 −10 0 10 20 30 40 50 60 70 80 90
Distance to start codon
E P A
Number of reads
50
40
30
20
10
0
Frame 0
Frame 1
Frame 2
54 57 60 63 66 72
Distance to start codon
Trends in Genetics

Translatome analysis in leukemia

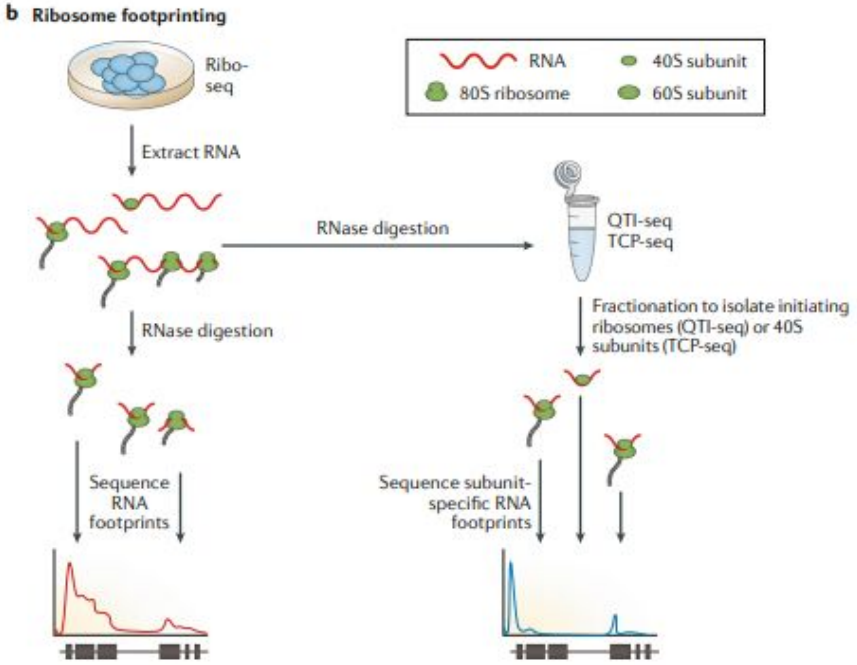
Simultaneous RNA-seq, PolyRNAseq and RFP-seq to capture changes in cancer samples



10.1038/s41467-019-10508-2

Ribo-seq modifications

Ribo-seq can be used to catch some particular subsets of ribosomes, for instance, ribosomes on the initiation stage

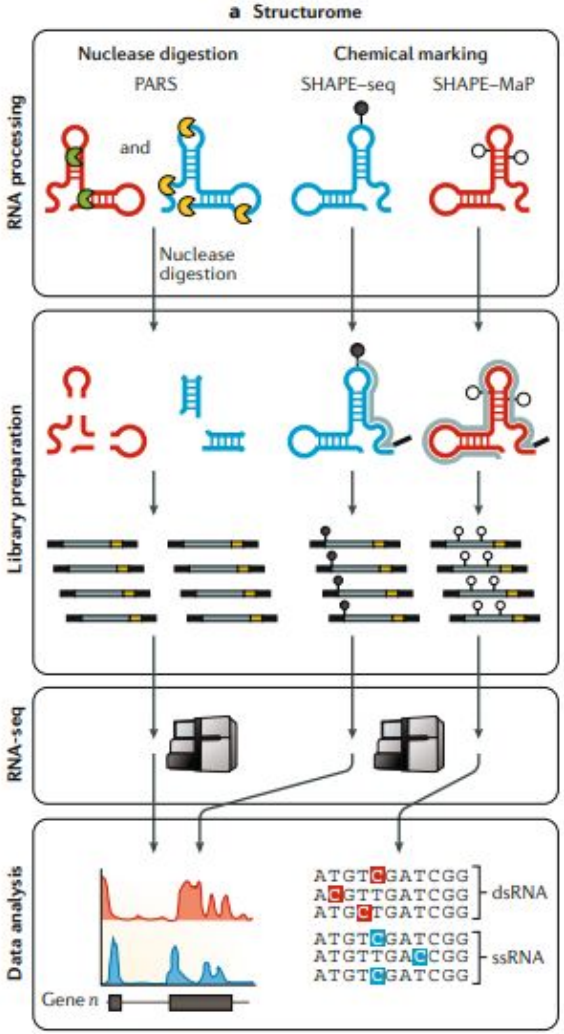


10.1038/s41576-019-0150-2

Structural RNA-seq

RNA structure may have an important role

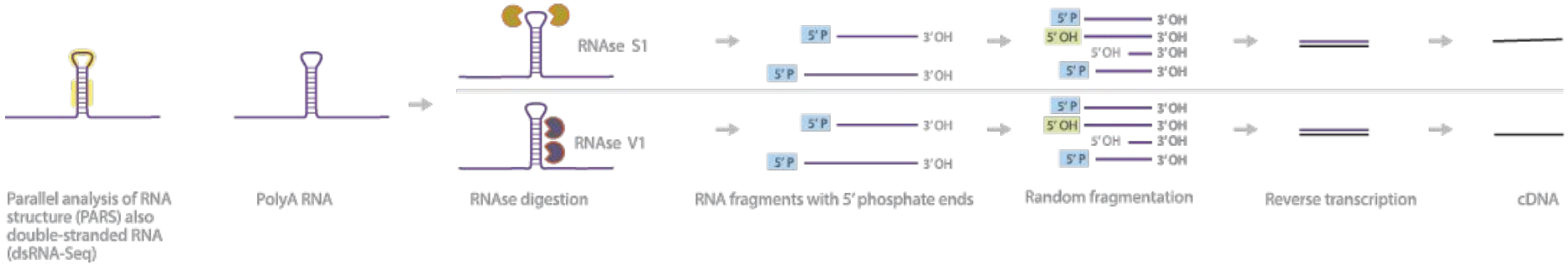
Structure may regulate alternative polyA-site



10.1038/s41576-019-0150-2

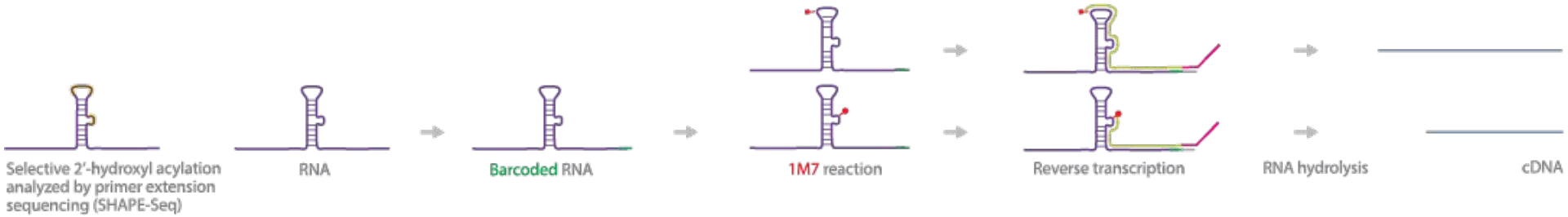
RNA-seq with nucleases

Depending on the nuclease choice either single-strand or double-strand RNA fragments are cleaved



Chemical modification RNA-seq

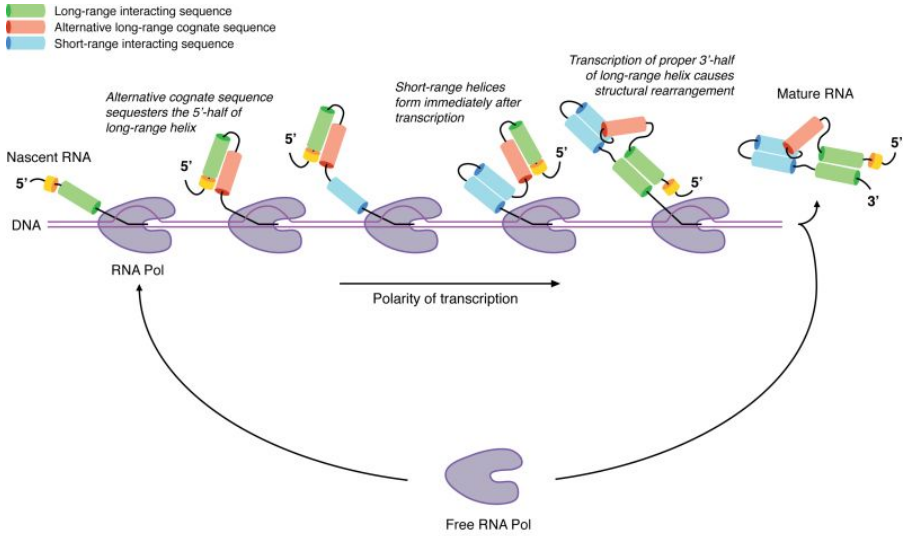
Chemical modifications affect only single- or double-strand RNA and interfere with reverse transcription



Cotrascriptional folding of RNA

SPET-seq

Authors have shown that cotranscriptional folding is a general feature for RNAs in bacteria

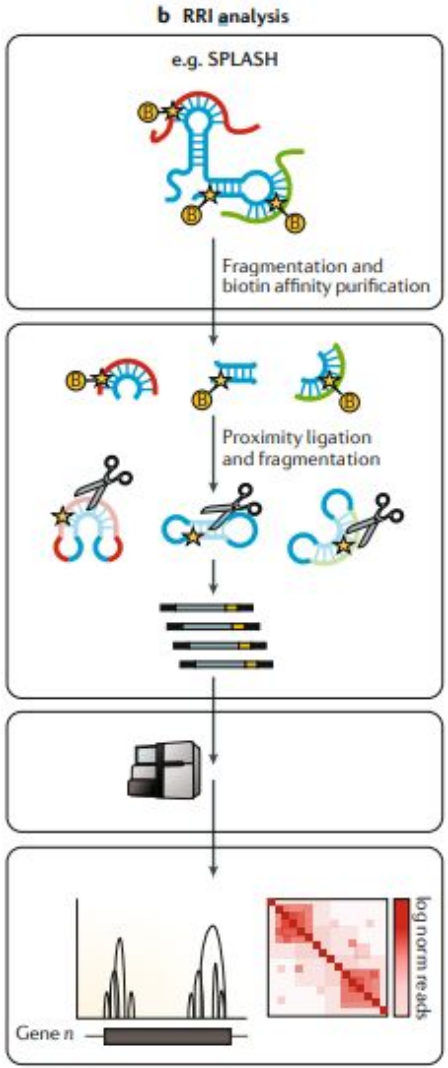


10.1093/nar/gkx617

Studying RNA-RNA interactions

Creating cross links between RNA molecules

Low efficiency, intramolecular interactions may obfuscate the result



10.1038/s41576-019-0150-2

Studying RNA-protein interactions

RIP-seq

RNA immunoprecipitation

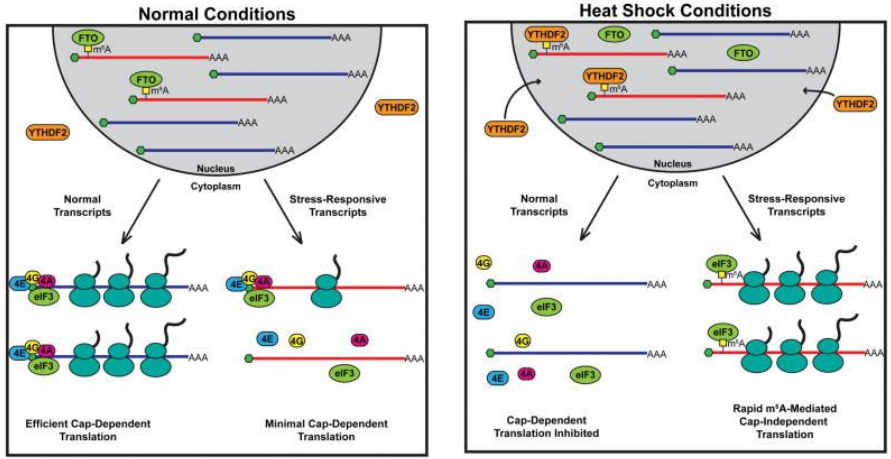
The main difference from ChIP - no cross links

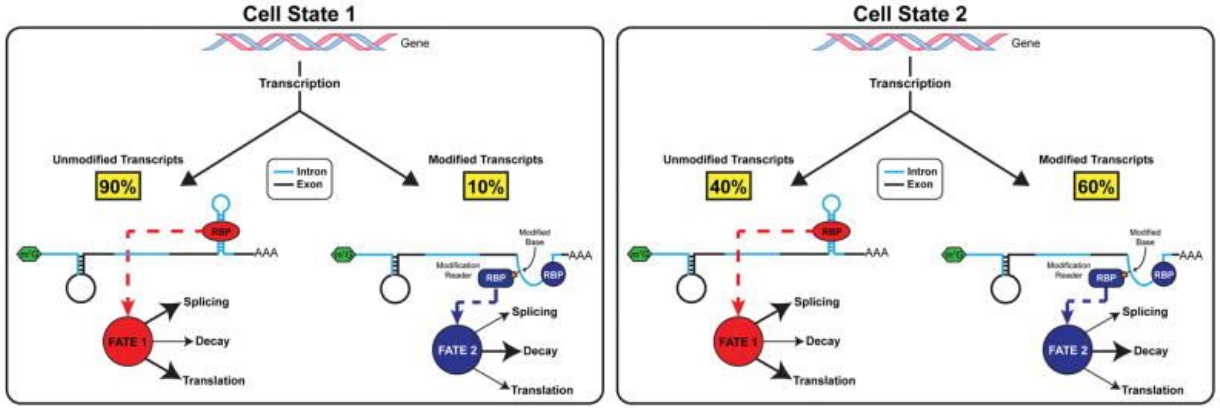
Weaker interactions result in more non-specific results

Epitranscriptome

Additional level of control over gene expression

Chemical modifications can regulate RNA secondary structure and translation



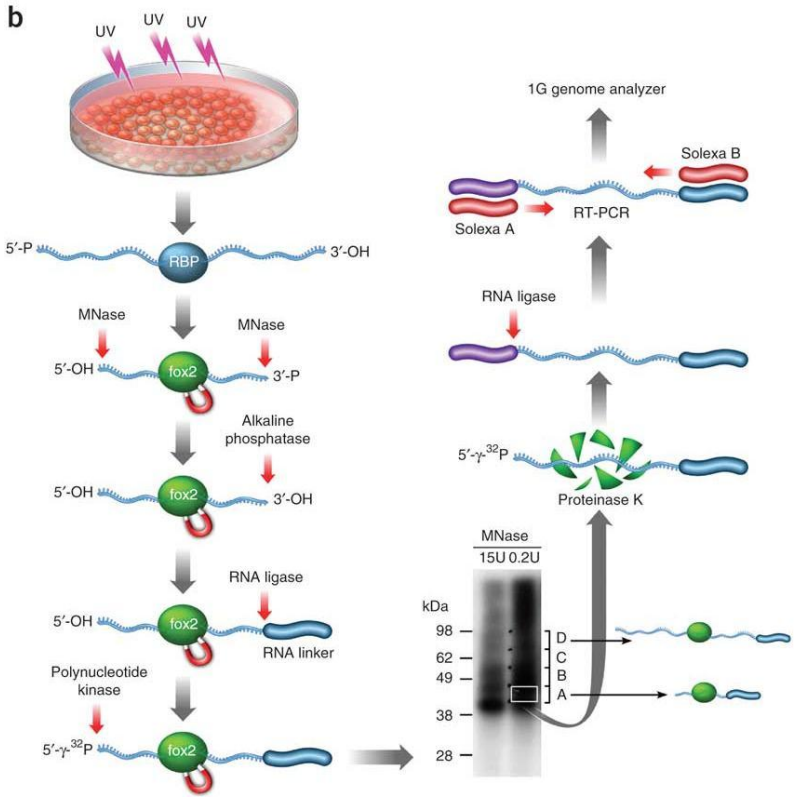


10.1038/nrm.2016

CLIP

UV to form RNA-protein cross links

Many modifications that allow resolution up to single nucleotides



Questions, suggestions?