Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes

Benjamin J. Blencowe,
Sidrah Ahmad, and
Leo J. Lee1

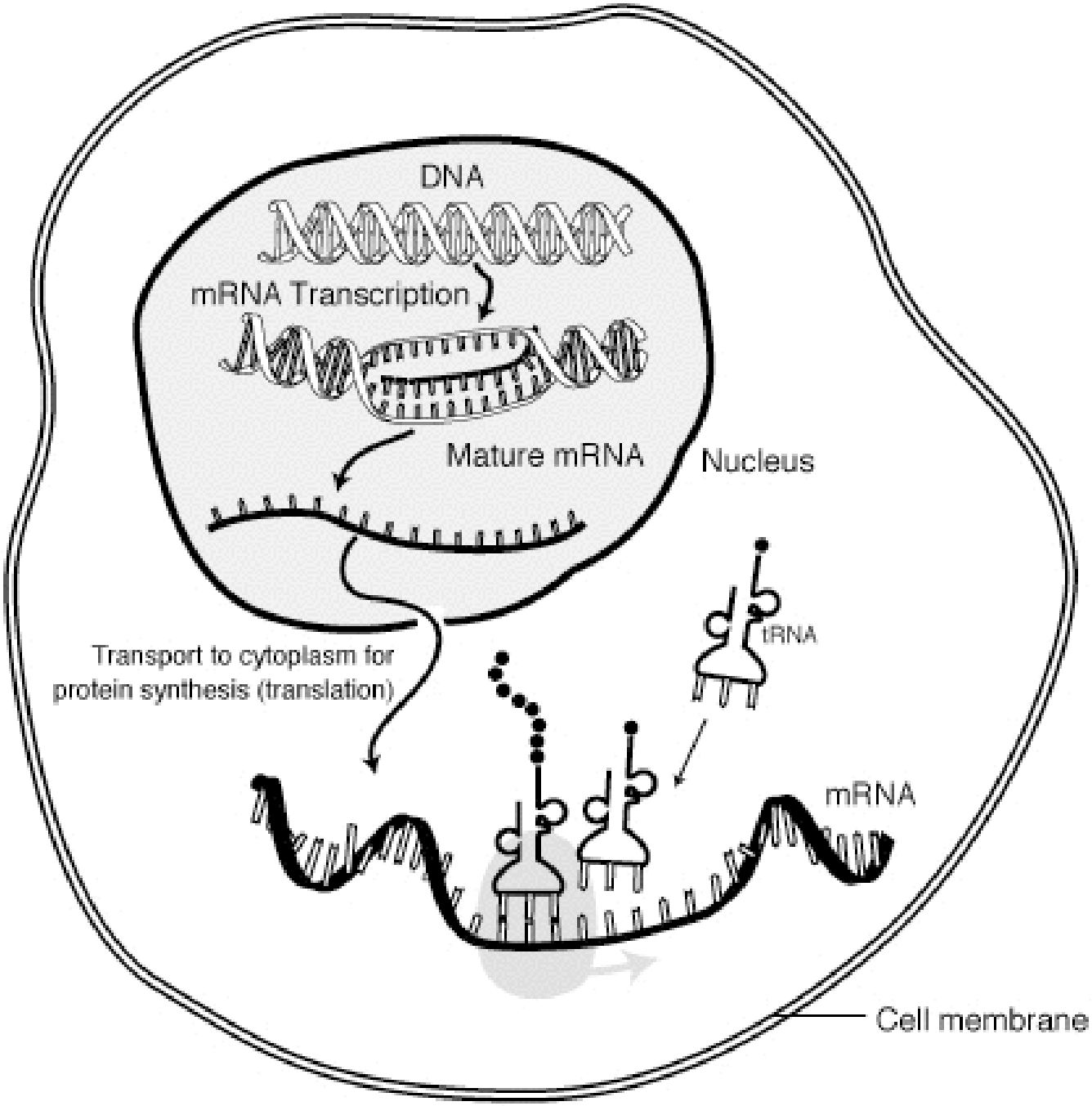
Presentor: Seyed

Introduction

- HTS is a powerful tool in profiling coverage and quantitative accuracy
- detection of substantial new transcript complexity
- elucidation of binding maps and regulatory properties of RNA-binding proteins
- new insights into the links between different steps in pre-mRNA processing

Introduction

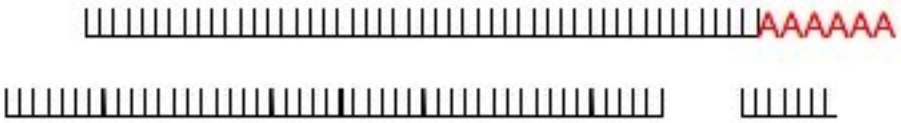
- RNA transcript can serve in a critical regulatory or enzymatic capacity
- An ultimate goal is to accurately predict functional properties of RNAs from sequence features alone and also how these functions are altered in human diseases
- HTS methods for analyzing RNA populations, a.k.a “RNA-Seq” (“mRNA-Seq” for mRNA) are a major step.


DNA
mRNA Transcription
Mature mRNA
Nucleus
Transport to cytoplasm for
protein synthesis (translation)
tRNA
mRNA
Cell membrane

Introduction

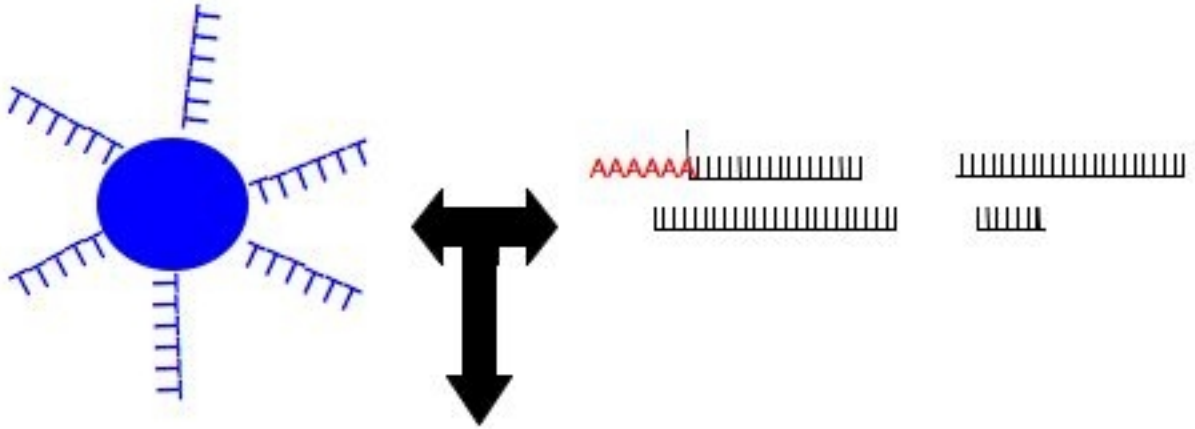
- Microarray profiling systems employ glass slides containing thousands to millions of anchored oligonucleotides designed to hybridize to transcript sequences.
- Approach is limited by the probe sets available for specific hybridization on the microarray
- Limited sensitivity and specificity by detection is indirect, typically measured as a fluorescence signal
- RNA-Seq provides a relatively unbiased
- RNA-Seq is performed using tagged libraries of short cDNAs
- no prior knowledge of the sequences to be profiled is required

Isolate Total RNA

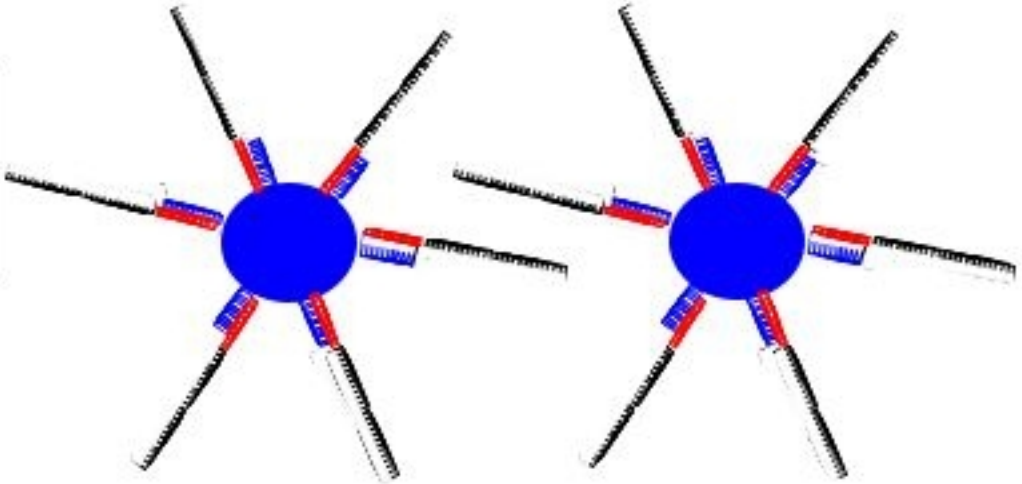


Fragmentation and/or Isolation

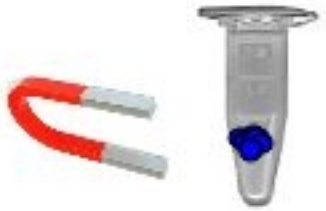
In this case, isolation via Poly(T) coated magnetic beads



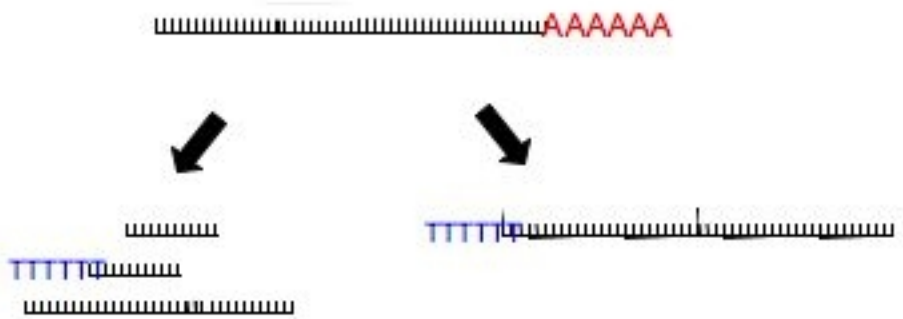
Poly(A) RNA molecules bind to the Poly(T) magnetic beads



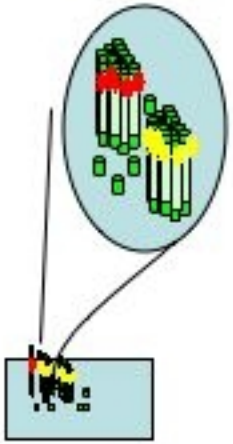
Magnetically isolate and wash beads



Fragment and/or Reverse Transcribe



Fragmentation (if not done already), size selection, and sequence



Illumina Solexa, Roche 454, or ABI SOLiD

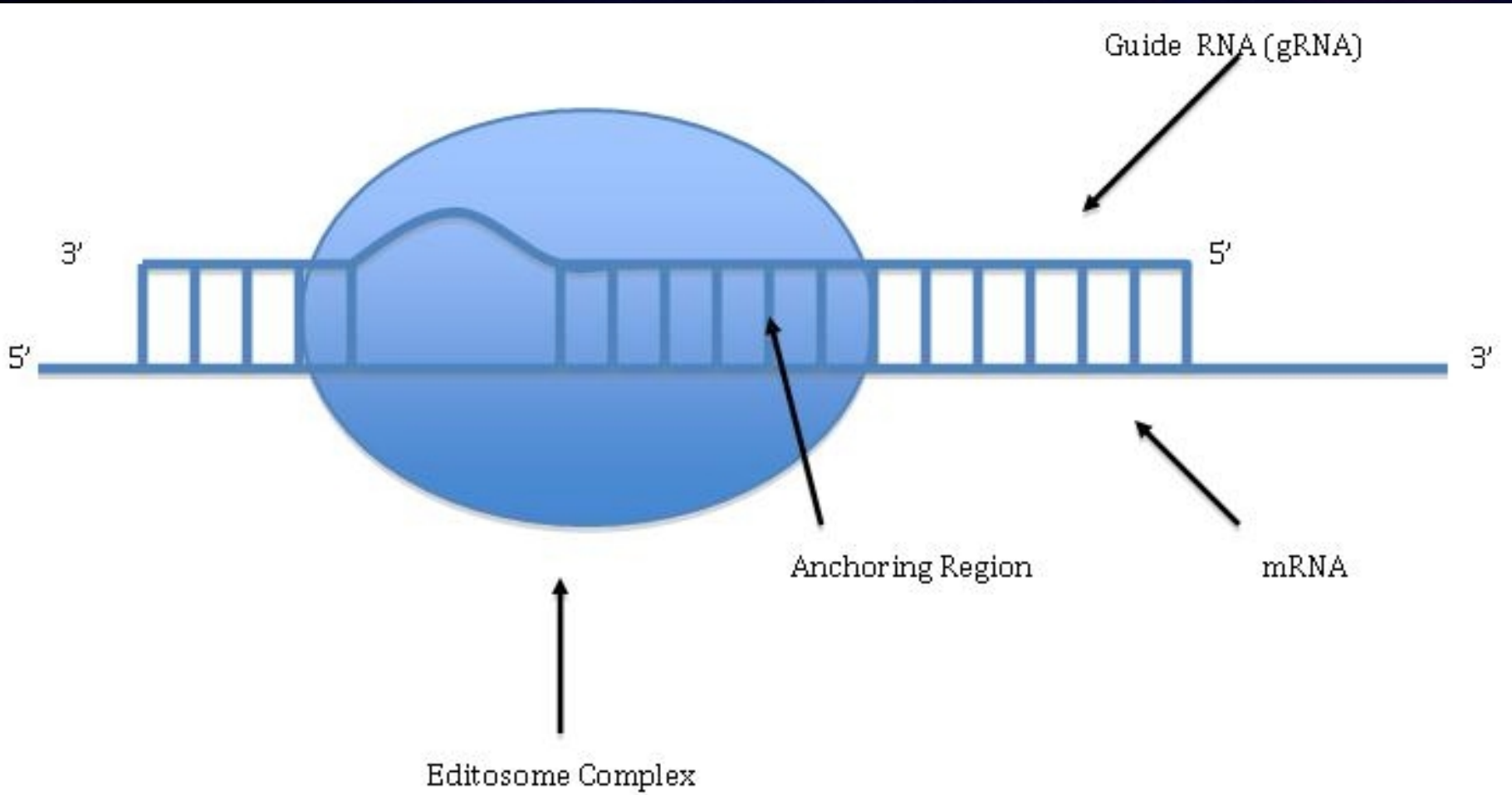
Graphic shown here is Illumina

Introduction

- A drawback of HTS is the high cost to obtain the data, but it has a high quality.
- The Illumina and AB produce tens of millions of reads (~50 nt per read, 2- to 3-d run)
- The Roche system generates a few hundred thousand reads (400–500 nt per read)

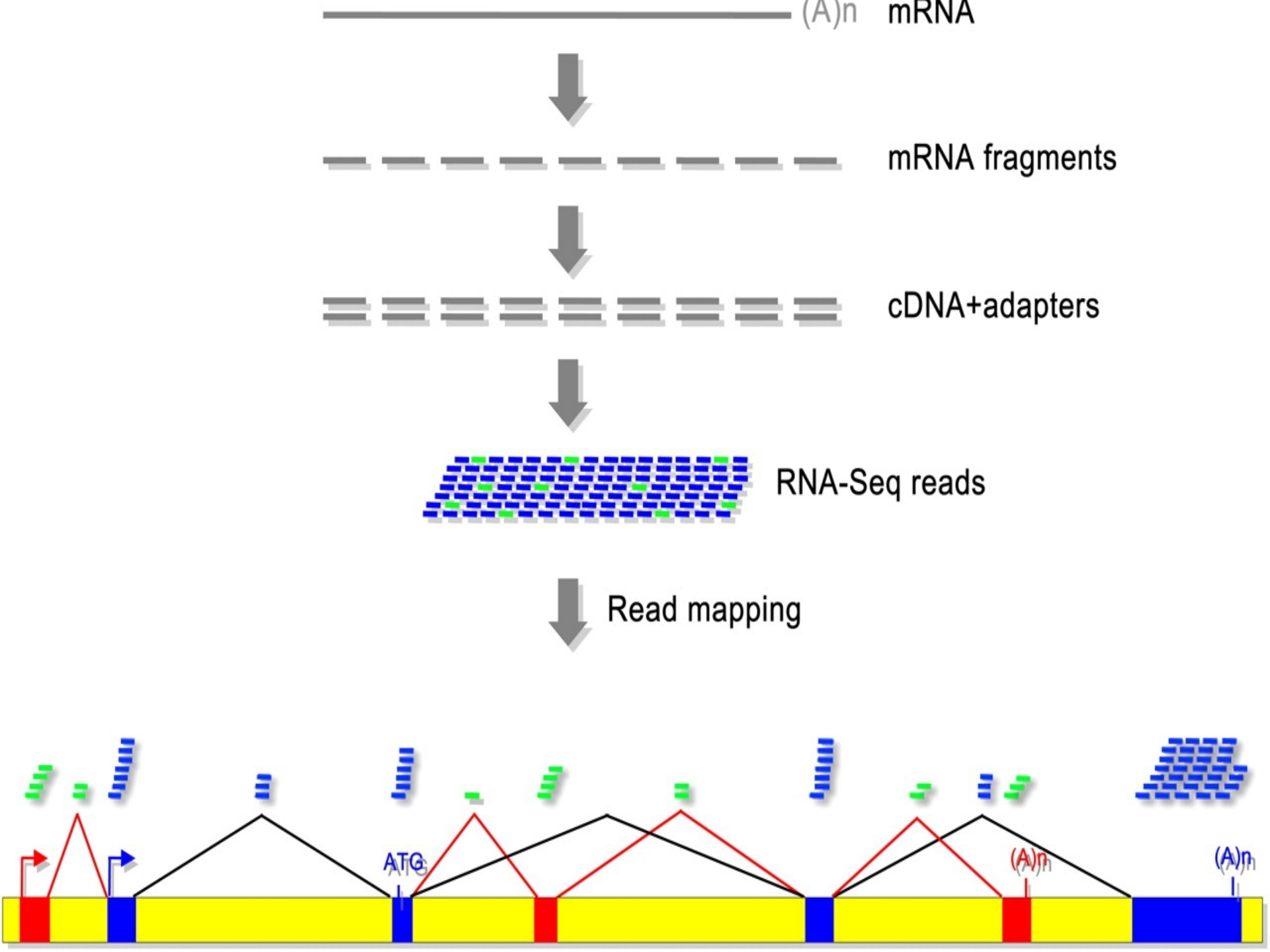
Deep surveying of mRNA processing complexity and regulation

- Pre-mRNA transcripts undergo a series of modification subject to regulation leading to transcript diversification.
- Additional processes such as RNA editing, in which individual bases are altered, can lead to further transcript diversity.
- transcriptomes of eukaryotes are highly complex
- Previously ESTs and longer cDNAs were employed for analyses of mRNA population.
- alternative splicing and other forms of transcript processing developed.
- lack of sufficient EST/cDNA coverage from individual cell and tissue types.
- custom microarrays with probe sets overcame many of the obstacles.


Guide RNA (gRNA)
3'
5'
5'
3'
Anchoring Region
mRNA
Editosome Complex

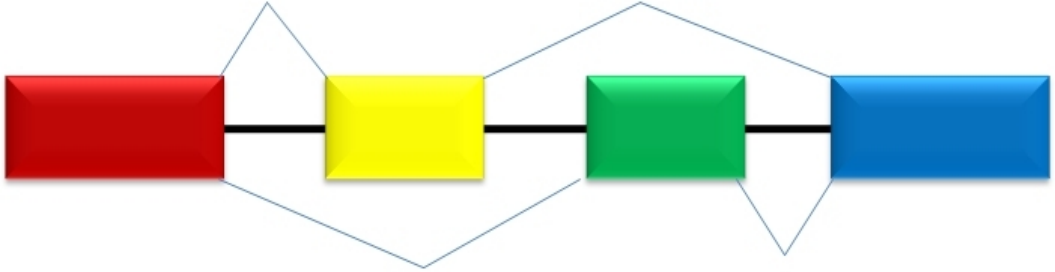
Deep surveying of mRNA processing complexity and regulation

- RNA-Seq represents the best tool to characterize transcriptomes
- first step identify reads that uniquely align to the genome and transcriptome
- no “gold standard” so far.
- Grimmond and Wold laboratories describes initial results from analyzing mammalian mRNA-Seq
- Mortazavi et al analyzed mouse tissue transcriptomes and identified unique mapping reads corresponding to ~17,000 previously unannotated regions of known genes, ~145,000 distinct splice junctions were detected and alternative splicing was detected in ~3500 genes.


(A)n
mRNA
mRNA fragments
cDNA+adapters
RNA-Seq reads
Read mapping
ATG
(A)n
(A)n

Deep surveying of mRNA processing complexity and regulation

- using the AB SOLiD system Cloonan et al mouse transcriptomic changes were profiled using ~100 million 25-nt.
- Approximately one-third of unique mapping reads were detected outside of known exons.
- Sultan et al reported analyses RNA-Seq on human embryonic kidney.
- detected extended 5′ and 3′ UTR sequences.
- Four-thousand-ninety-six putative novel junctions were detected in 3106 genes.
- Wang et al detected new exons and junctions by mapping sequence reads to a library of computationally predicted exons and splice junctions
- thousands of “high-confidence” new candidate splice junctions detected.
- alternative splicing occurs in 92%–96% of human genes, or ~98% or more of multiexon genes.



pre-mRNA

splicing

alternatively spliced mRNAs

translation

protein isoforms


http://go.funpic.hu

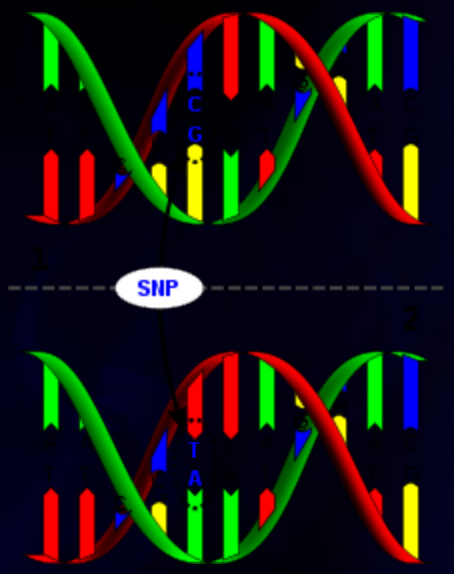
Deep surveying of mRNA processing complexity and regulation

- Our group analyzed 17–32 million Illumina read data sets from six diverse human tissues
- Estimated alternative splicing occurs in transcripts from 92%–97% of multiexon human genes
- Frequency of detection of alternative splicing per exon is independent of the number of exons per gene
- Number of alternative splicing events per gene increased in a near linear fashion
- Genes with higher numbers of introns are statistically more often associated with human disease.

Cell-, tissue-, and individual-specific transcript variants

- Wang et al study transcripts from 105,000 alternative splicing events mined from cDNA/EST data
- majority displayed tissue-dependent variation in alternative exon inclusion levels
- scoring an absolute inclusion ratio change of at least 10% result in detection of substantially greater number of tissue-dependent differences.
- Wang et al. found human alternative splicing events that undergo the most pronounced tissue-dependent changes are significantly more often frame-preserving
- tissue-dependent alternative splicing events are more widespread than recognized previously with more conserved functions

Linking RNA regulation with trans-acting factors



- a subset of SNPs located within exons or neighboring intronic sequences are associated with individual-specific variation in alternative splicing levels
- Wang et al estimated that 10%–30% of alternative splicing events exhibited interindividual-specific variability
- individual-specific variation in alternative splicing is less frequent than tissue-dependent variation in alternative splicing
- remarkable sensitivity and quantitative nature of RNA-Seq
- a powerful source of data for linking individual- and population-specific genetic variation
- as well as disease-associated mutations
- Between 0.5% and 1% of human genes contain RBDs (e.g. RRMs, homology (KH) domains).
- However not many been functionally characterized
- accurately map RBD proteins physiologically relevant binding sites.
- RNA-Seq is proving to be a powerful tool for this.

Linking RNA regulation with trans-acting factors

- RNA-Seq have revealed “splicing regulatory networks” (SRNs).
- Darnell and colleagues described such an SRN on Nova-2 a brain-specific alternative splicing regulator.
- genes containing Nova-2 found to be enriched in functional annotations associated with synapse biology.
- Where studied, it was generally found that specific motifs corresponding to known binding sites of the targeted alternative splicing regulators are enriched in exons and/or flanking intron sequences of the coregulated alternative exons
- HTS is contributing tocis- and trans-acting factor-dependent global regulation of RNA processing.
- Darnell used HTS to provide a high-resolution map of Nova-2 binding in mouse neocortex tissue.

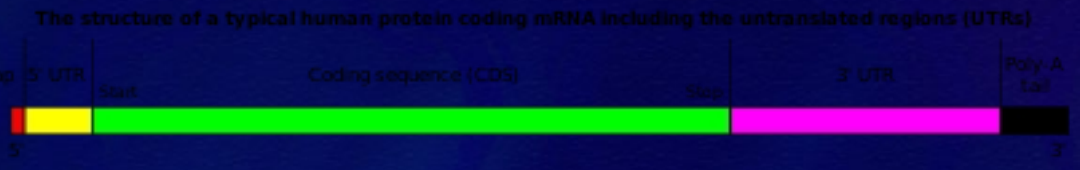
POKER CAT

Is Not Amused.

http://go.funpic.hu

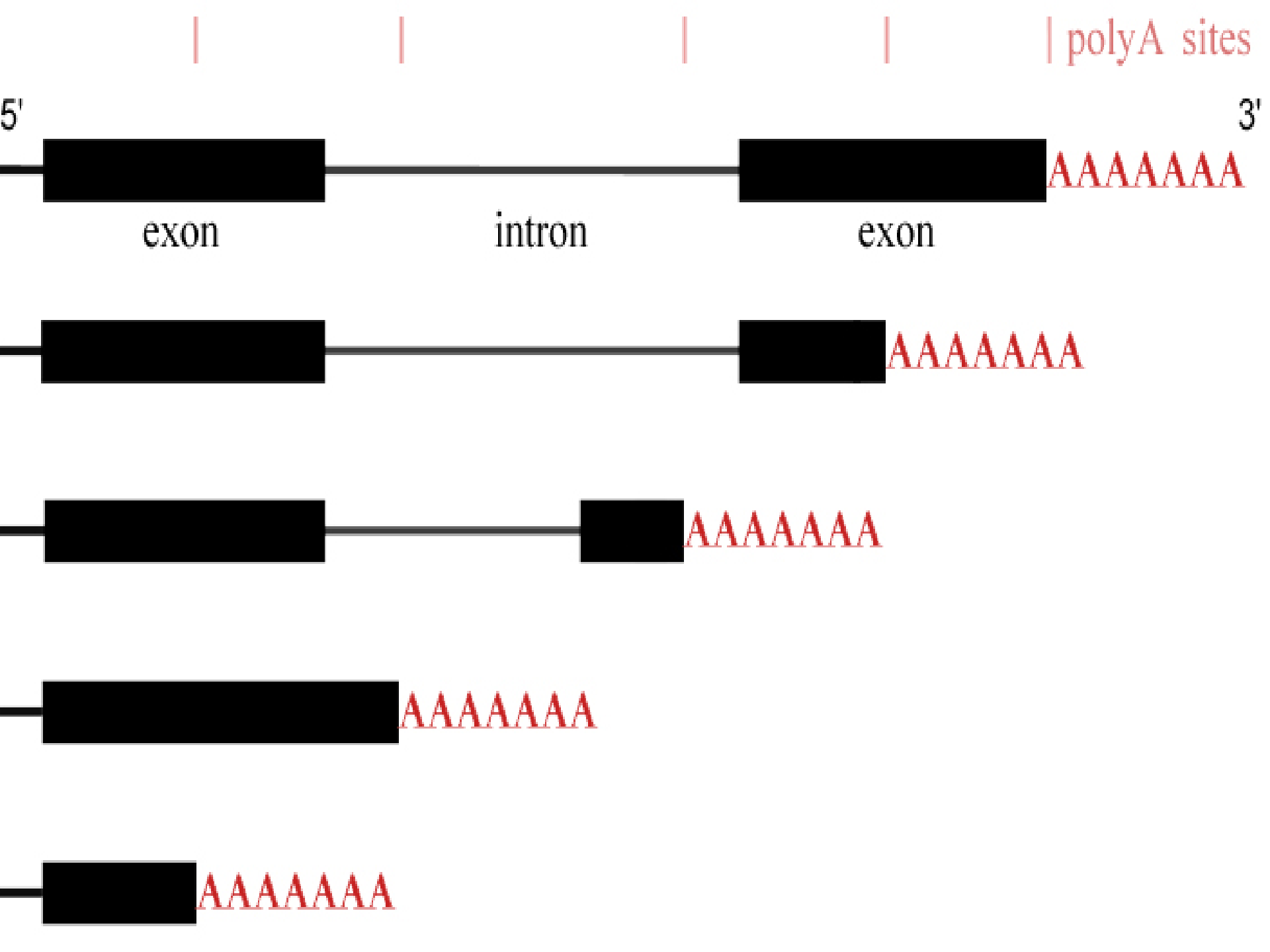
Coordinated RNA processing events and regulatory factor multitasking

- RBD proteins function in more than one step in the generation of mature mRNA transcripts.
- Some regulatory factors can directly impact more than one step leading to mRNA translation.
- RNA-Seq promises accelerate the discovery of “unexpected” functions for RBD proteins.
- Splicing and polyadenylation are closely coupled.
- Recent RNA-Seq- and microarray-based profiling revealed that alternative poly(A) site selection is a common process.
- In Burge et al experiment 86% of mapped alternative poly(A) sites in UTRs exhibited a directional shift resulting in shorter 3′ UTRs coinciding with a late stage of differentiation following stimulation of resting T cells.



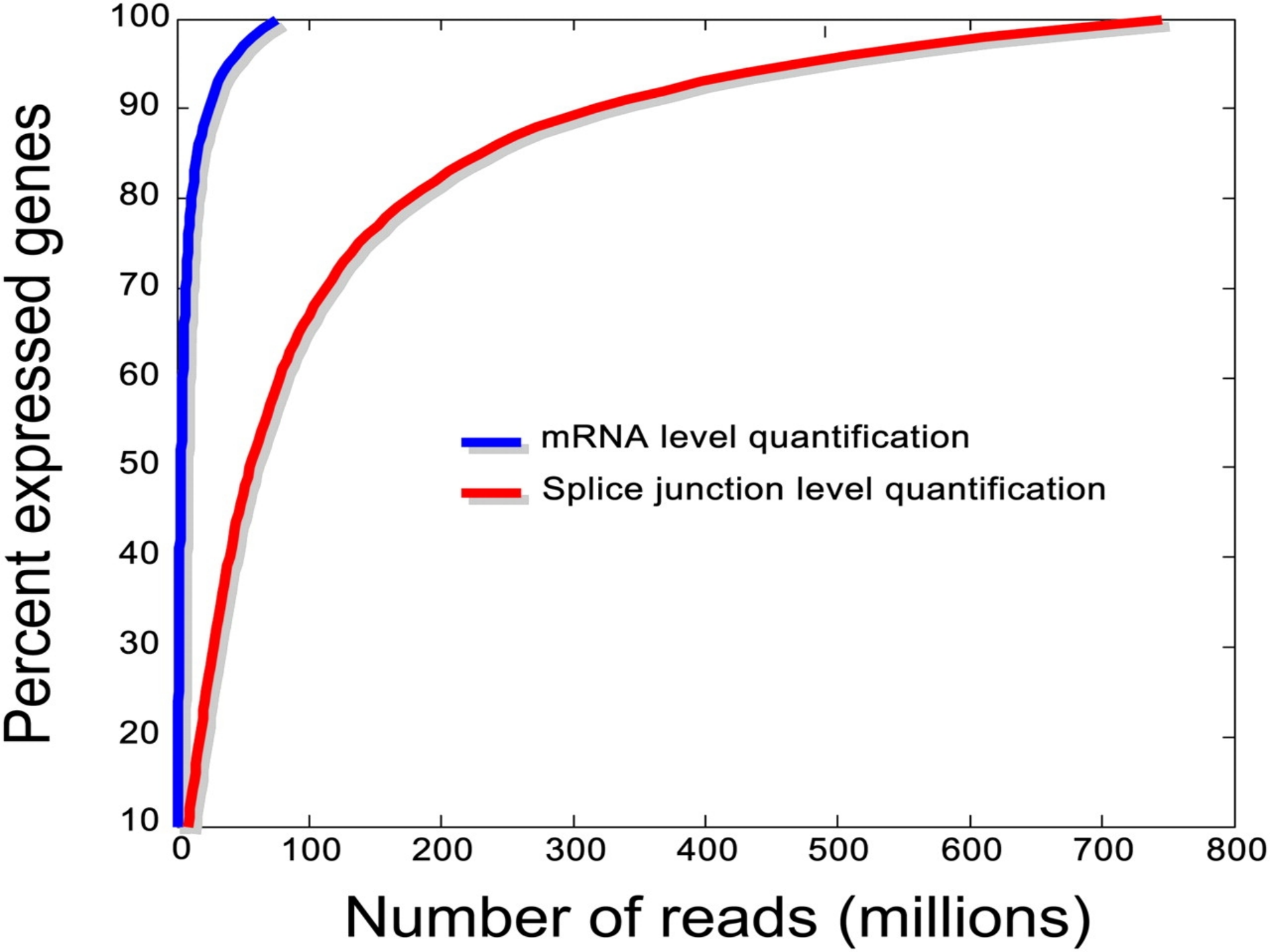
Coordinated RNA processing events and regulatory factor multitasking

- Burge et al showed differential alternative polyadenylation usage between tissues is more frequent than different types of tissue variable alternative splicing.
- Wang et al studied the linkages between alternative splicing and alternative polyadenylation
- A subset of the enriched motifs was common to the two types of RNA processing events.
- there may be widespread roles for alternative splicing regulators in the regulation of polyadenylation
- Licatalosi et al found 297 changes involving alternative 3′ UTR sequences.
- Combination of RNA-Seq profiling of transcriptomes with CLIP-Seq is a powerful tool !!!


polyA sites
5'
3'
AAAAAAA
exon
intron
exon
AAAAAAA
AAAAAAA
AAAAAAA
AAAAAAA

Current and future challenges in the emerging transcriptomics era

- HTS goal is to derive a “unified” network of gene expression regulatory steps.
- Different subsets of genes are regulated at the transcriptional and RNA processing levels
- physical coupling mechanisms may generally serve to temporally coordinate and enhance the kinetics of individual steps in transcription and processing in a cell/tissue- or condition-independent fashion.


Percent expressed genes
100
90
80
70
60
50
40
30
20
10
0
100
200
300
400
500
600
700
800
Number of reads (millions)
mRNA level quantification
Splice junction level quantification

Current and future challenges in the emerging transcriptomics era

- current generation systems do not yet provide an efficient means to comprehensively define the full complement of transcript isoforms in an RNA sample
- short-read profiling
- “paired-end” (PE) sequencing: sequences at 5′ and 3′ ends of the same cDNA molecule can be determined
- systems currently in development are expected to permit single-molecule sequencing
- Accurate measurements of exon inclusion levels require requires on average ~400 reads (at ~35 nt) per kilobase
- shotgun sequencing of mRNA samples is too expensive
- A possible solution: generate pools of primers directed to specific transcript regions of interest

Helicos BioSciences Corporation

Current and future challenges in the emerging transcriptomics era

- to what extent the transcript variants generated by different RNA processing steps are functionally significant?
- one can expect a spectrum of functional importance ranging from essential for viability to neutral activity and potential fodder for evolutionary adaptation.
- which transcripts are most likely to be translated?
- employment of RNA-Seq to characterize yeast mRNA sequences that are bound and protected by polyribosomes recovered by affinity purification
- With HTS we can systematically address the specific roles of the myriad of transcript variants

References

1. Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes

Benjamin J. Blencowe,

Sidrah Ahmad, and

Leo J. Lee1

2. Wikipedia