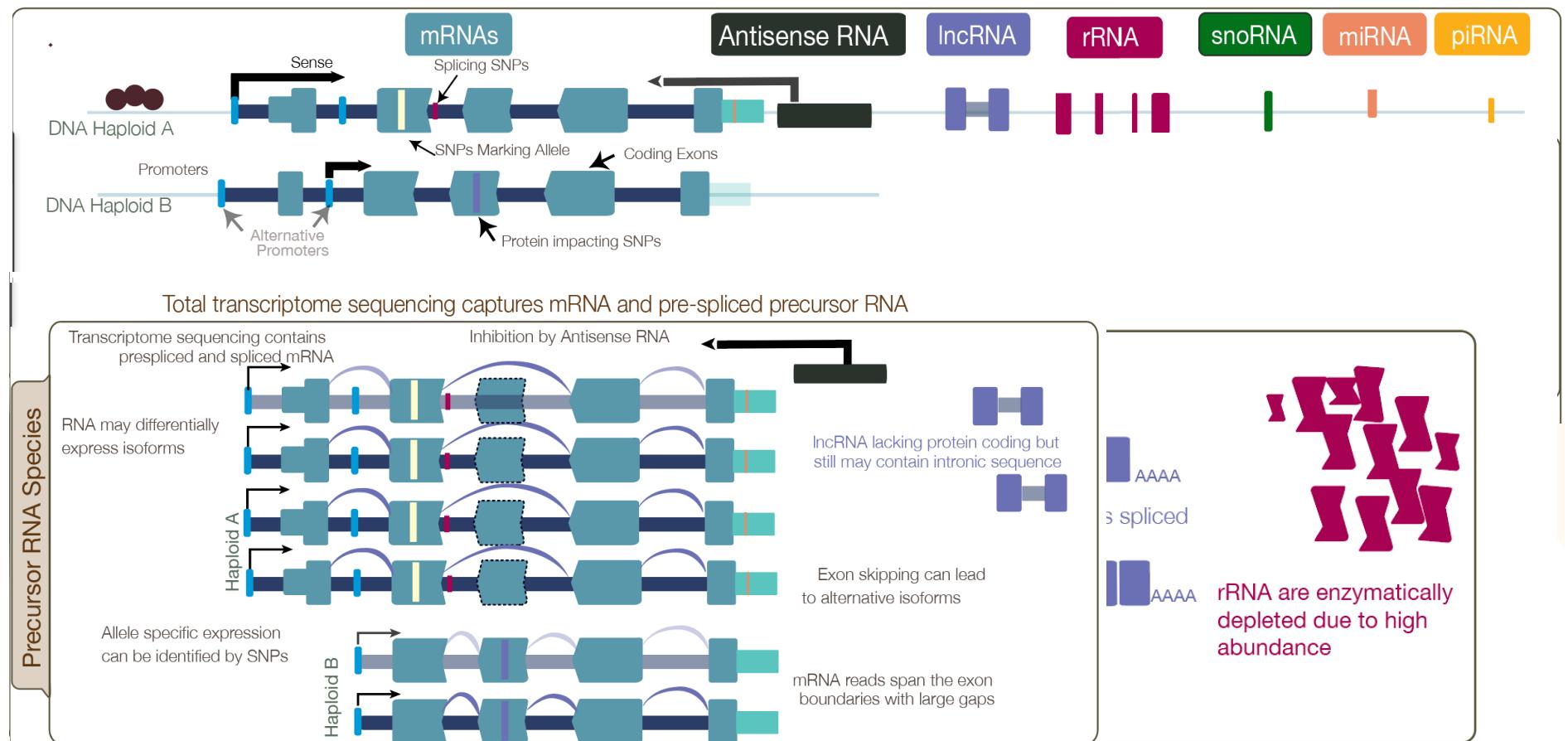


HETROSCEDASTICITY

BEYOND MEASURING MESSENGER RNA



BIOMEDICAL INFORMATICS

WEEK OF APRIL 2ND: PART 2 ON RNA

RNA

Small RNA sequencing

miRNA (~18–24 nucleotides)
Potential biomarker in progression of neurological disease and cancer prognosis

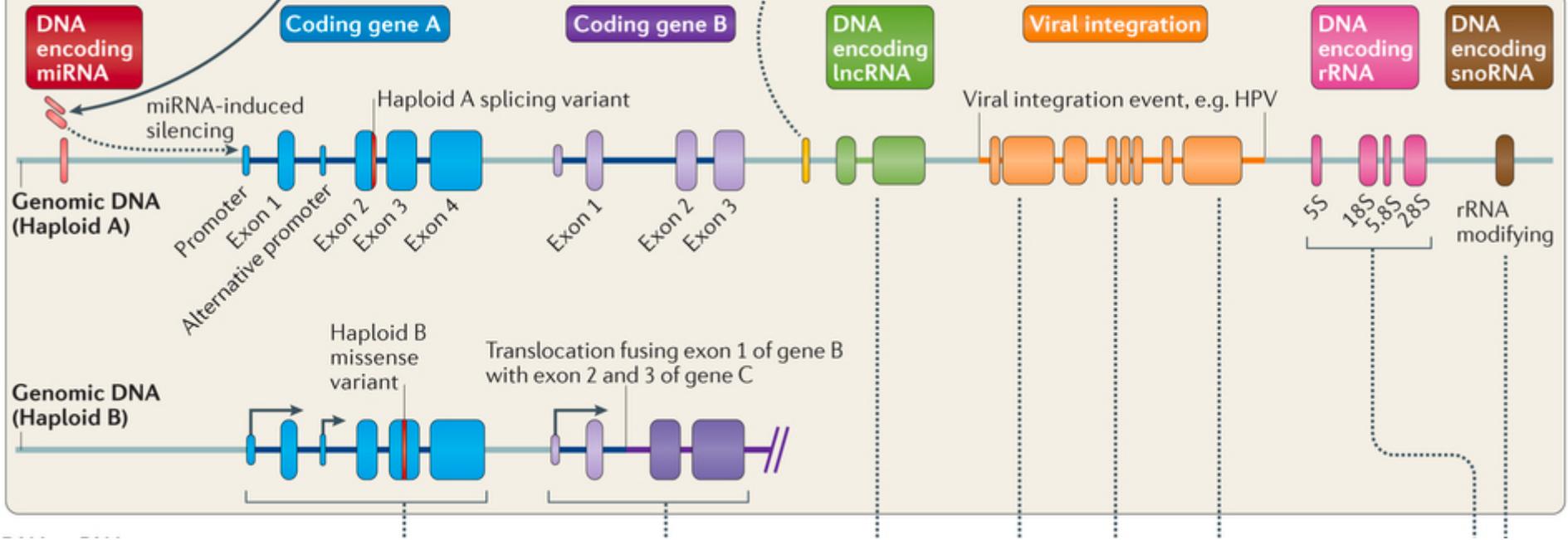
piRNA (~26–32 nucleotides)
Potential biomarker in disease progression

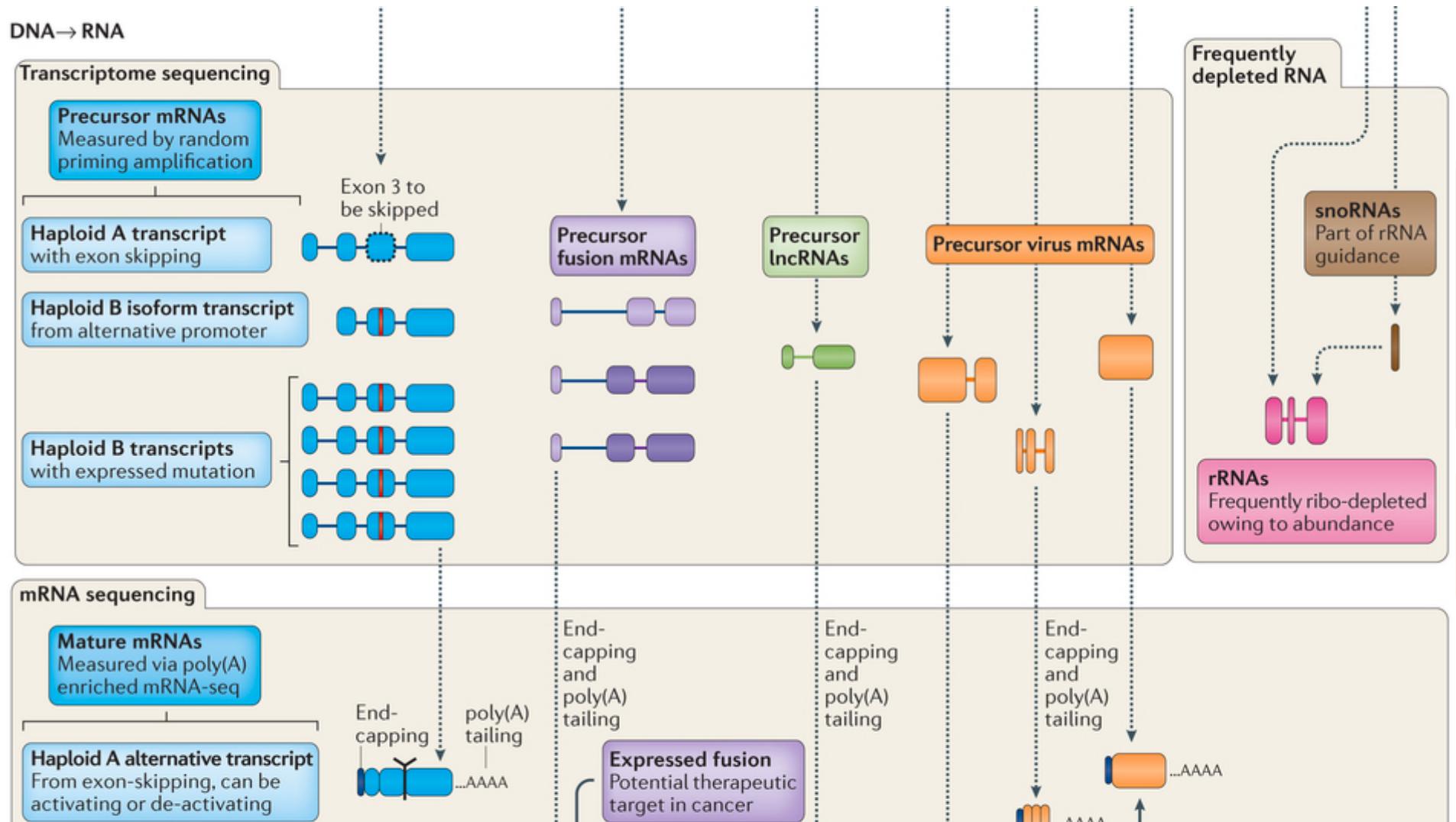
APPLICATIONS OF NEXT-GENERATION SEQUENCING

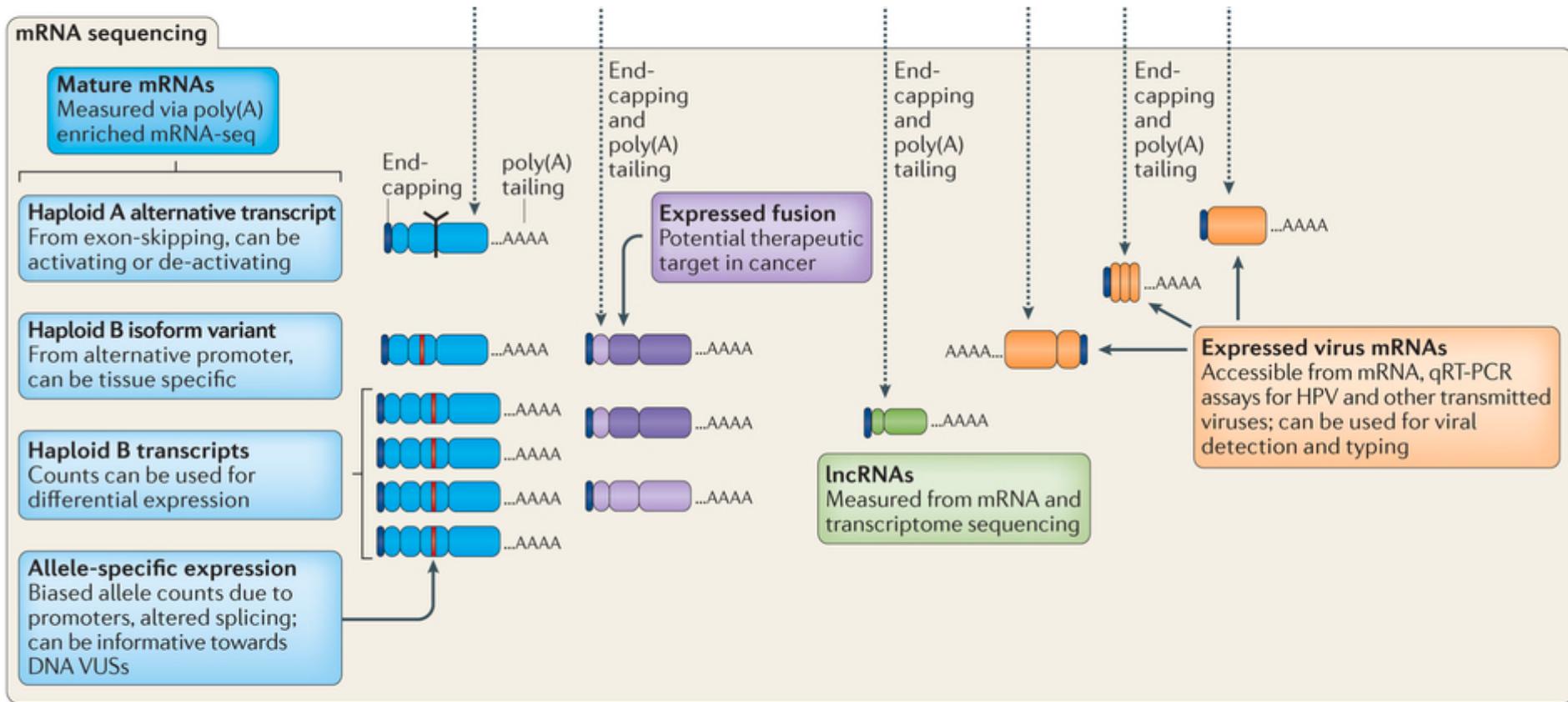
Translating RNA sequencing into clinical diagnostics: opportunities and challenges

Sara A. Byron¹, Kendall R. Van Keuren-Jensen², David M. Engelthaler³, John D. and David W. Craig⁴

Genomic DNA sequencing







MYTH BUSTING: RNA-SEQ AND SPLICING

Exon skipping as a treatment for DMD



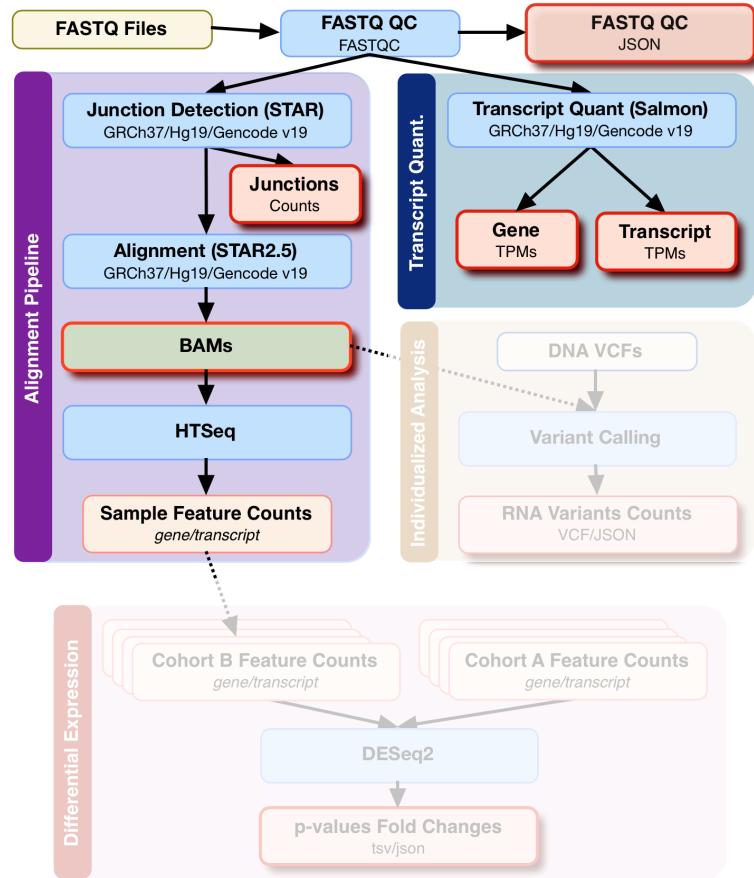
PRIMARY DATA PROCESSING & ANALYSIS

PPMI Long RNA (Current)

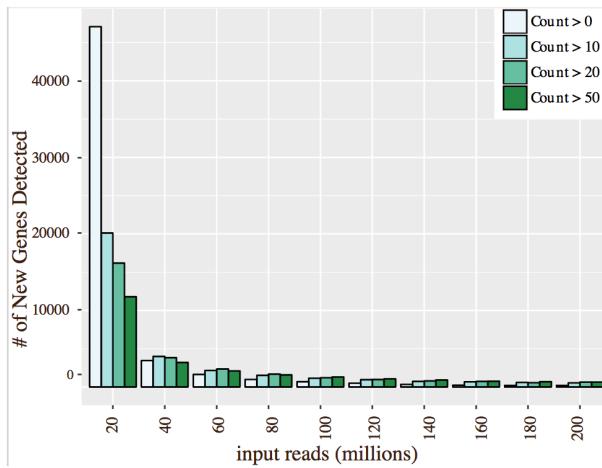
- Mirrors GTEX Gencode19/GRCh37
- Alignment: STAR
- Reference-free Transcript Quantification: Salmon
- Counts: Feature Count

Analysis (Nov+)

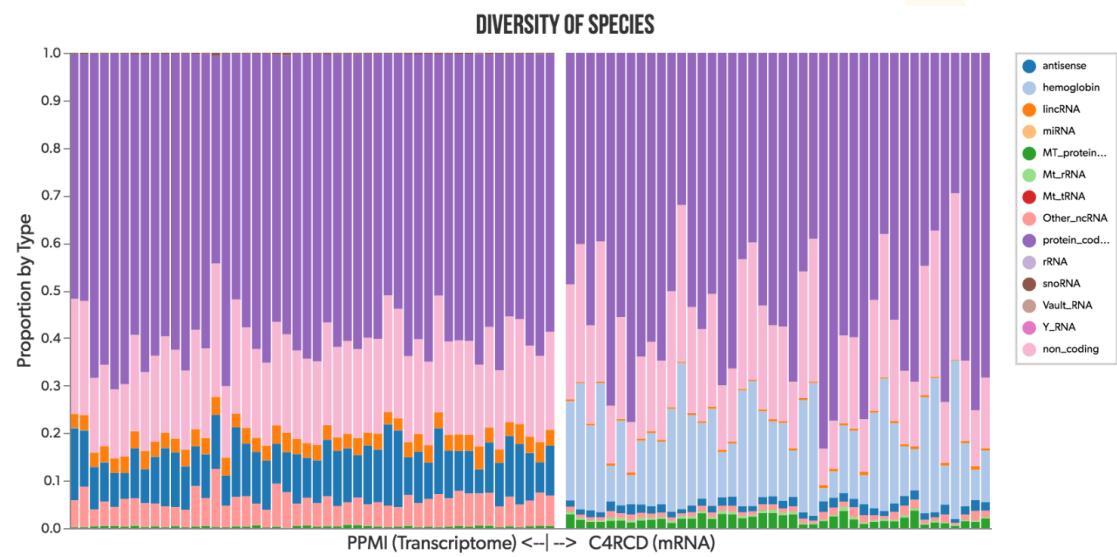
- Measurements
 - Gene and Transcript TPMs
 - Pre-computed p-values/log₂(FC) by transcript & gene
 - Junctions, Allele Counts at DNA
- Additional Analysis
 - Individual or personalized annotation (haplotype aware)
 - Allele Specific Expression
 - Non-sense mediated decay



BEYOND CODING



Value in Depth to 100M Pairs
(200M Reads)

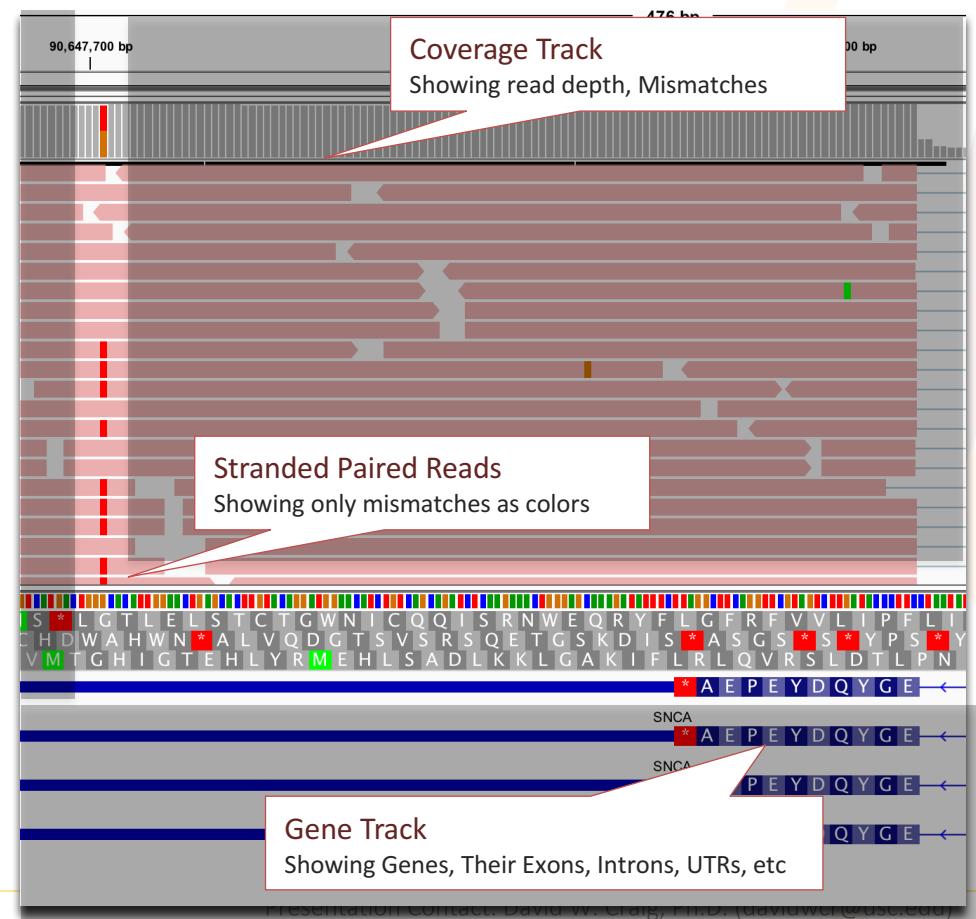


Transcriptome plus depletion provided greater diversity

TRANSCRIPTOME: BEYOND QUANTIFICATION

Raw Data View:

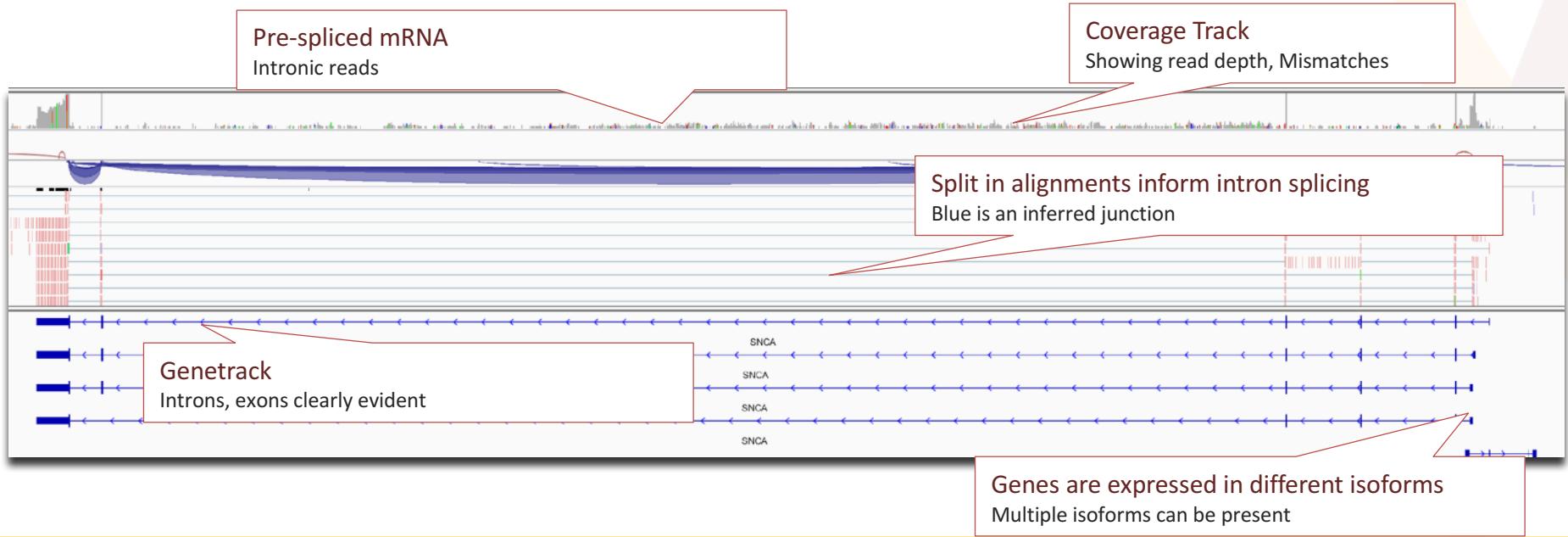
- Pre-spliced, Spliced, Strandedness
- Allele counts, etc.



TRANSCRIPTOME: BEYOND QUANTIFICATION (ZOOM OUT)

Transcript Quantification:

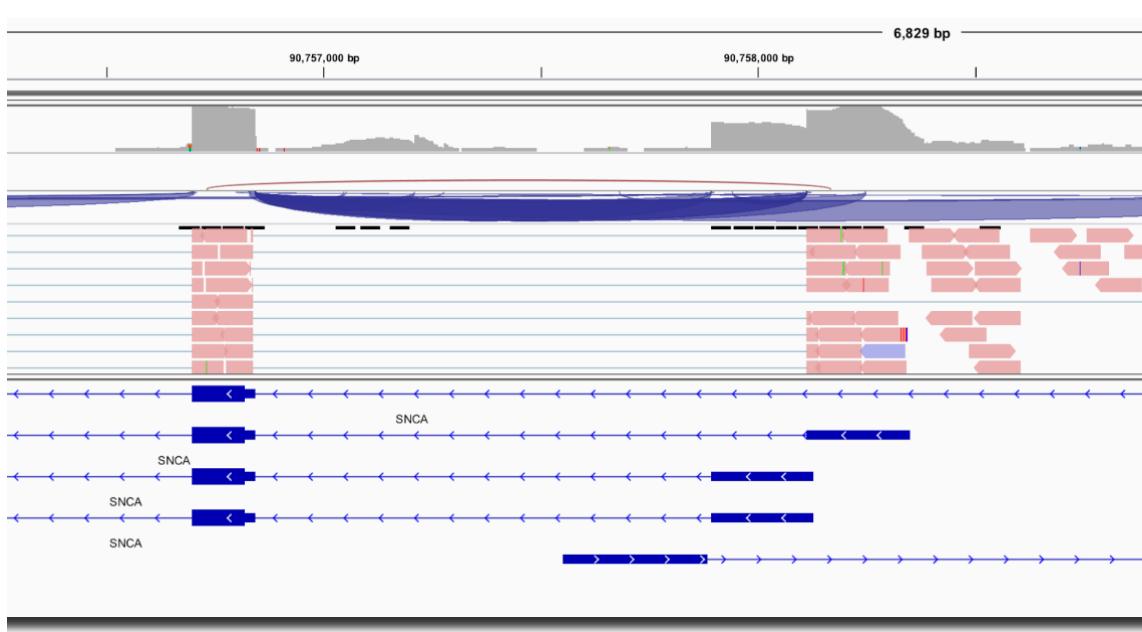
- Pre-spliced, Spliced, Strandedness



TRANSCRIPTOME: BEYOND QUANTIFICATION

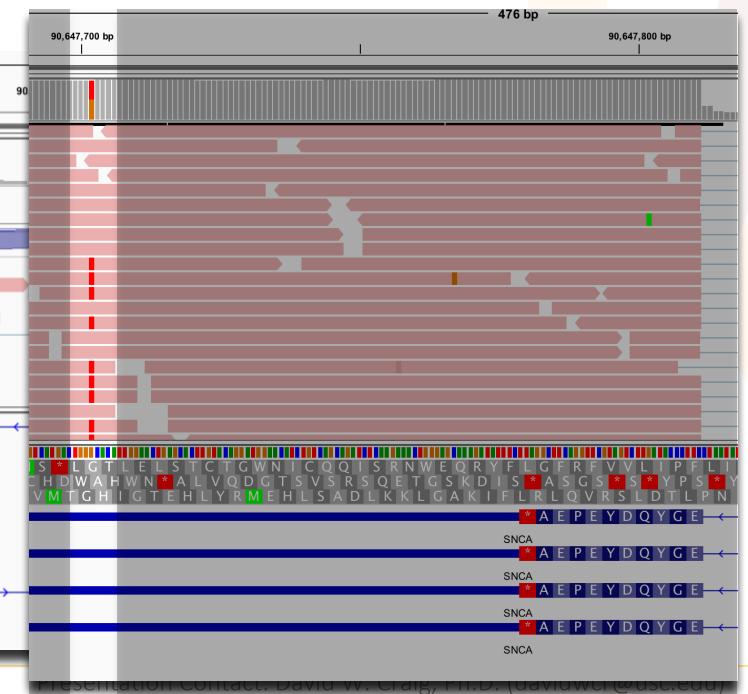
Junctions/Isoforms

- Alternative start-sites
- Integrated to individual w/ DNA



SNPs

- Allele specific expression
- Non-sense mediated decay, eQTLs



PER INDIVIDUAL: UP TO 1 BILLION READS @ 5 TIME POINTS

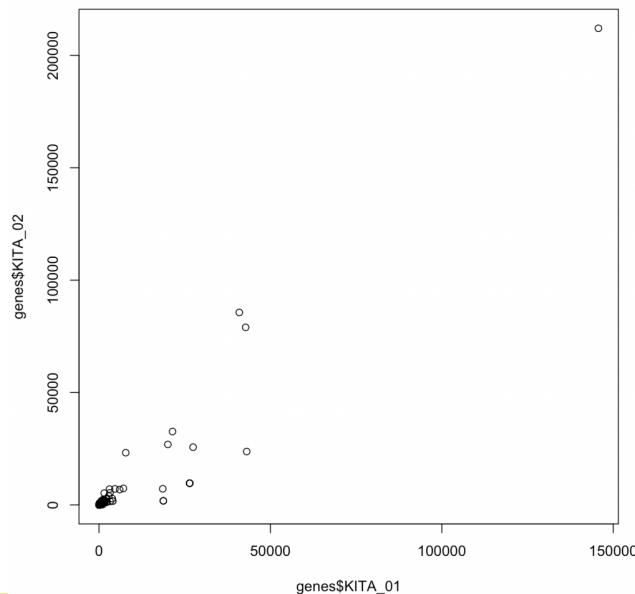


HETEROSCEDASTICITY –

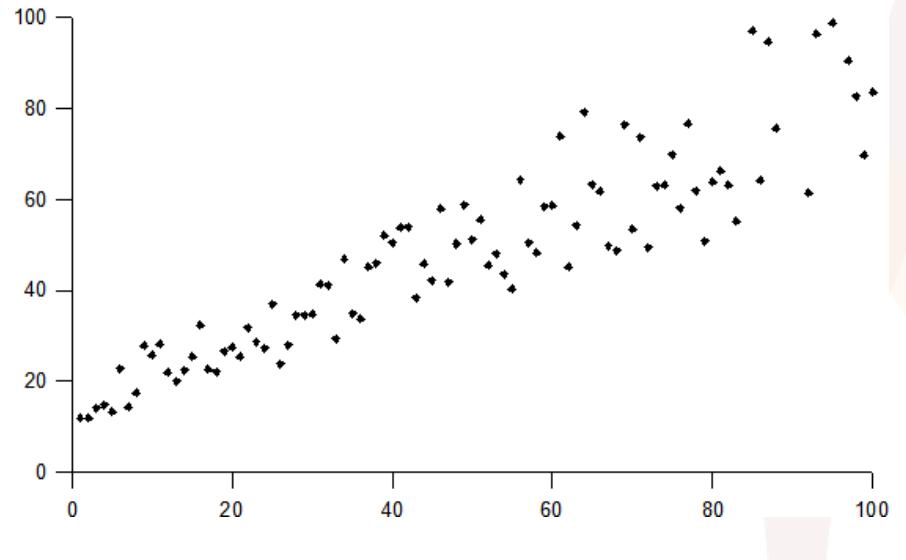
When error is proportional to the value, and when we are measuring over several orders of magnitude

Gene expression

- Typically, thousands of genes measured over several orders of magnitude.



Heteroscedasticity



DATA TRANSFORMATION: LOGGING

