

CAGI slide pack guidelines

I agree to my slides being used according to the following terms: *Scott Adamson*

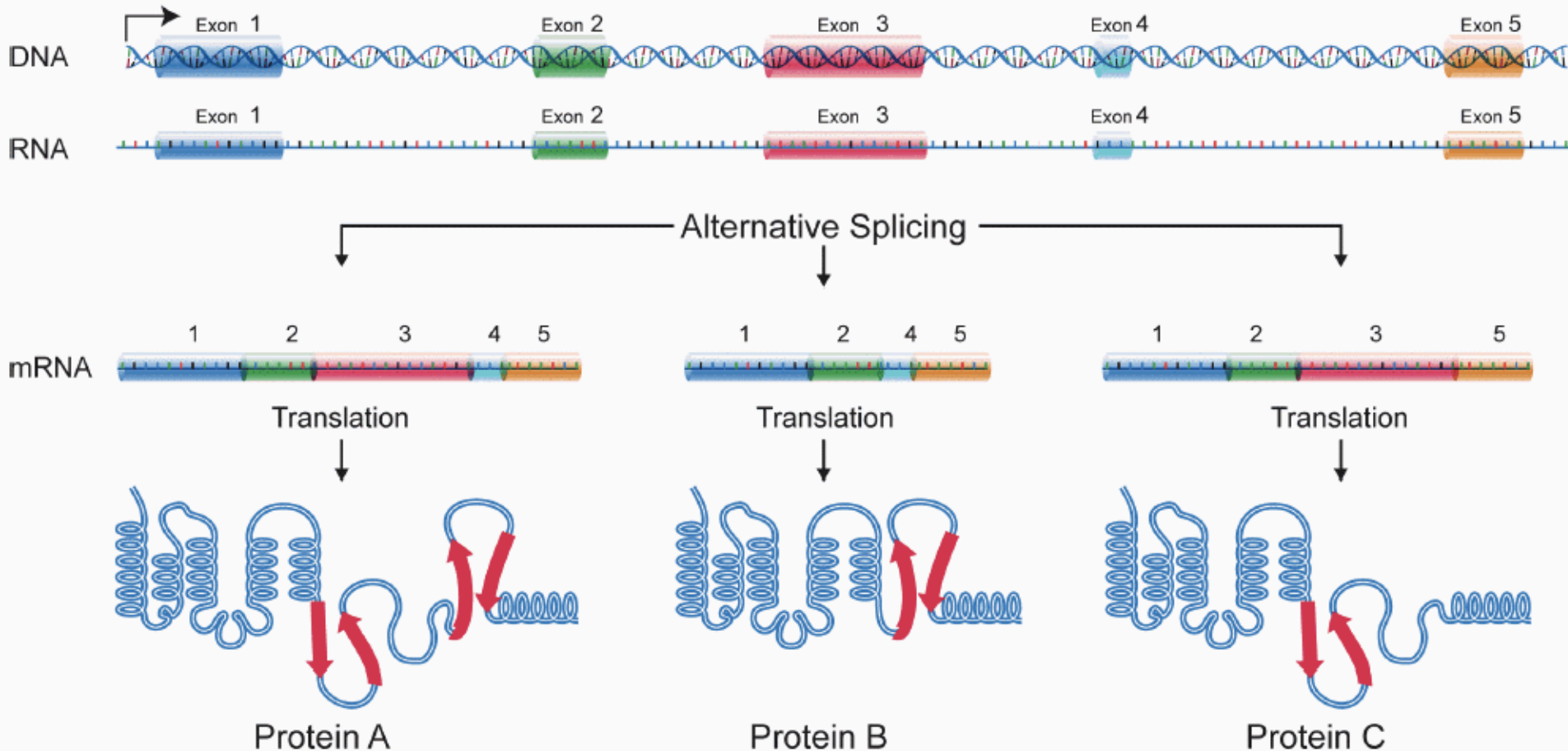
- Authors: please include THIS slide at the beginning of your presentation to indicate that you agree with reuse of your slides in CAGI presentations, according to the terms below, and digitally sign by writing your name above.
- To disseminate the CAGI results, I encourage all CAGI participants to remix these CAGI slides for inclusion in their oral and poster presentations
- CAGI participants may use or publish any content of these slides only in compliance with the CAGI Data Use Agreement
- CAGI participants may not cite or publish any content of these slides except with the written permission of the originating author or in the primary CAGI publications
- CAGI participants must acknowledge the original authors of these slides and CAGI 2018. Include the acknowledgement banner across the bottom
- CAGI participants must include the CAGI credits slide in their presentations
- Slides with a red slash (no remix) may not be included in a presentation unless all attendees are registered CAGI participants who signed the data use agreement
- Slides with a red X over the entire slide may not be used in any circumstances
- Slides with a blue slash (embargoed) may not be used unless the embargo is explicitly lifted
- Slides without slash or X may be tweeted unless flagged with the “No tweet” icon.
- CAGI participants using these slides must explain the “No tweet” icon, which means that the slide content should not be disseminated outside the presenting conference hall. If the presentation venue permits pervasive tweeting, you may not include these slides in your talk
- Note for slide authors: we expect all CAGI participants to abide by these restrictions, and will make best effort to ensure they are followed. However, but bear in mind that we have limited means of enforcing them, and therefore the restrictions cannot be guaranteed.



Vex-seq: High-Throughput Identification of the Impact of Genetic Variation on pre-mRNA Splicing Efficiency

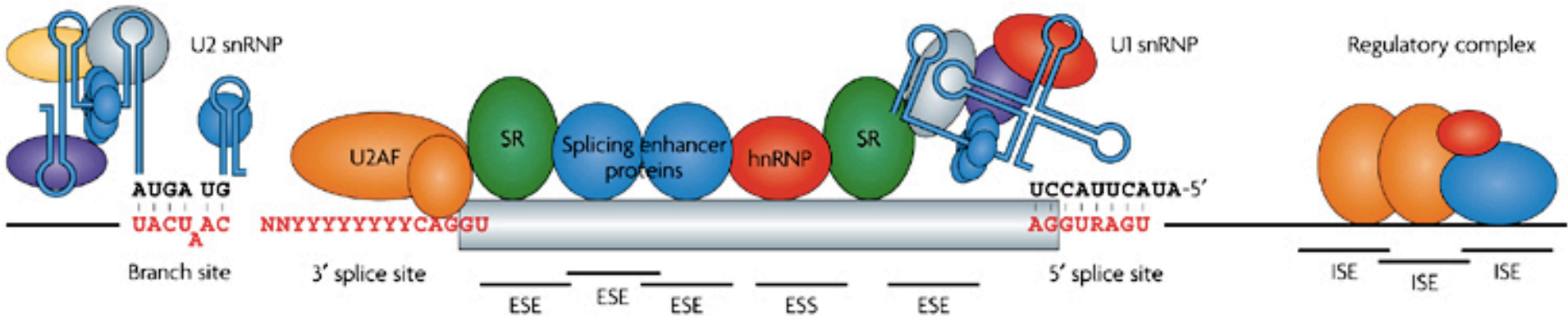
Scott I. Adamson, Lijun Zhan, and
Brenton R. Graveley
UConn Health

pre-mRNA Splicing



(NHGRI)

Splicing is Regulated by Different Proteins and Complexes



Nature Reviews | Genetics

(Wang & Cooper, 2007)

- Spliceosome machinery
- Exonic/Intronic splice enhancers and silencers

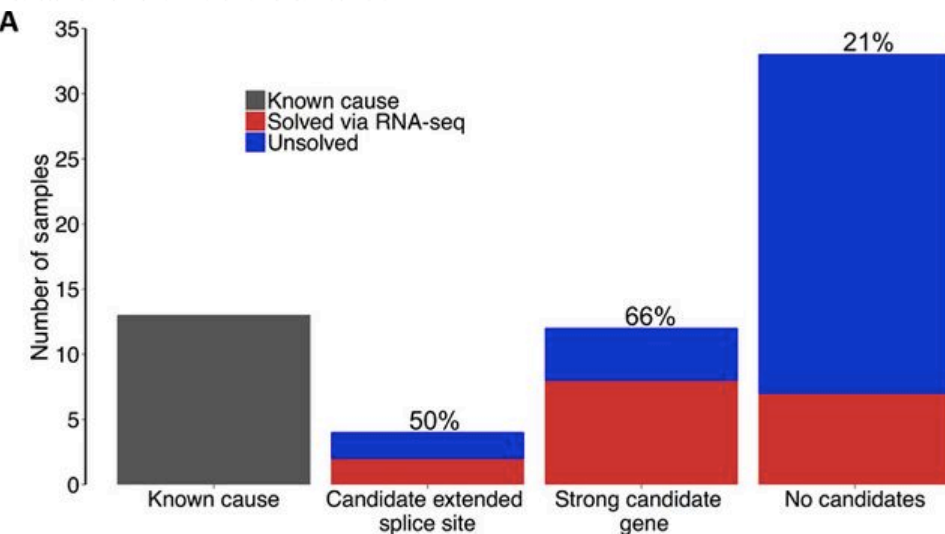
Aberrant Splicing Contributes to and Underlies Disease Phenotypes

Evidence of selection on splicing-associated loci in human populations and relevance to disease loci mapping

Eric R. Gamazon , Anuar Konkashbaev, Eske M. Derks, Nancy J. Cox & Younghee Lee 

Improving genetic diagnosis in Mendelian disease with transcriptome sequencing

Beryl B. Cummings^{1,2,3}, Jamie L. Marshall^{1,2}, Taru Tukiainen^{1,2}, Monkol Lek^{1,2,4,5}, Sandra Donkervoort⁶, A. Reghan Foley⁶, Veronique Bolduc⁶, Leigh B. Waddell^{4,5}, Sarah A. Sandaradura^{4,5}, Gina L. O'Grady^{4,5}, Elicia Estrella⁷, Hemakumar M. Reddy⁸, Fengmei Zhao^{1,2}, Ben Weisburd^{1,2}, Konrad J. Karczewski^{1,2}, Anne H. O'Donnell-Luria^{1,2}, Daniel Birnbaum^{1,2}, Anna Sarkozy⁹, Ying Hu⁶, Hernan Gonozakzy¹⁰, Kristl Claeys¹¹, Himanshu Joshi⁵, Adam Bournazos^{4,5}, Emily C. Oates^{4,5}, Roula Ghaoui^{4,5}, Mark R. Davis¹², Nigel G. Laing^{12,13}, Ana Topf¹⁴, Genotype-Tissue Expression Consortium, Peter B. Kang^{7,8}, Alan H. Beggs⁷, Kathryn N. North¹⁵, Volker Straub¹⁴, James J. Dowling¹⁰, Francesco Muntoni⁹, Nigel F. Clarke^{4,5,*}, Sandra T. Cooper^{4,5}, Carsten G. Bönnemann⁶ and Daniel G. MacArthur^{1,2,†}



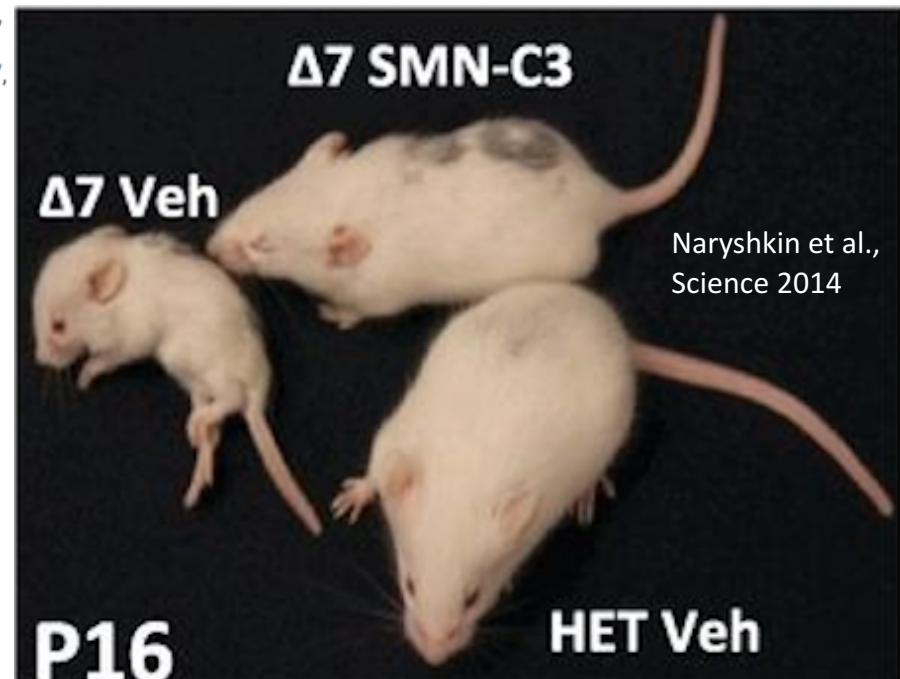
Created for CAGI 2018

genomeinterpretation.org

RNA splicing is a primary link between genetic variation and disease

Yang I. Li,¹ Bryce van de Geijn,² Anil Raj,¹ David A. Knowles,^{3,4} Allegra A. Petti,⁵ David Golan,¹ Yoav Gilad,^{2*} Jonathan K. Pritchard^{1,6,7*}

Spinal Muscular Atrophy SMN2 gene exonic mutation



Scott Adamson

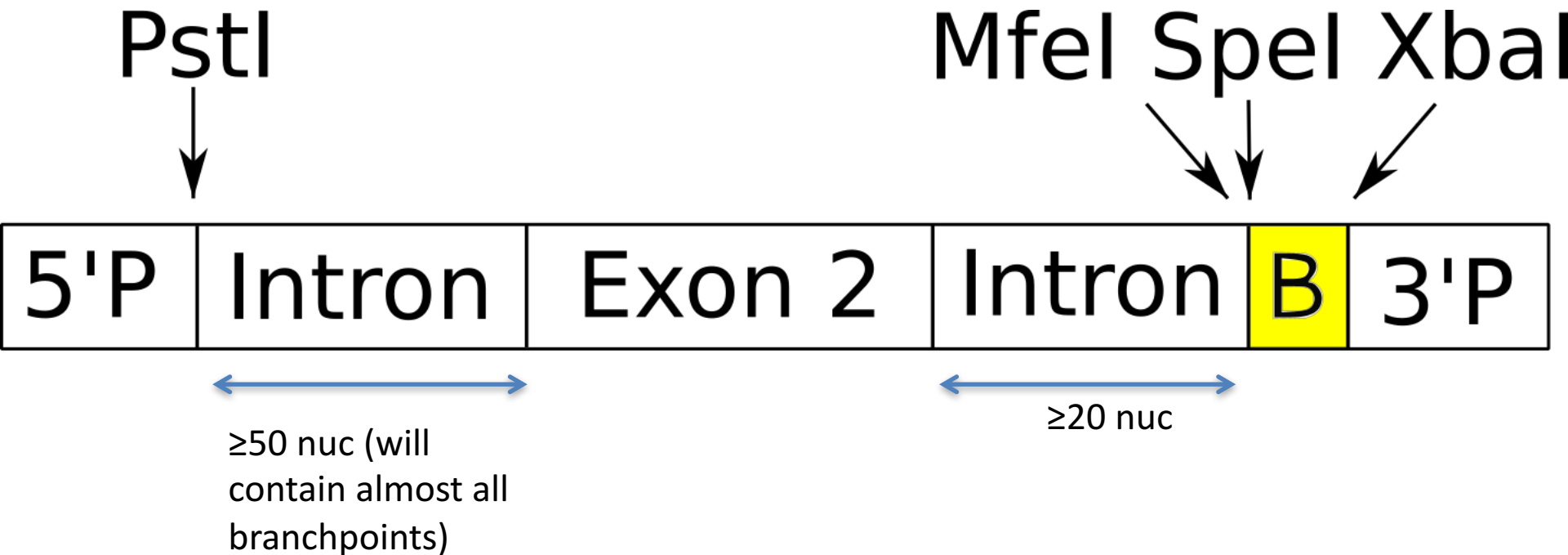
Specific Aims

- Aim 1: Develop, construct and test reporter splicing reporter assay to simultaneously test the impact of thousands of intronic and exonic variants on splicing efficiency (Variant exon sequencing/Vex-seq)
- Aim2: Analyze splicing outcomes
 - Provide insight into candidate mechanisms of variant induced splicing changes
 - Identify potential biases and strengths of approach

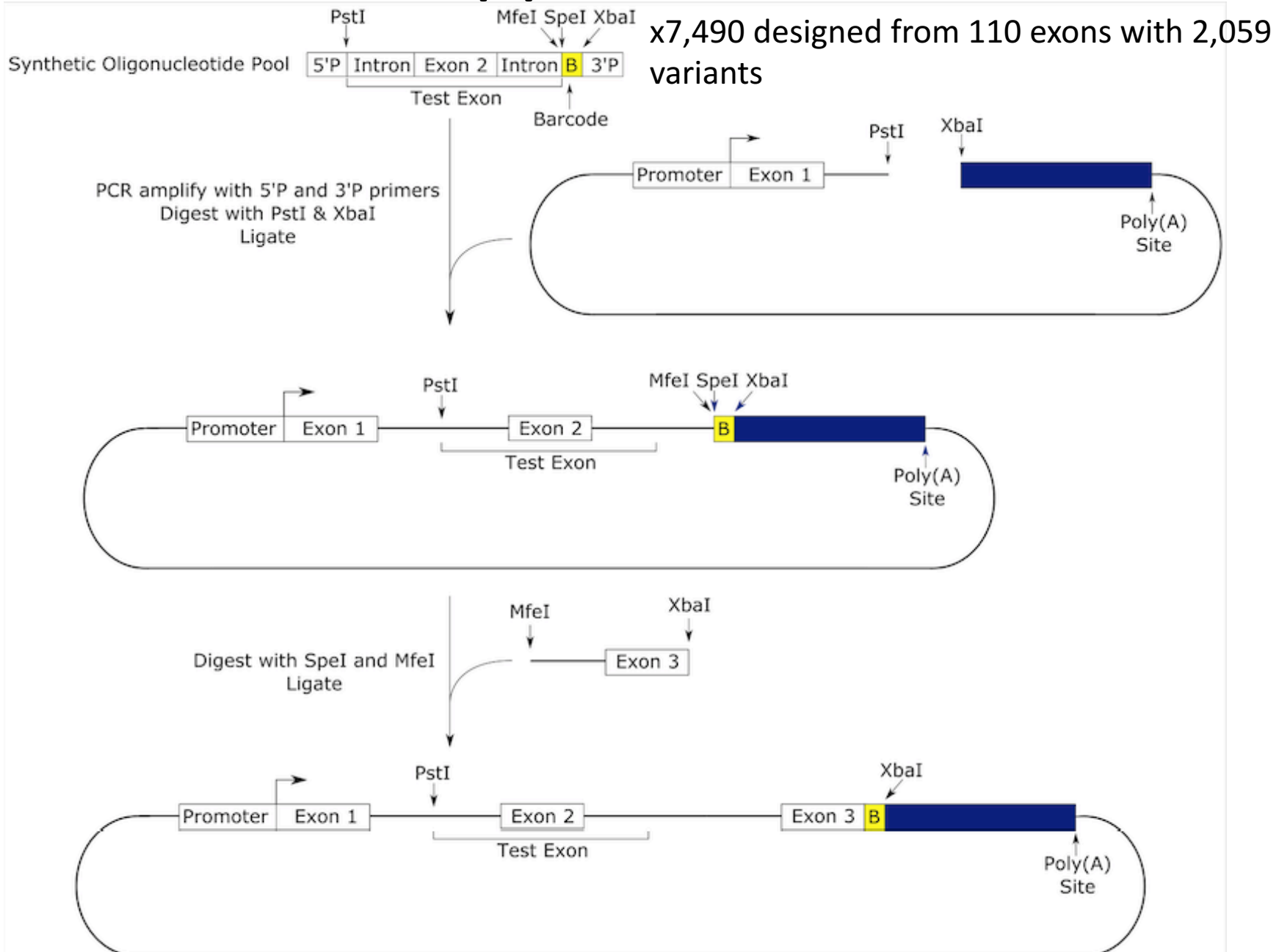
Test Exon Design

- Exon selection (110 total exons)
 - Cassette exons based on Gencode v19 annotations
 - 68-97 bases long (limited by synthesis)
- Variant selection (2,059 variants)
 - Based on ExAC database (Lek et al., Nature 2016)
- Controls
 - Reference sequences
 - Barcodes in technical triplicates
 - Splice site sequence controls
- Synthesis platform
 - Agilent assay synthesis 7,500 oligos x 230 nucleotides long

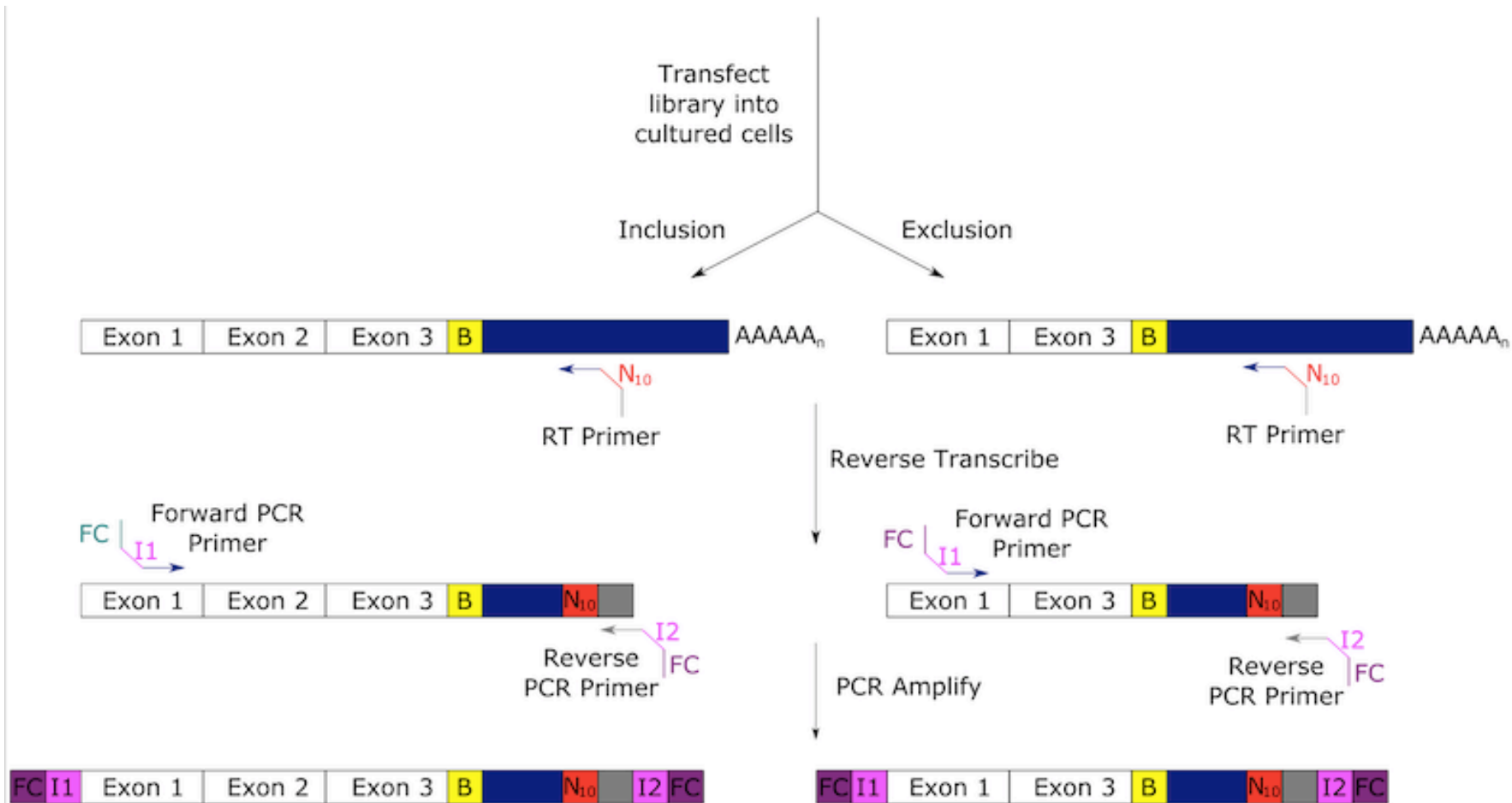
Test Exon Design



Approach



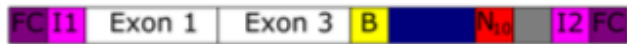
Approach



Computational Pipeline of Results



+



Assign Reads to Cell Line
with Indices (I1 and I2)



Filter Out Reads Without
Exon 3



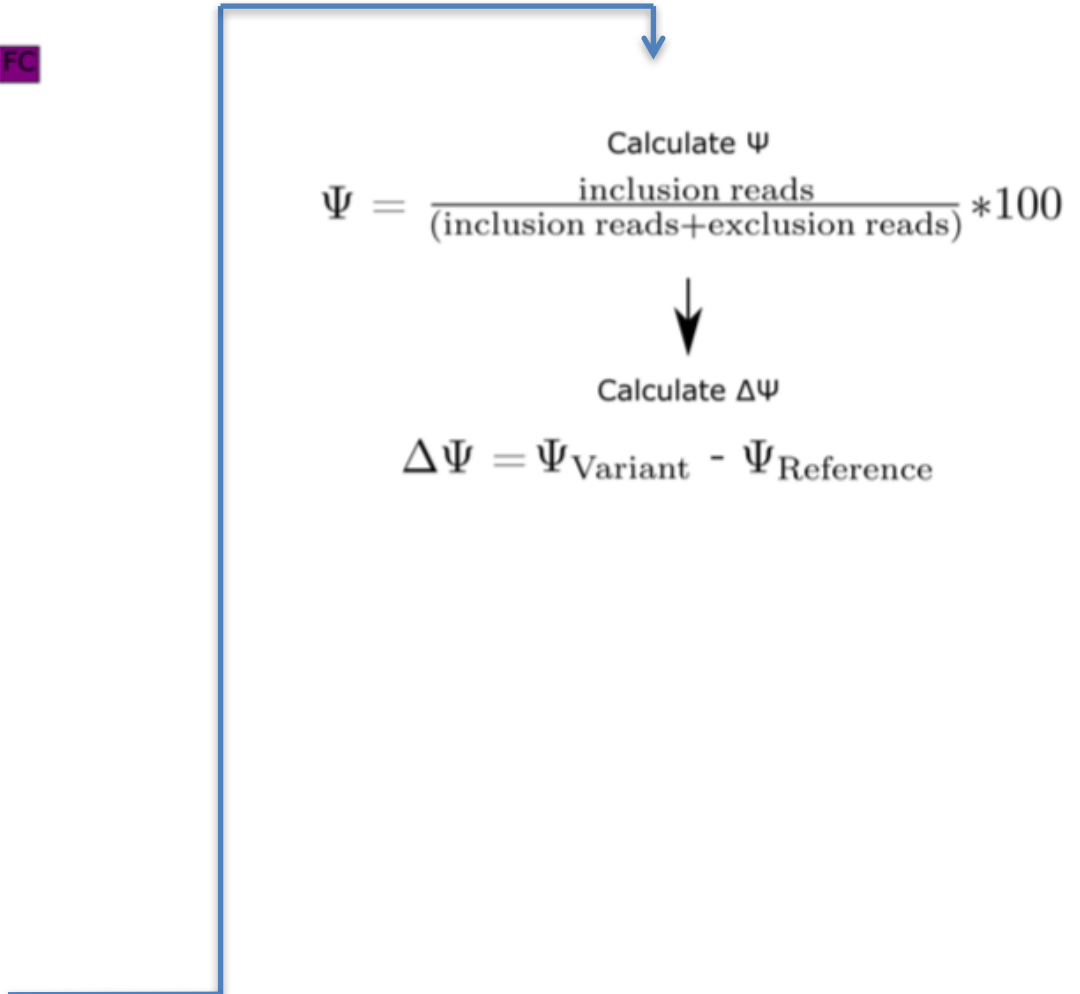
Bin Reads by Barcodes (B)



Collapse PCR Duplicates with
10 Nucleotide UMI (N10)
and Barcode

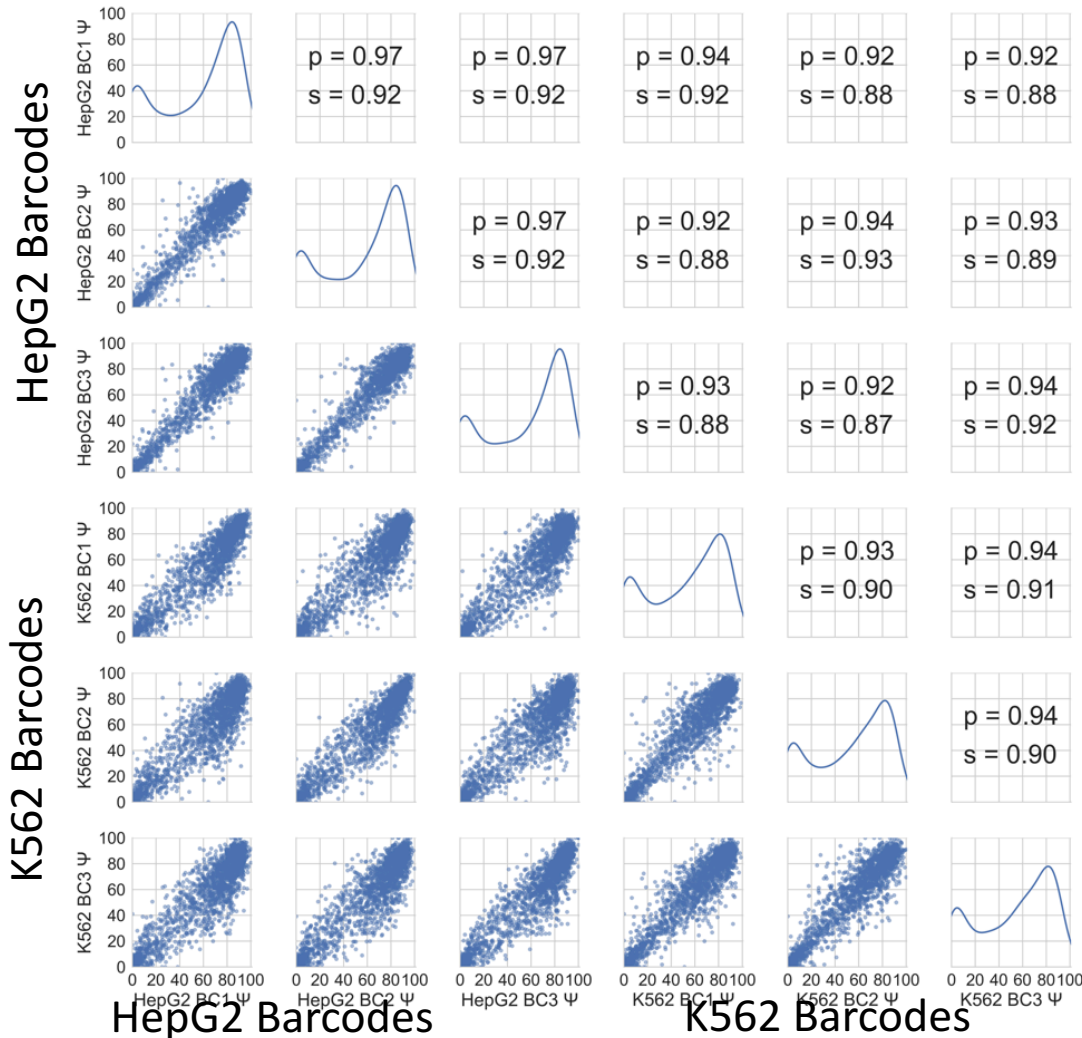


Align Reads to Reference



Splicing Quality Control: Barcode

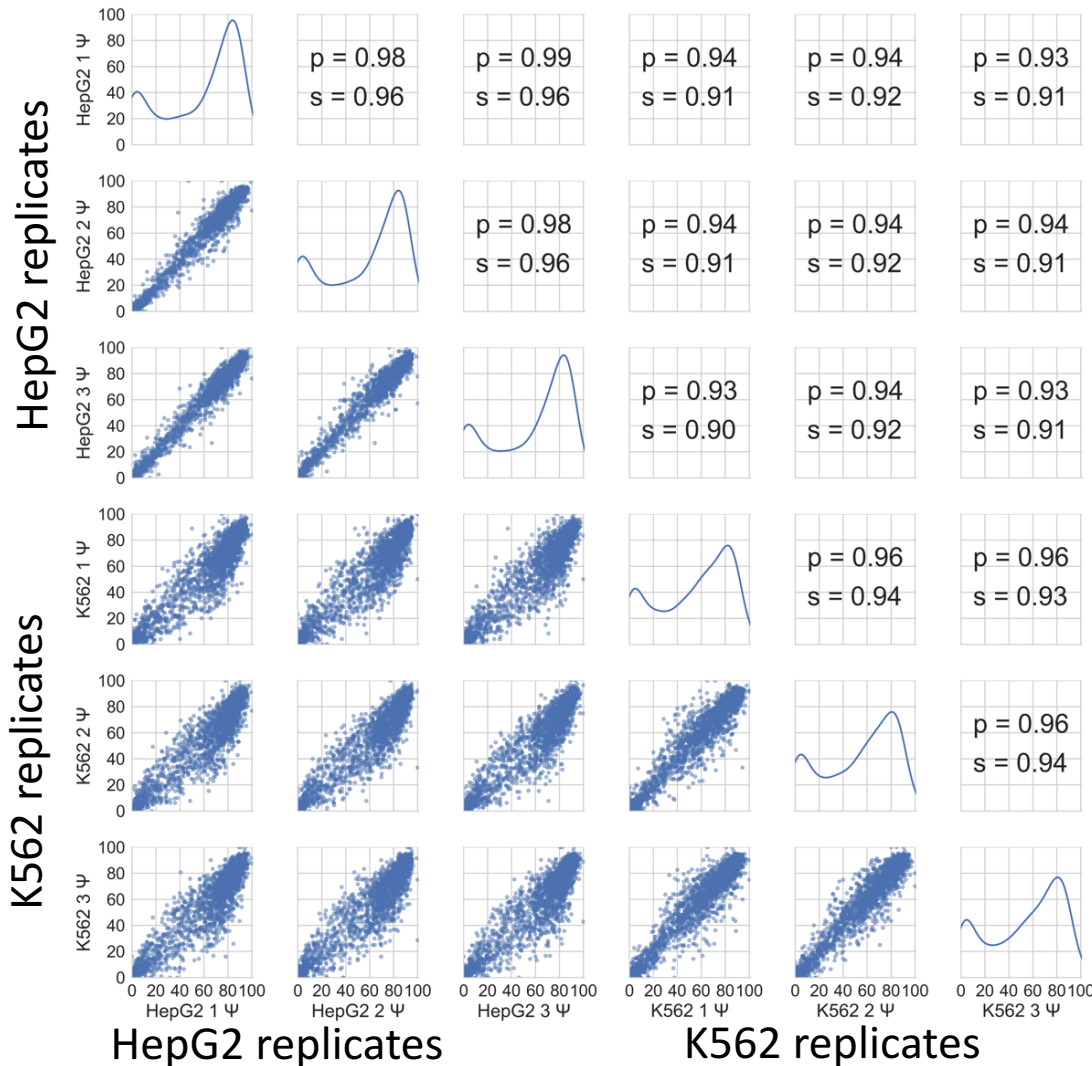
Replicate Splicing is Well Correlated



p = pearson correlation coefficient
s = spearman correlation coefficient

$$\Psi = \frac{\text{inclusion reads}}{(\text{inclusion reads} + \text{exclusion reads})} * 100$$

Splicing Quality Control: Biological Replicate Splicing is Well Correlated

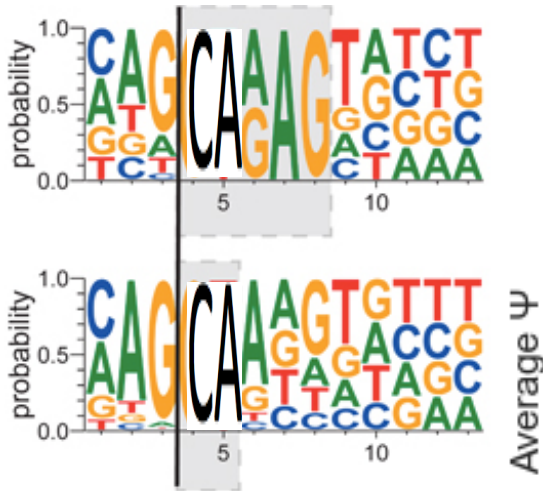


p = pearson correlation coefficient
s = spearman correlation coefficient

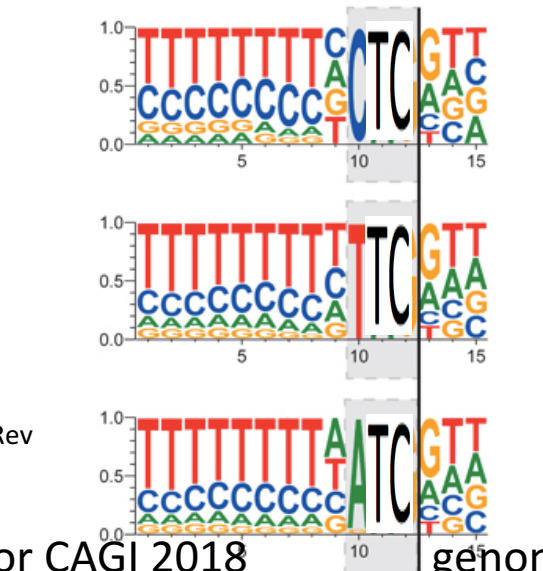
$$\Psi = \frac{\text{inclusion reads}}{(\text{inclusion reads} + \text{exclusion reads})} * 100$$

Positive and Negative Controls Exhibit Expecting Splicing Behavior

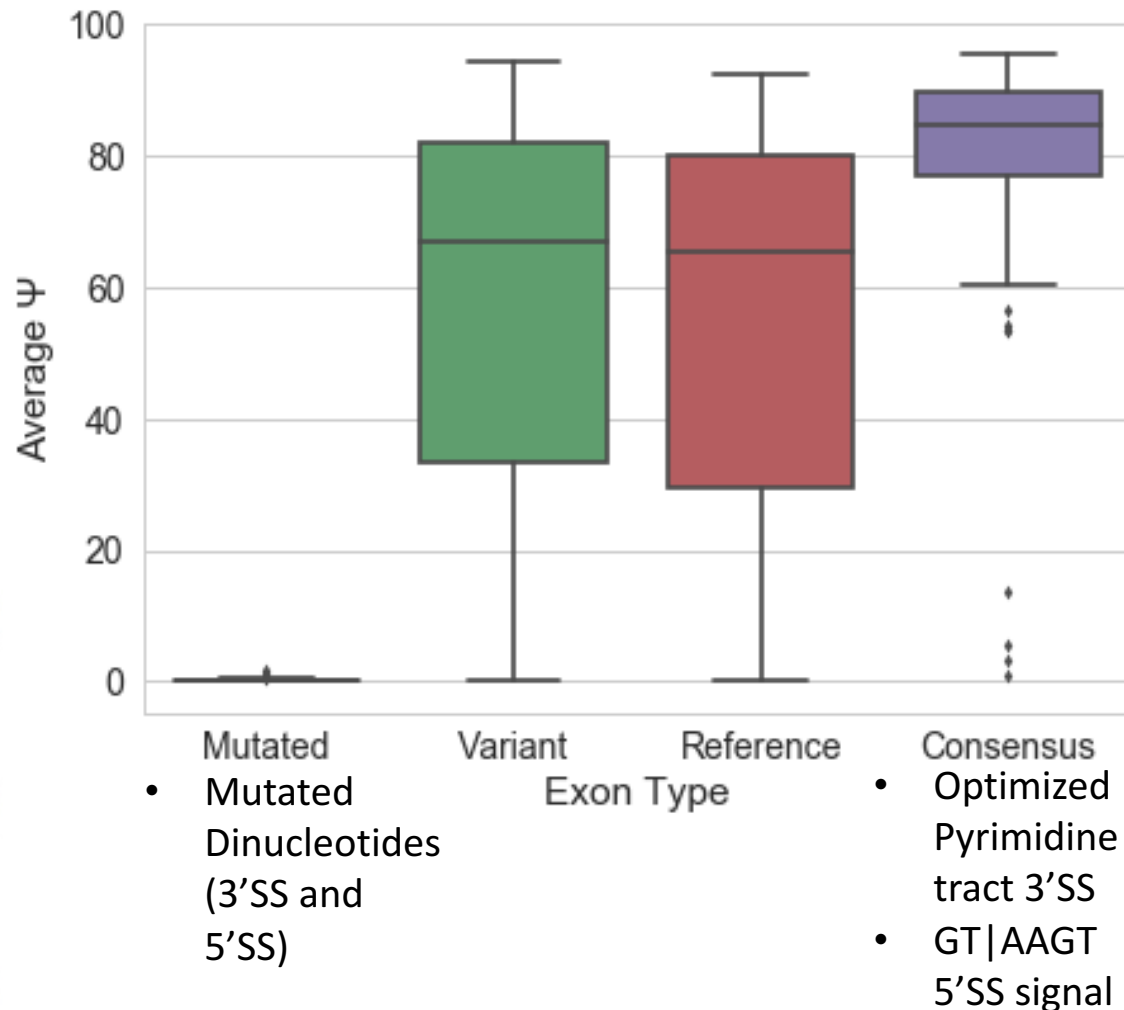
5' Splice Sites



3' Splice Sites

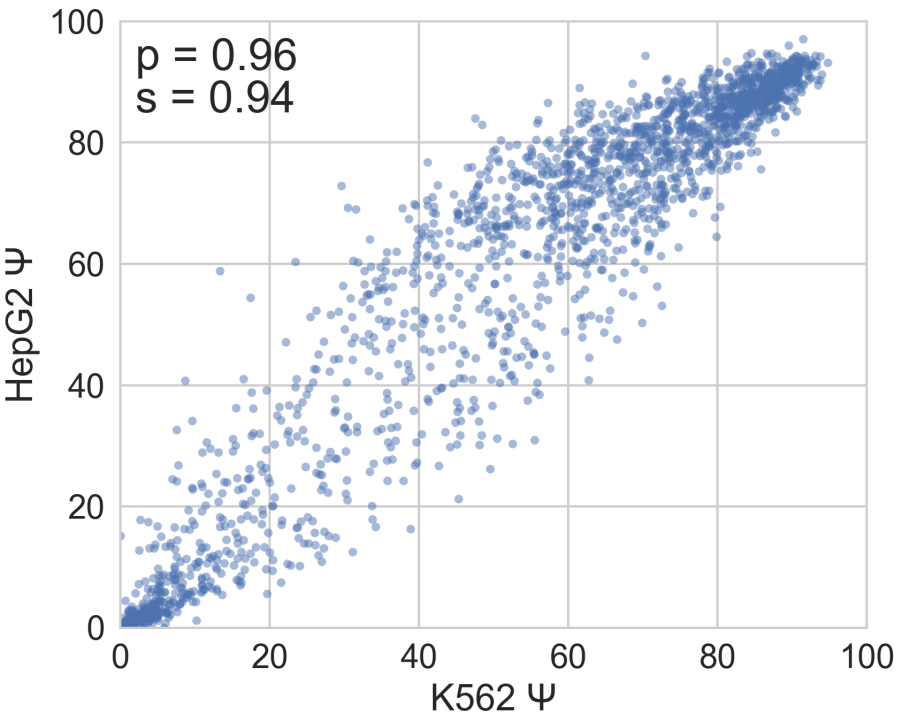


Adopted from:
Sibley CR et al., Nat Rev
Genet; 2016

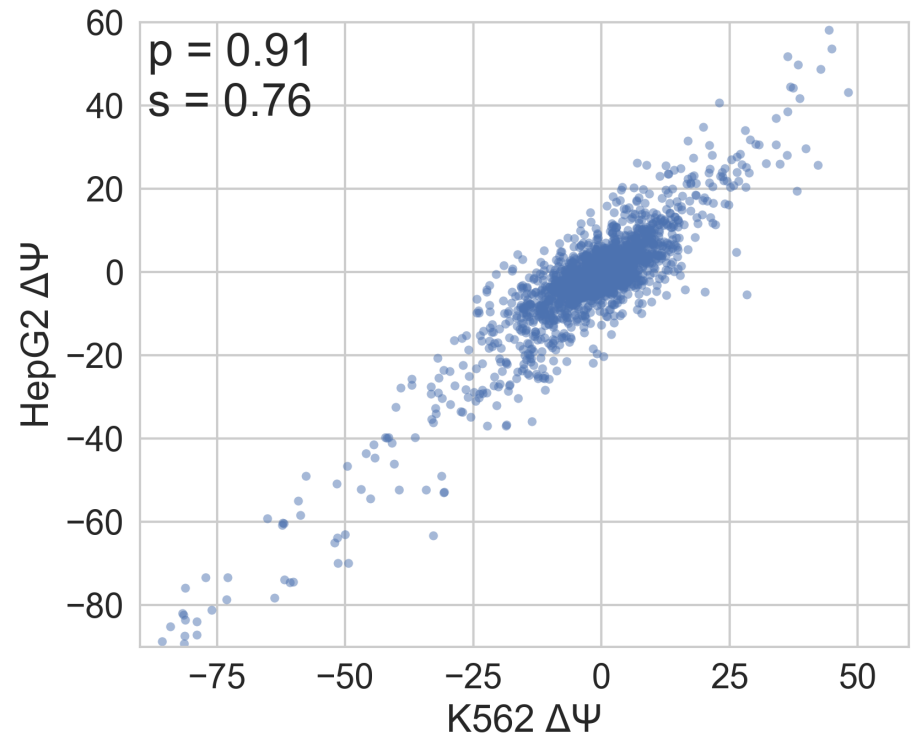


Similar Splicing Behaviors Between K562 and HepG2 Contexts

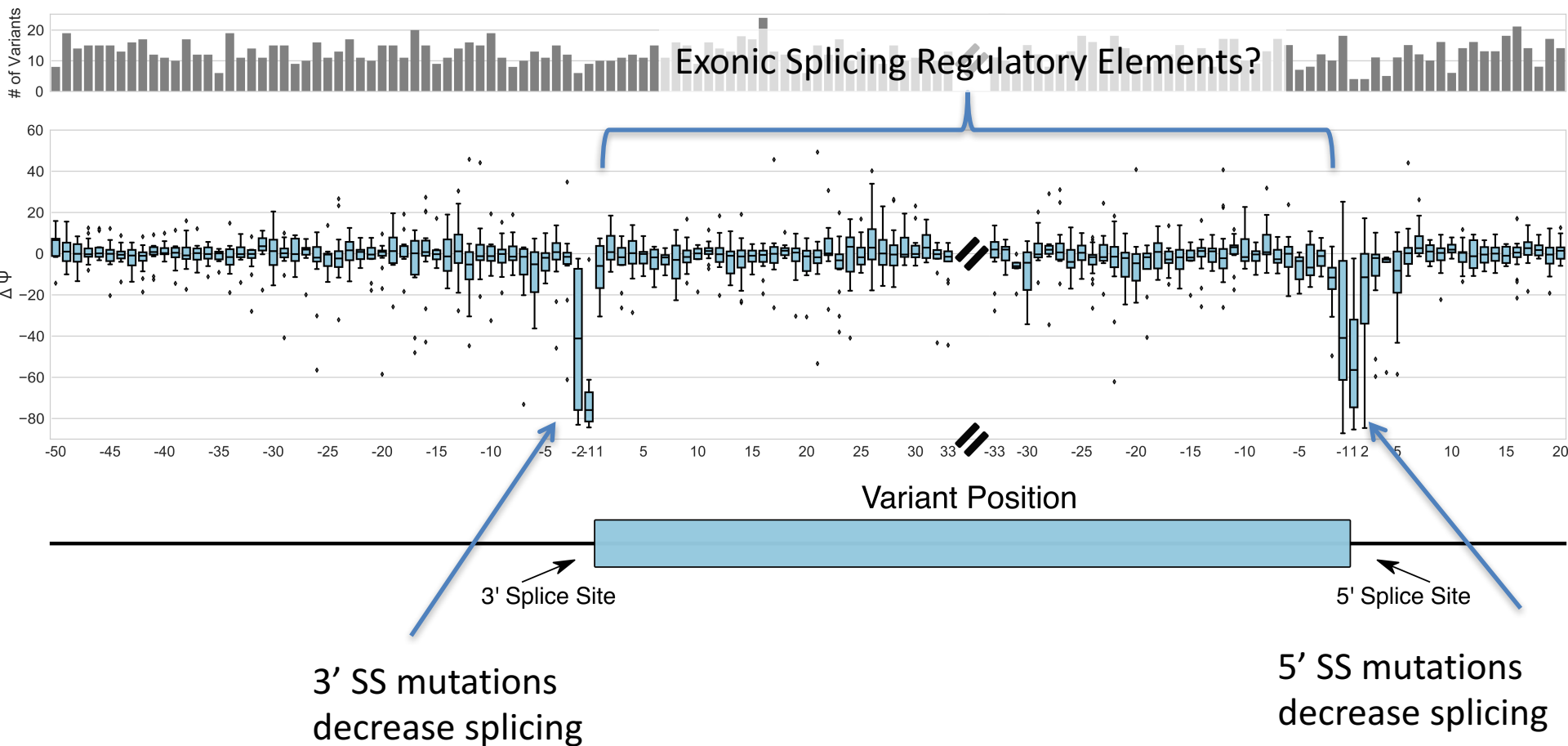
Ψ for Variant and Reference Sequences



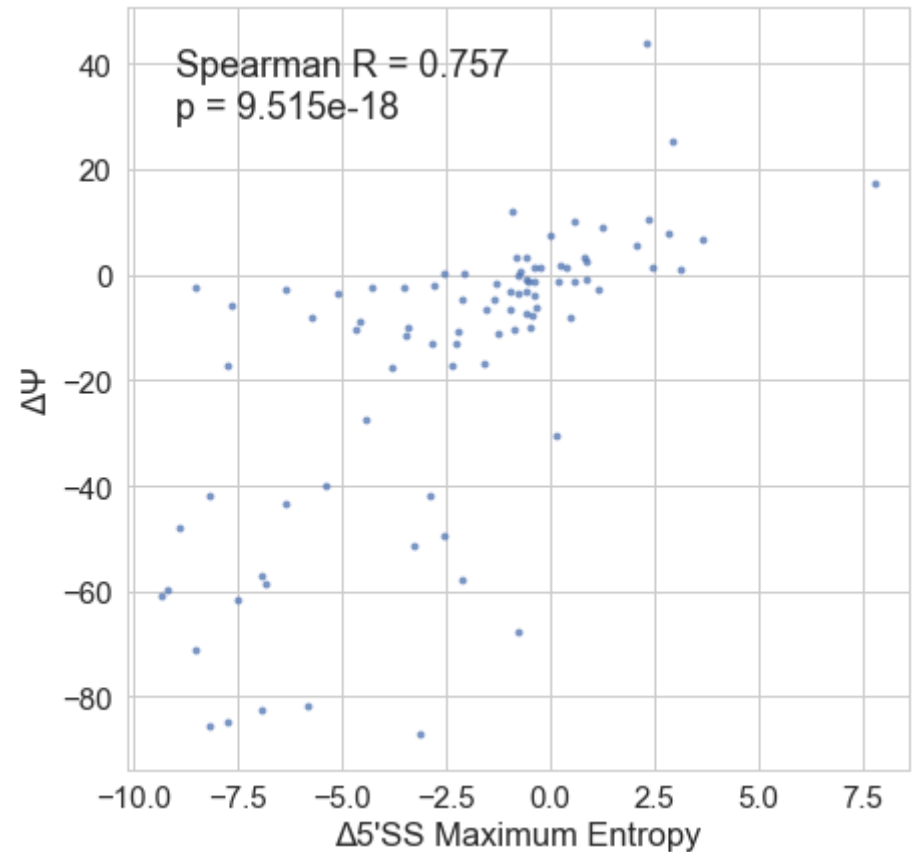
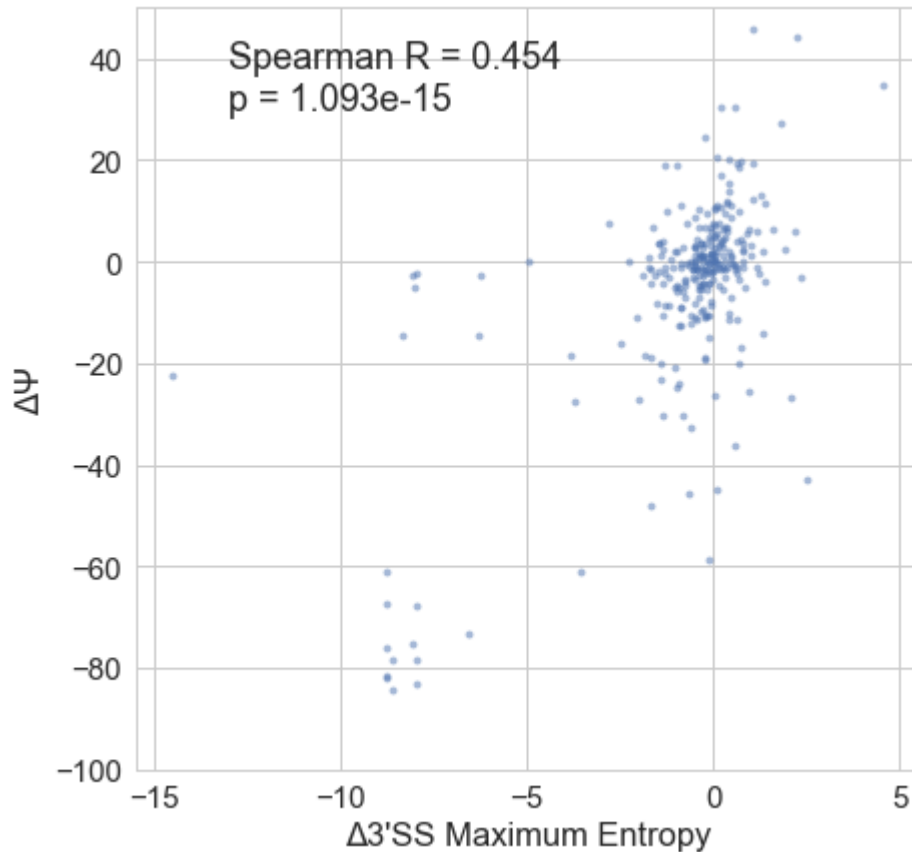
$\Delta\Psi$ for Variant Sequences



Changes in Splicing Result from Predictable Splice Site Alterations and Other Sources



Splice Site Strength is Significantly Correlated for to Splicing Changes at Splice Sites

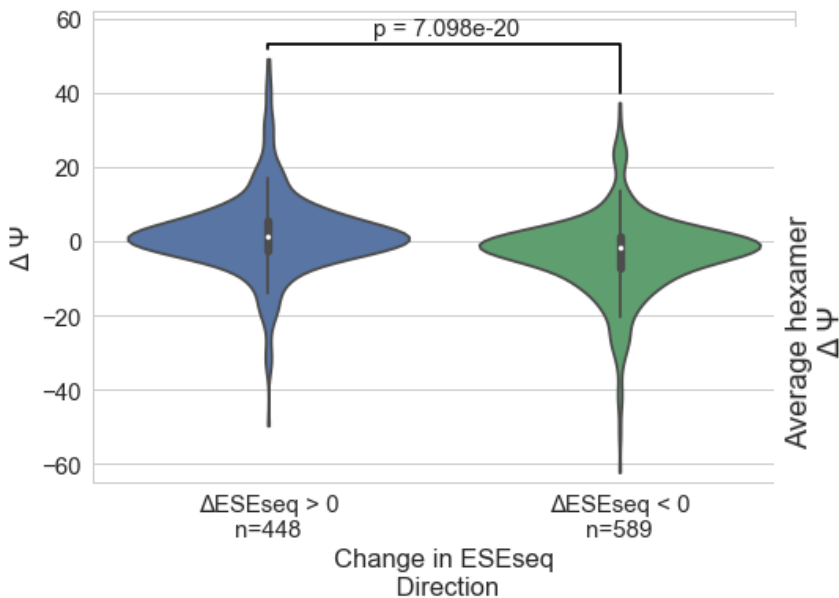


MaxEnt Scan. Yeo & Burge. 2004

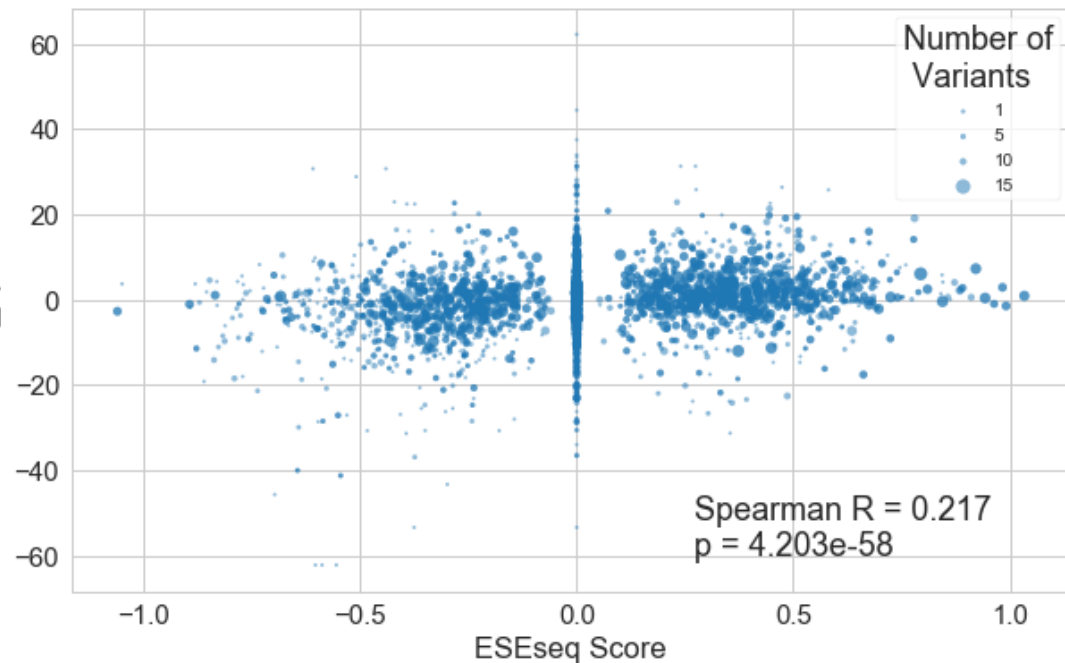
Exonic Splicing Regulatory Elements are Significantly Correlated with Splicing Changes

ESEseq dataset provides enrichment of all hexamers as exonic splicing enhancers or silencers (Ke et al., Genome Res; 2011)

Change in ESEseq Score (from reference to variant)



ESEseq score for each hexamer associated with splicing behavior of gained and lost hexamers between reference and variant sequences



Advantages

- Able to test intronic variants
- Known number of total reads for each construct
- Sequencing DNA construct, while useful is not necessary

Disadvantages

- Variant sampling differences*
- Non-sense Mediated Decay*
- Unanticipated splicing outcomes
- Limited reporter context

*More analysis in paper

Conclusions

- Successfully constructed a splicing reporter assay to interrogate the impact of variants on splicing behavior
- Investigated broad level influences from exonic splicing regulatory elements and splice site strength
- Investigated conservation of loci that change splicing

Technical CAGI Specific Notes

- Results in the paper (and presented here) reflect the average splicing behavior of K562 and HepG2 splicing outcomes
 - CAGI test was only on HepG2 data
- After training and test data were submitted to CAGI we did some barcode QC and filtered out a small number of barcodes (~1.82%)
 - I don't imagine this will change results qualitatively, but may explain some minor differences between CAGI $\Psi/\Delta\Psi$ values and published values

Interested in Studying K562 Data or Other Analysis?

- Paper:
 - <https://doi.org/10.1186/s13059-018-1437-x>
- Github (dig into pre-processed data / K562 $\Delta\Psi$ values)
 - <https://github.com/scottiadamson/Vex-seq>
- GEO (raw data available)
 - <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113163>



Thanks

Graveley Lab:

Lijun Zhan*

Sara Olson

Ahmad Osman

Sandy Garrett, PhD

Shane Lawson

Xintao Wei, PhD

Gopi Rajadinakaran, PhD

Former member: Michael Duff, PhD

Jeff Chuang Lab

UConn Health Biomedical Sciences

Graduate Program

JAX-GM

Funding: NHGRI R21 HG008799

John and Donna Krenicki

Endowment Fund to BRG

Scott Adamson

The UConn Health logo, featuring the words "UConn" and "HEALTH" in white, bold, sans-serif capital letters on a dark blue square background.

UConn
HEALTH

CAGI Organizers
CAGI Conference

Citations

- GS Wang and TA Cooper, Nat Rev Gen. 2007. <https://doi.org/10.1038/nrg2164>
- BB Cummings et al., Sci Transl Med. 2017. <https://doi.org/10.1126/scitranslmed.aal5209>
- ER Gamazon et al., 2017. Sci Rep <https://doi.org/10.1038/s41598-017-05744-9>
- YI Li et al., Science. 2016. <http://dx.doi.org/10.1126/science.aad9417>
- NA Naryshkin et al., Science 2014. <https://doi.org/10.1126/science.1250127>
- M Lek et al., Nature 2016. <http://dx.doi.org/10.1038/nature19057>
- CR Sibley et al., Nat Rev Genet 2016. <https://doi.org/10.1038/nrg.2016.46>
- W McLaren et al., Genome Biol. 2016. <http://dx.doi.org/10.1186/s13059-016-0974-4>
- G Yeo & CB Burge, J Comput Biol. 2004. <http://dx.doi.org/10.1089/1066527041410418>
- S Ke et al., 2011, Genome Res. 2011. <http://dx.doi.org/10.1101/gr.119628.110>
- KS Pollard et al., Genome Res. 2010. <https://doi.org/10.1101/gr.097857.109>