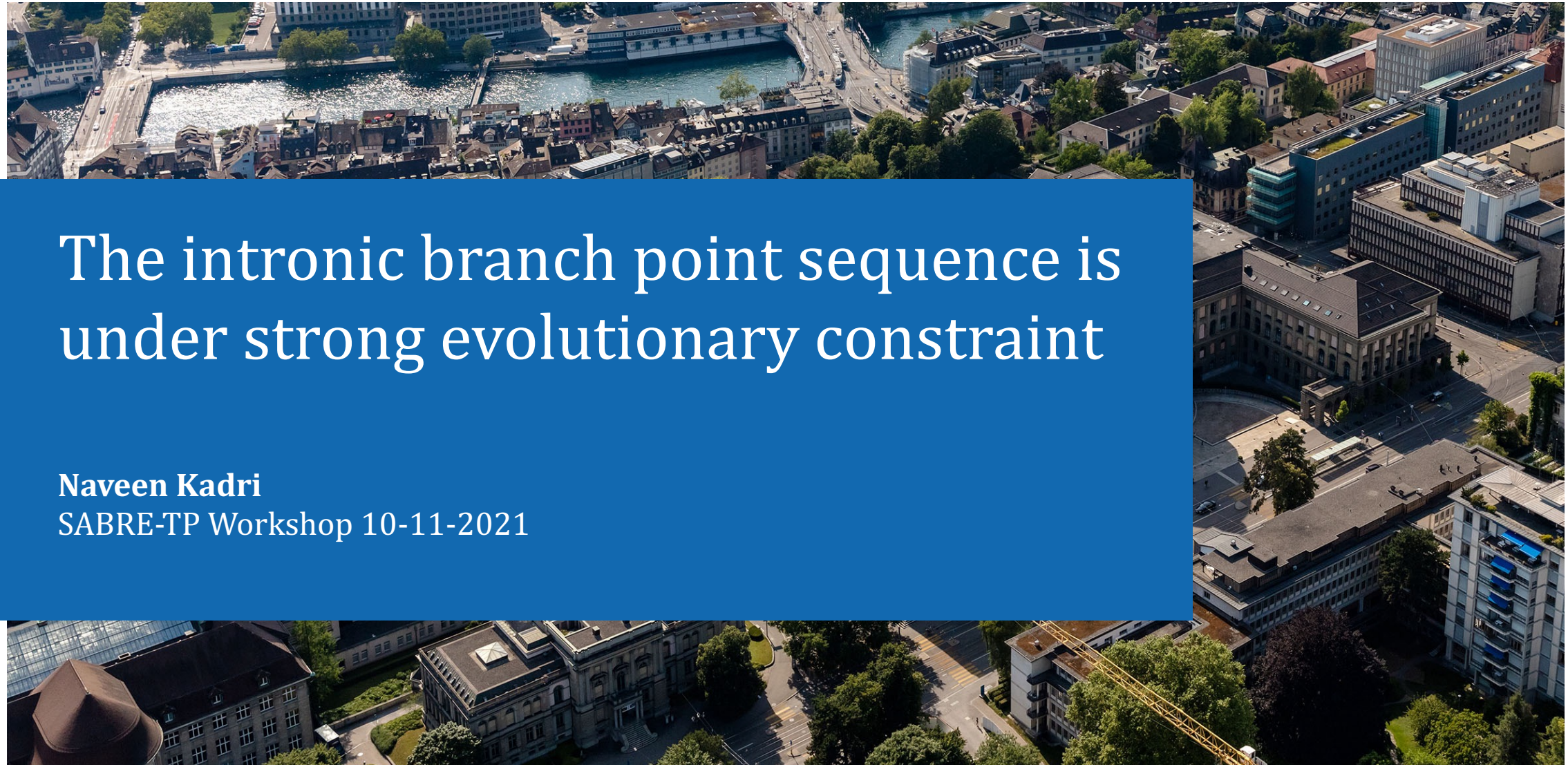
ETH zürich

D-USYS
Institute of Agricultural Sciences



The intronic branch point sequence is under strong evolutionary constraint

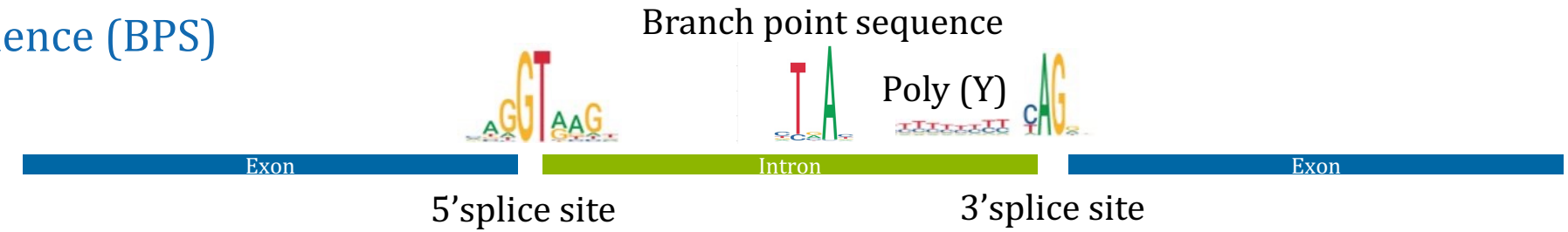
Naveen Kadri

SABRE-TP Workshop 10-11-2021

Introduction

- Noncoding sequences are an important source of genetic variation
  - Regulatory elements → gene expression & alternative splicing
  - QTL / significant contribution to h^2

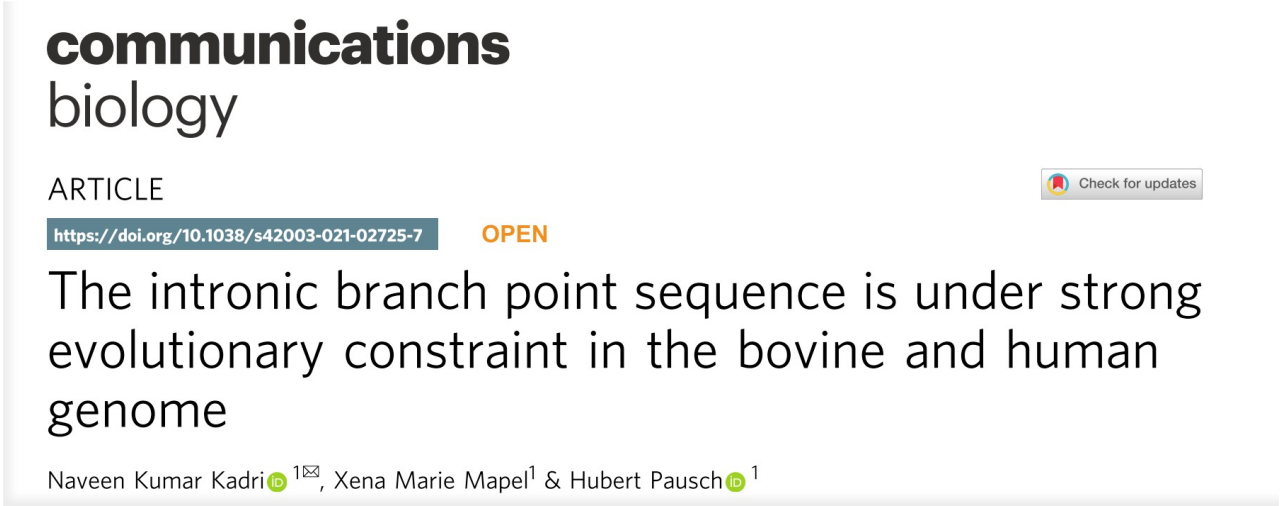
- m-RNA splicing Removal of (noncoding) introns and ligation of (coding) exons]
  - Spliceosome → evolutionarily conserved across eukaryotes
  - Requires motifs contained in intronic and exonic sequences (cis-acting)
    - Splice acceptor and donors
    - Polypyrimidine tract
    - Branch point sequence (BPS)



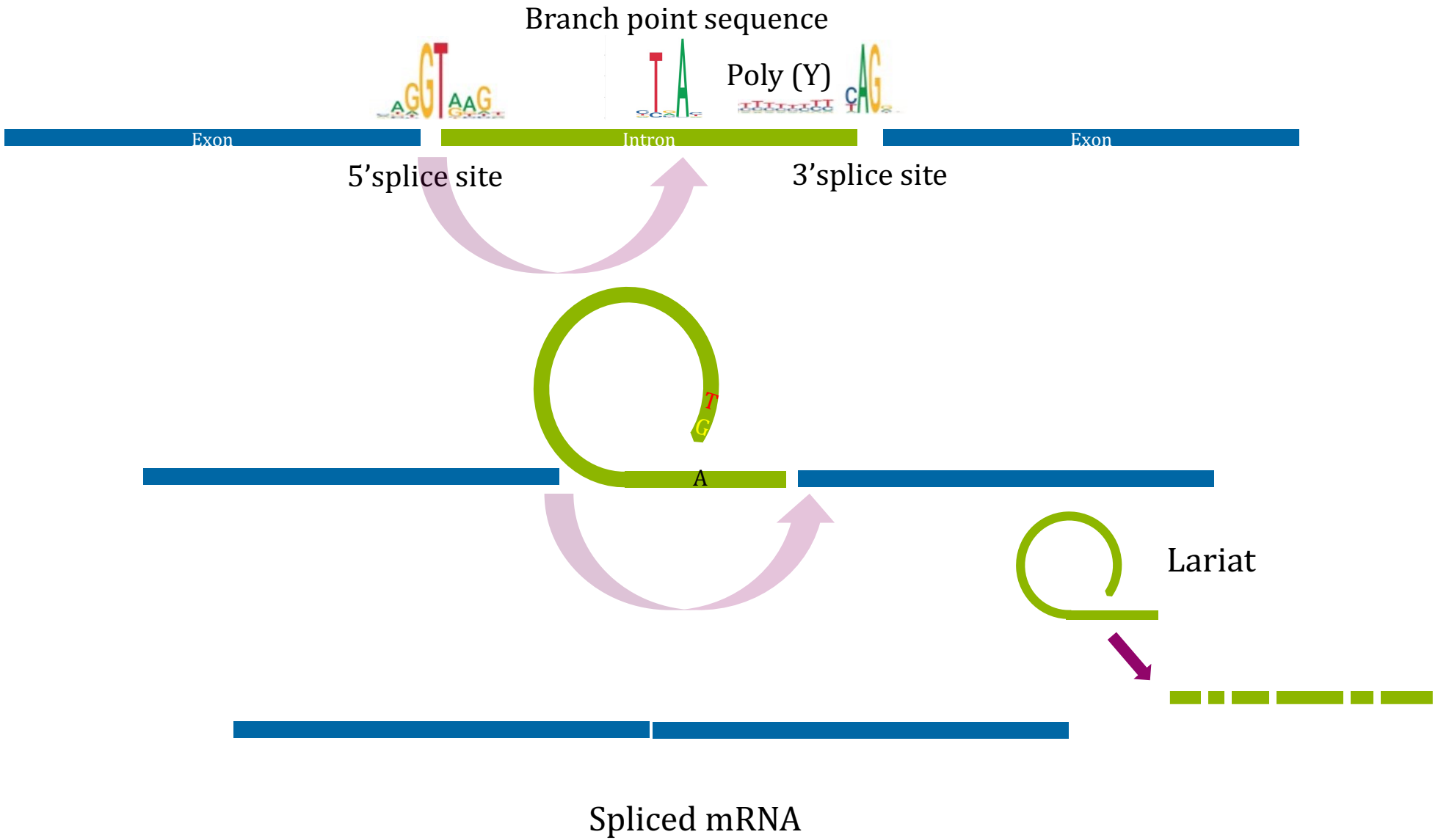
Introduction

- This study → focus on BPS
  - So far neglected feature
  - Associated with human diseases
  - Programs predicting functional consequences (e.g., VEP) are blind

  - Characterize bovine BPS
  - Strong evolutionary constraint
  - Splicing QTL overlapping bovine BPS
  - Confirmation in human BPS



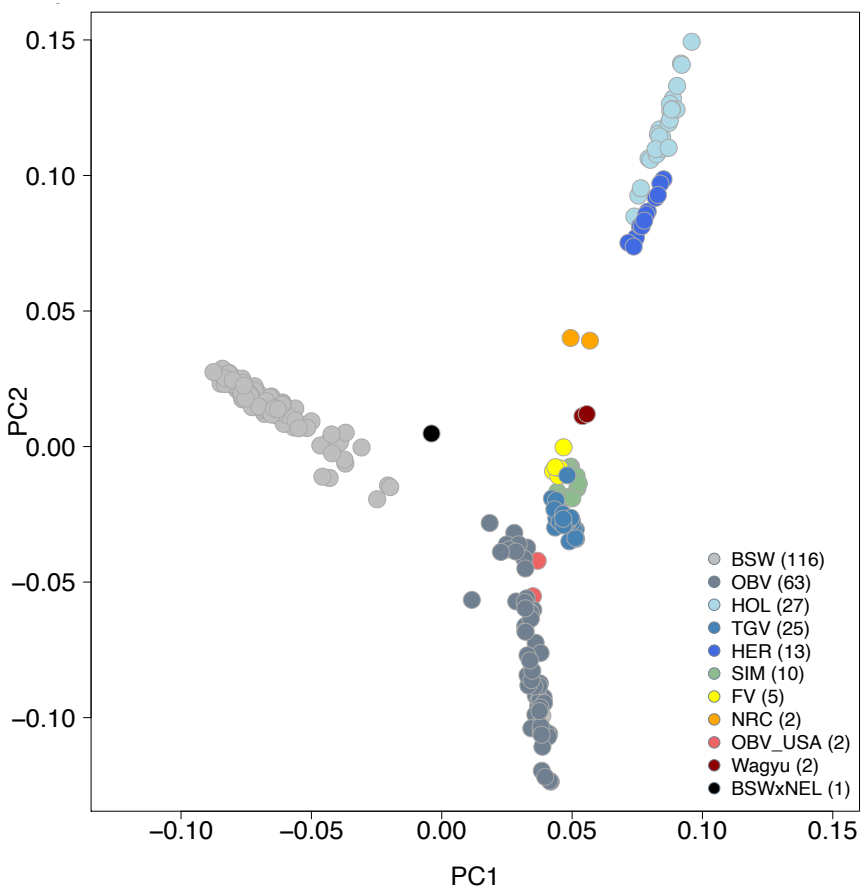
Introduction



Data

- 266 cattle from 11 breeds
  - Sequenced to an average depth of 16.28
- 29,227,950 (bi-allelic) and 190,200 (multi-allelic)
- One variant site every 85 bp

- Functional features
  - Ensemble (release 104) annotation of ARS-UCD1.2



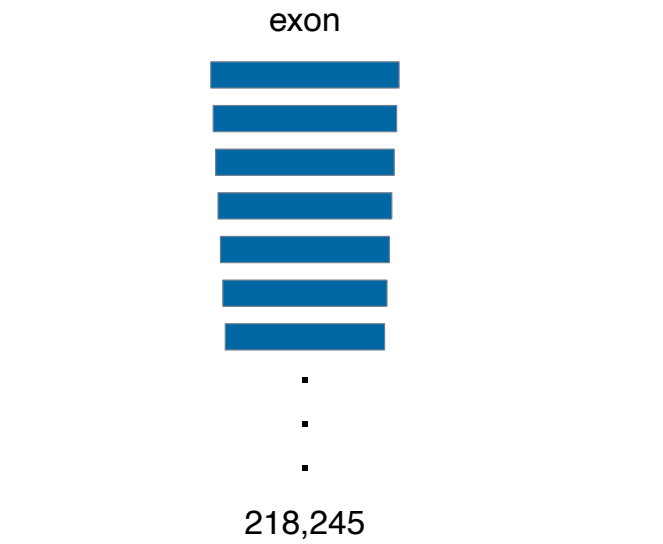
Constraint on annotated features

- 35,848 transcripts from 20,785 genes

- Variation expressed as variants /100 bp

Constraint on annotated features

- 35,848 transcripts from 20,785 genes

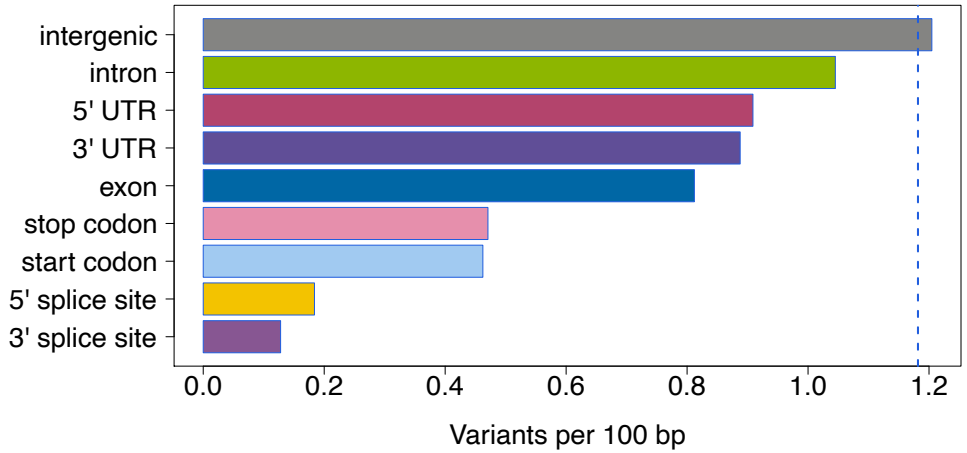


267,564 / 32,947,475
0.81 variants per 100 bp

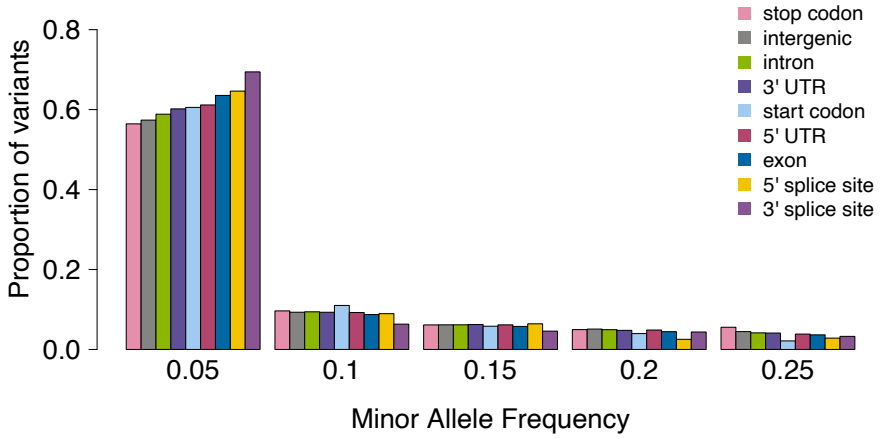
- Variation expressed as variants /100 bp

Constraint on annotated features

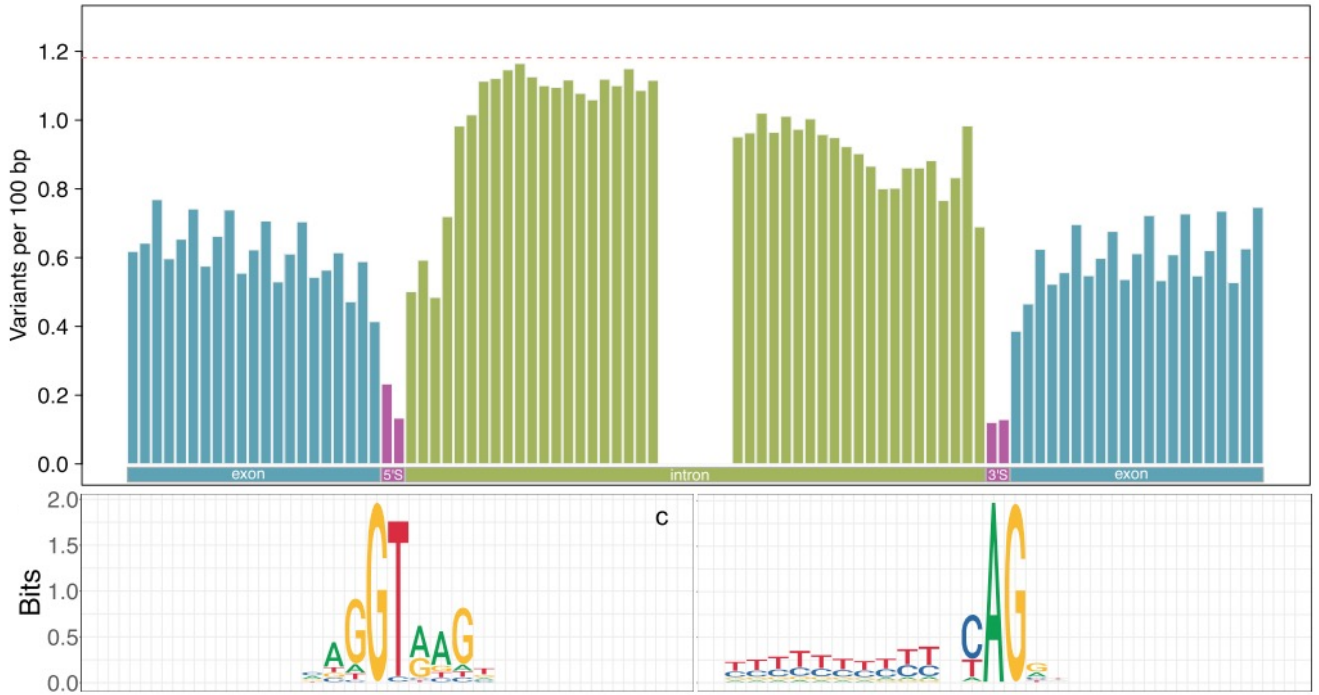
- Intergenic region more variable
- Genic features less variable
- Splice sites least variable



- Enriched for low-frequency/rare variants



Local variation around the Splice sites



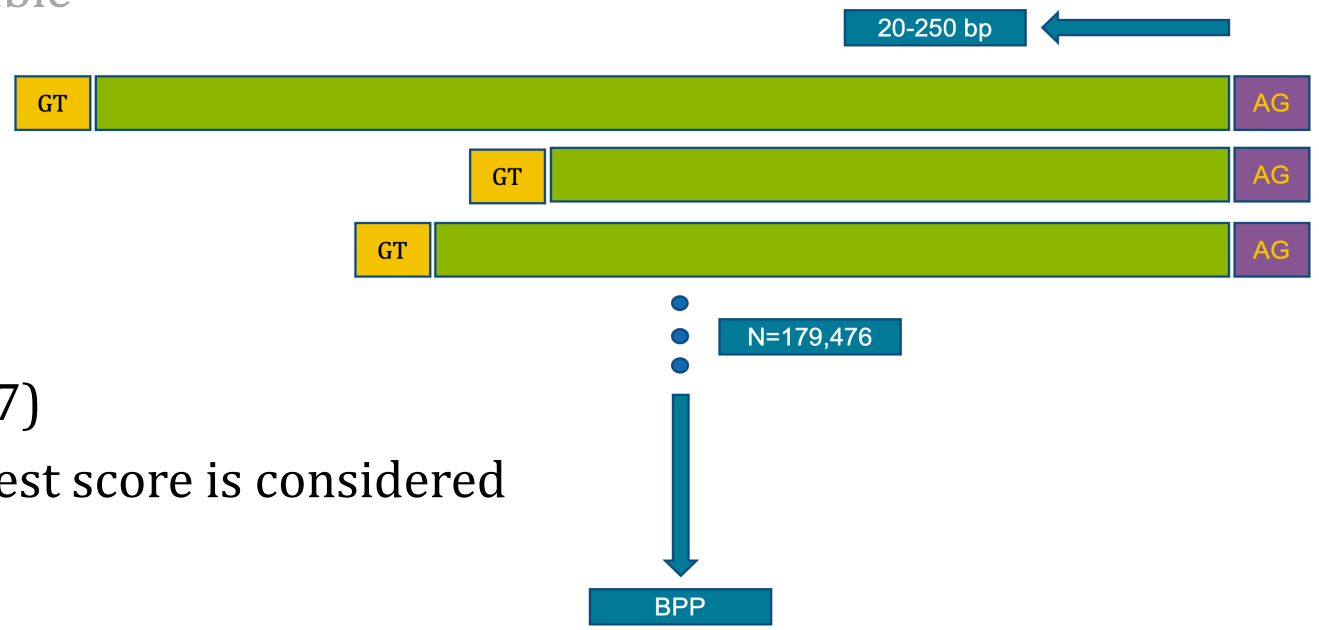
- Spice sites indeed less variable than surrounding intronic & exonic sequences
- Known patterns retrieved
  - Wobble at the third position of triplet codon

Prediction of Bovine branch point sequence

- Hitherto neglected → not available from GTF files
  - Computational prediction is possible

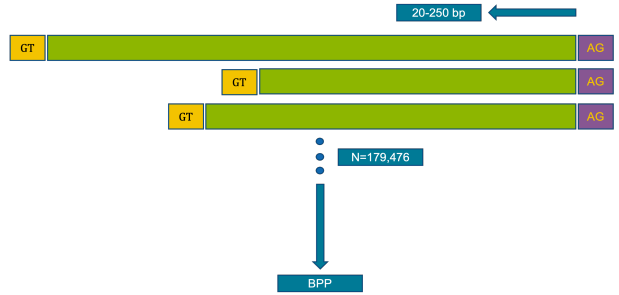
- 179,476 bovine introns >20 bp long
  - Up to 250 bp from the 3'SS

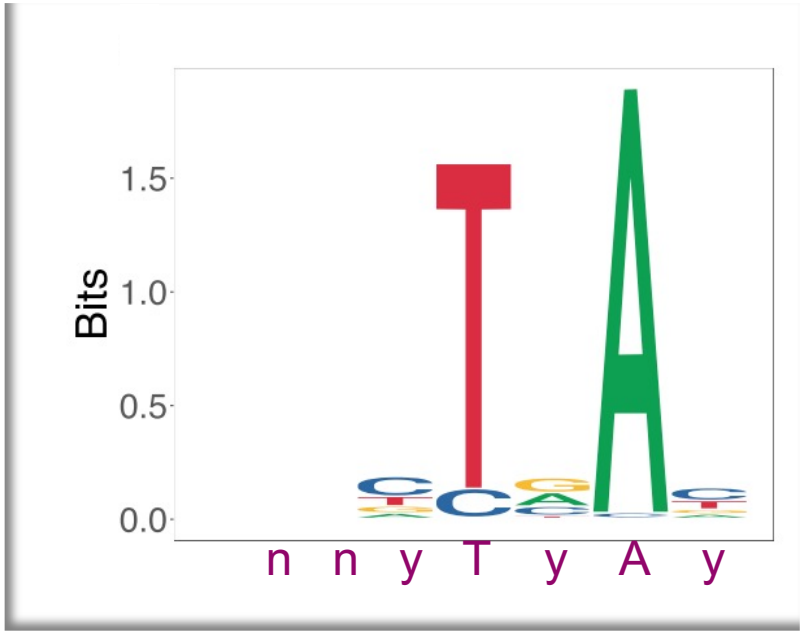
- Prediction using BPP (Zhang et al., 2017)
  - Only one BP/intron with the highest score is considered



Bovine branch point sequence

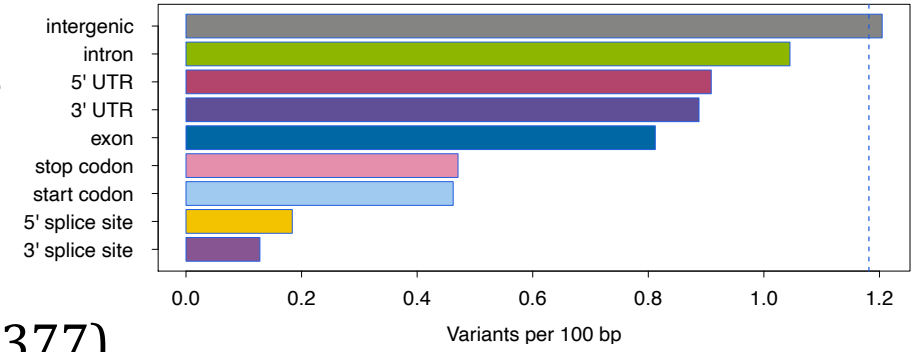
- Location 14-145 bp upstream of 3'ss (mean = 28.8 ± 11.2)
  - 82% between 18-37 bp
- Predicted branch point was predominantly Adenine (98.64%)
  - Cytosine (1.28%) or Thymine (0.08%)
- Degenerate heptamer → nnyTyAy
  - Highly conserved Thymine at position 4 (91.56%)
- >90% BPS were "TNA" at position 4 to 6 (canonical)
  - "CNA" → ~8%

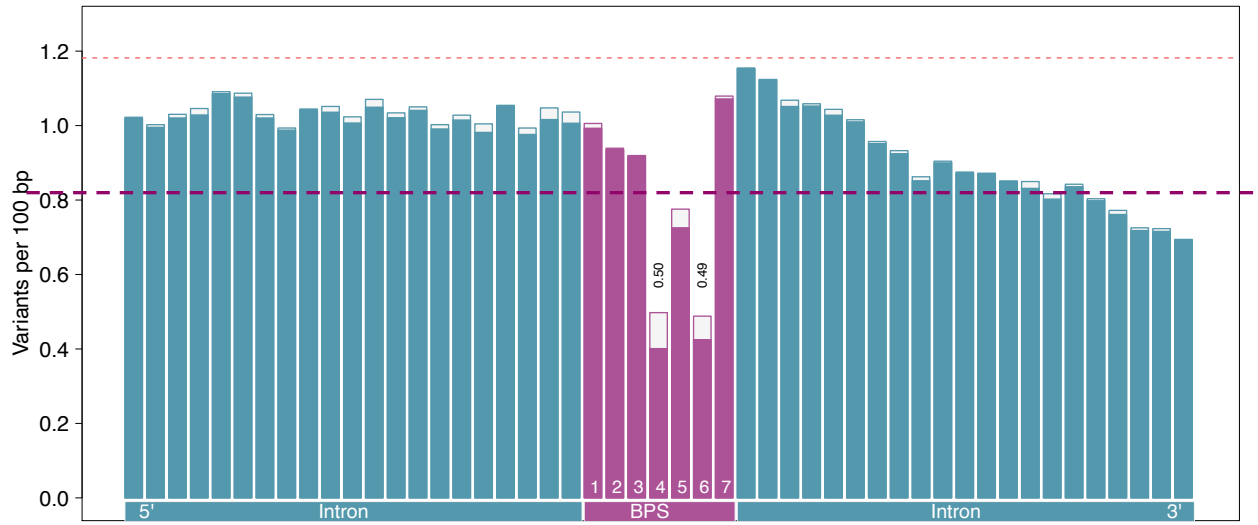




Constraint on bovine Branch point sequence

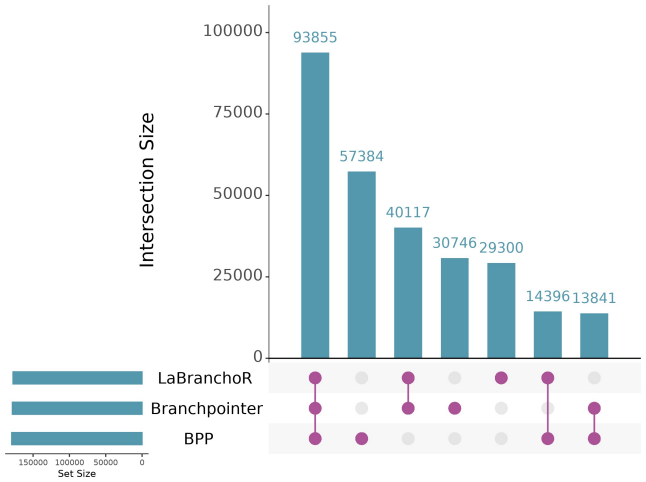
- Less variable than the surrounding intronic sequence
  - Highly conserved 4(T) and 6 (A)
- Lower than coding sequence
- Stronger constraint in canonical ("TNA") BPS (n=162,377)

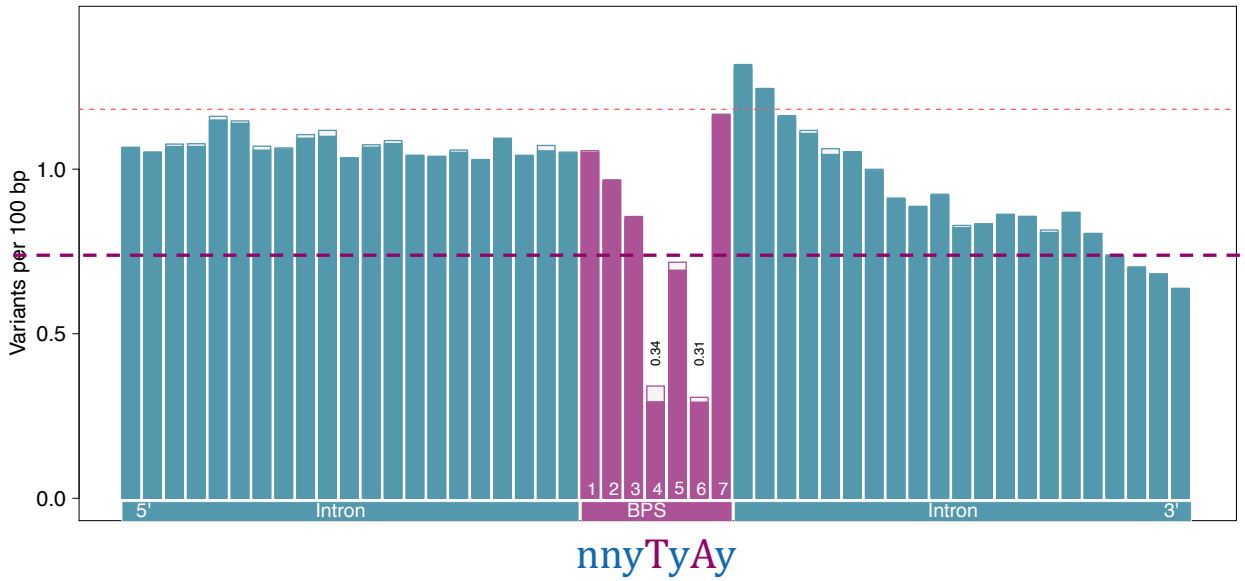




Prediction tools

- Prediction from two other tools → branchpointer[1] and LaBranchOR[2]
- Matching ~53% (n=93,855)
- Even higher constraint on these 93,855 BPS





[1]Paggi et al., RNA 24, 1647–1658 (2018). [2]Signal et al., Bioinformatics 34, 920–927 (2018).

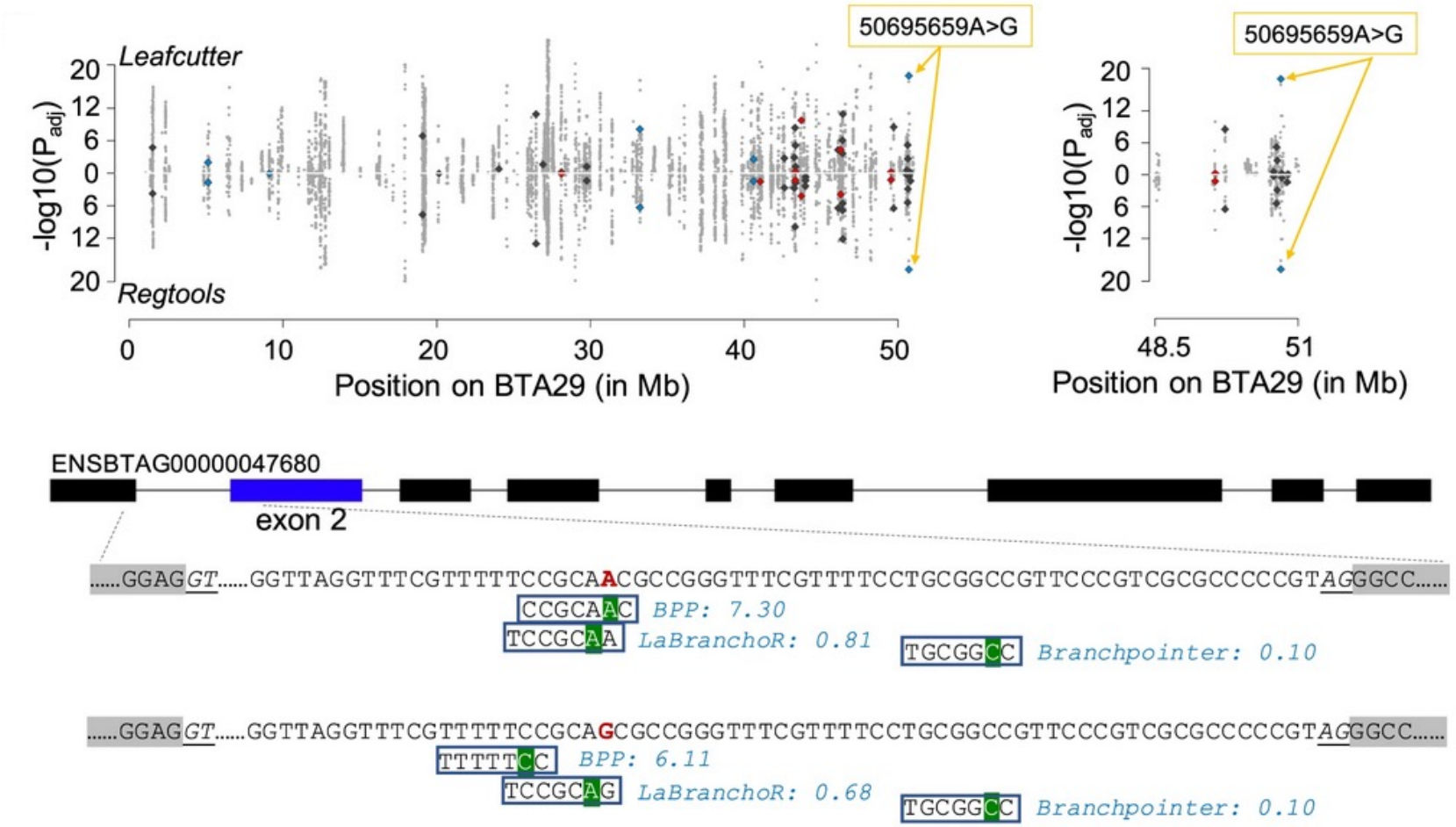
Splicing QTL

- sQTL cohort
  - Testis tissue sampled from 76 mature bulls
  - DNA sequencing to an average of 12.6-fold depth
  - Deep total paired-end RNA sequencing (283.6 million reads per sample)

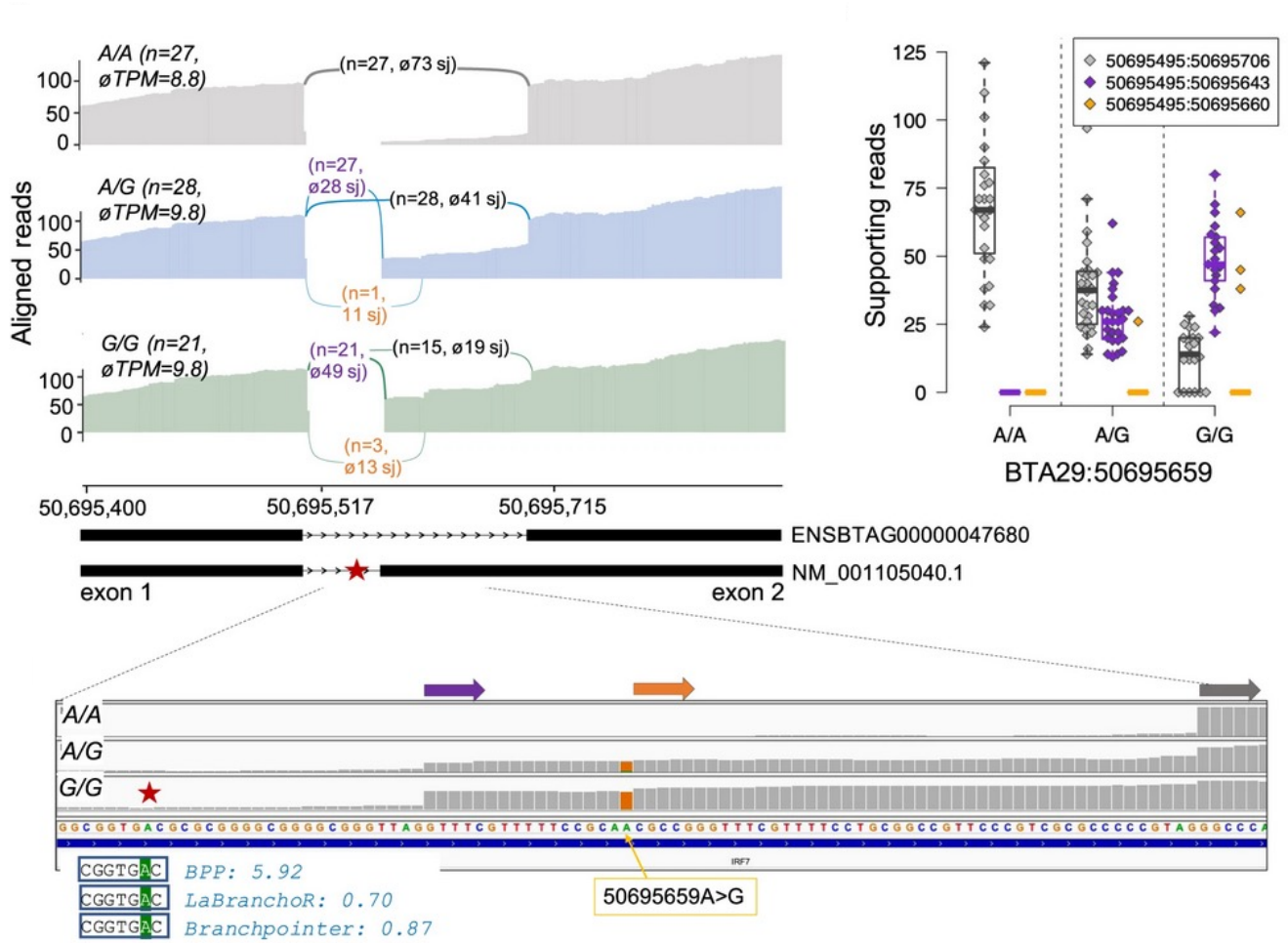
- Splicing events characterized using *Leafcutter*[1] and *Regtools*[2]
- cis-sQTL detection using *FastQTL*[3]

[1]Li et al., ***Nat. Genet***. 50, 151–158 (2018), [2]Cotto et al., ***bioRxiv*** https://doi.org/10.1101/436634 (2021), [3]Ongen et al., ***Bioinformatics*** 32, 1479–1485 (2016)

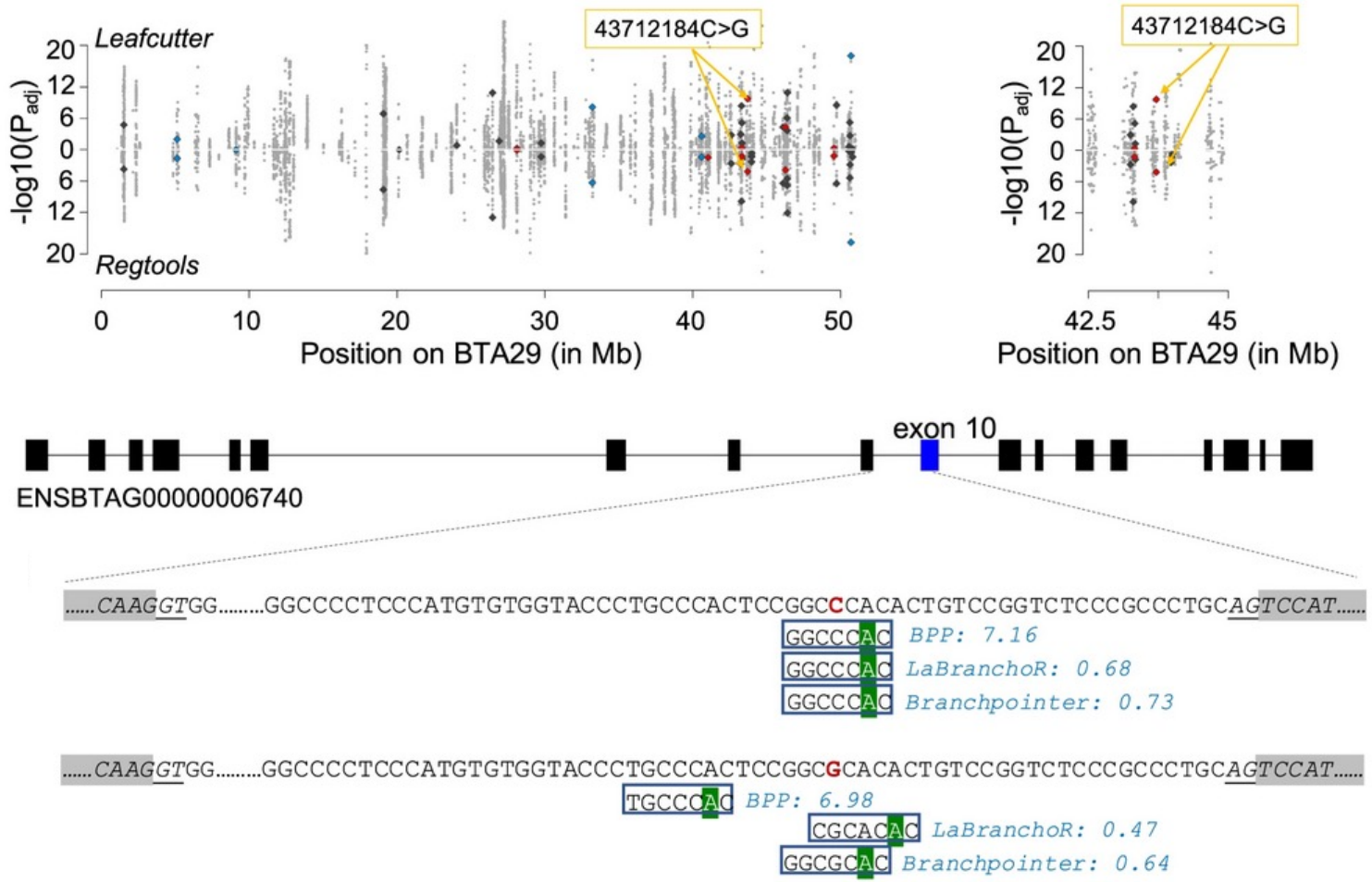
A branch point mutation leads to utilization of alternative 3’ splice acceptor sites in *IRF7*



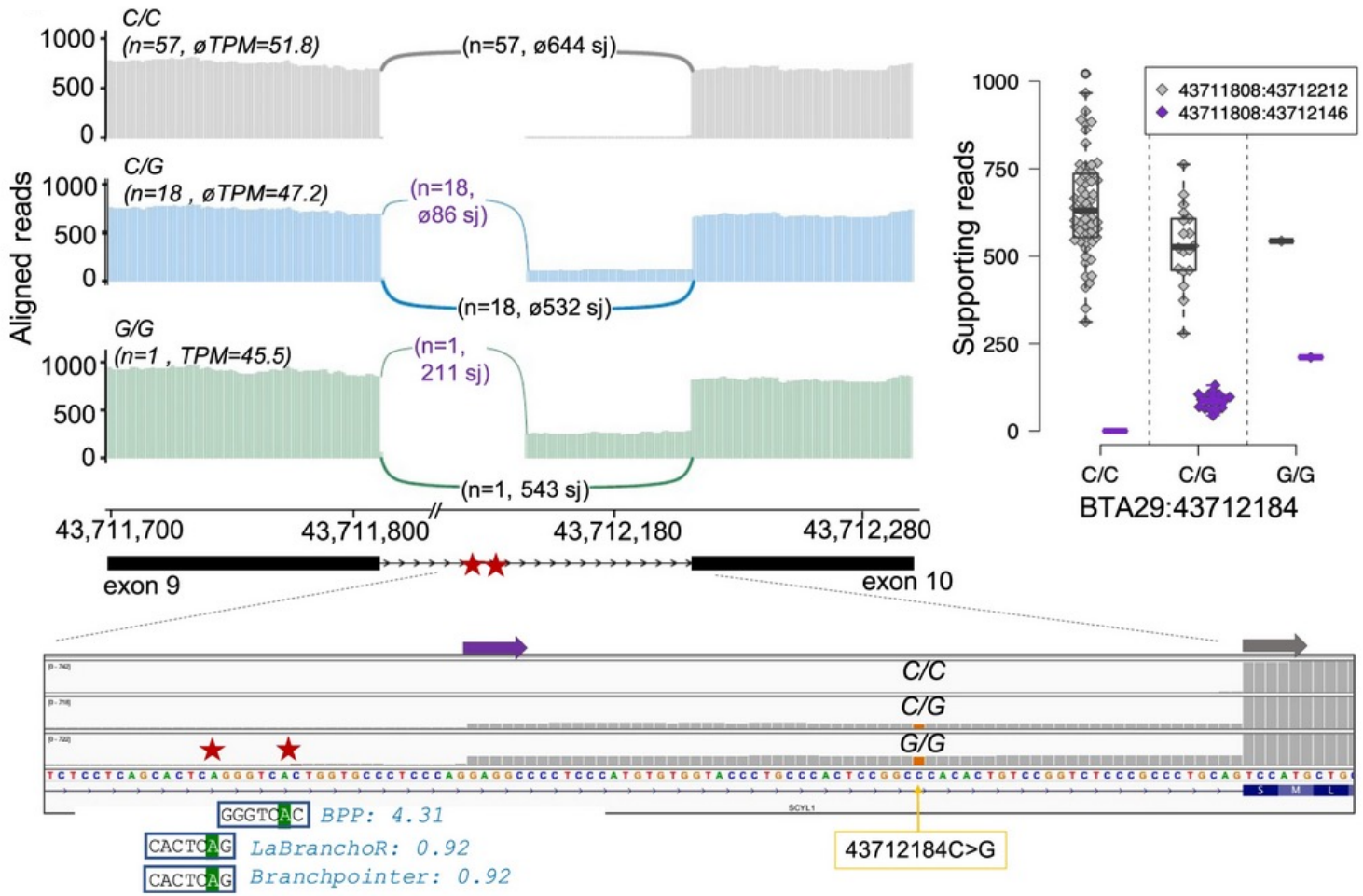
A branch point mutation leads to utilization of alternative 3’ splice acceptor sites in *IRF7*



A branch point sequence mutation leads to alternative 3’ splicing of *SCYL1* exon 10

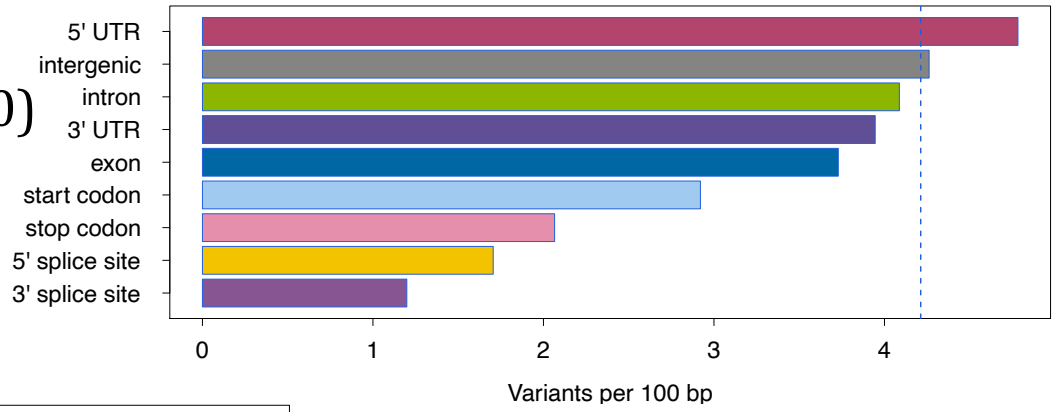


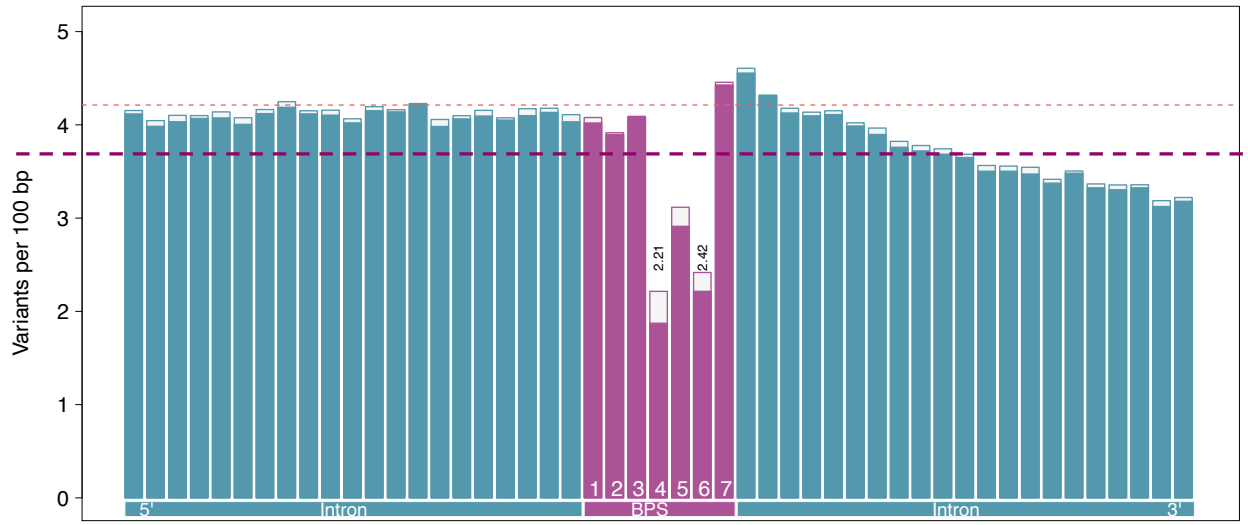
A branch point sequence mutation leads to alternative 3’ splicing of *SCYL1* exon 10



Confirmation in Human functional features

- Variant catalogue of 115,640,370
  - 3942 human genomes (HGDP + 1KG; gnomAD v3.0)
  - 4.21 variants/ 100 bp
- BPS predicted for 201,832 introns → nnyTyAy
- Higher constraint on the Thymine at position 4





Genome-wide Exonic

nnyTyAy

Conclusions

- First description of Bovine BPS
  - Features very similar to other species

- Strong constraint → stronger than coding sequence
  - Possible false positives, true constraint might be even stronger

- Variants associated with alternative mRNA splicing

- Ignored as candidate casual variants
  - We provide annotations
  - Genotyping BPS variants in large cohorts with customized SNP chip

Acknowledgements

Thank you …

BovReg
Understanding cattle genomes

European Commission
Horizon 2020
European Union funding
for Research & Innovation

AgroVet Strickhof
Eine Kooperation in Bildung und Forschung:
strickhof ETH zürich University of Zurich

ETH zürich
University of Zurich
Functional Genomics Center Zurich

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Federal Office for Agriculture FOAG

swissgenetics