Computational Systems Biology
Deep Learning in the Life Sciences

6.802 6.874 20.390 20.490 HST.506

Boryana Doyle

February 28, 2017

The human splicing code reveals new insights into the genetic determinants of disease

H.Y. Xiong, B. Alipanahi, L.J. Lee, H. Bretschneider, D. Merico, R.K.C. Yuen, Y. Hua, S. Gueroussov, H.S. Najafabadi, T.R. Hughes, Q. Morris, Y. Barash, A.R. Krainer, N. Jojic, S.W. Scherer, B.J. Blencowe, B.J. Frey

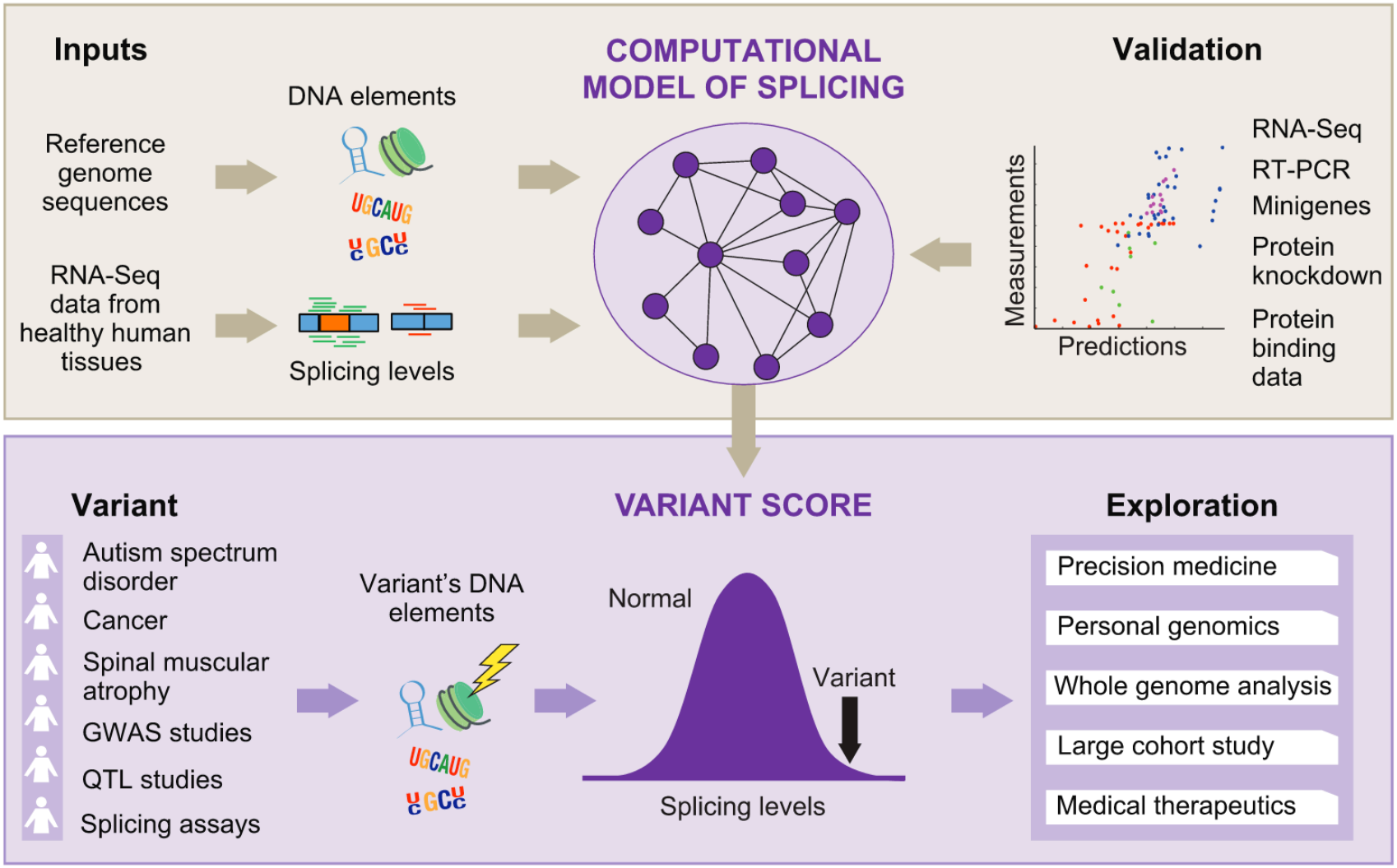
Massachusetts Institute of Technology

http://mit6874.github.io

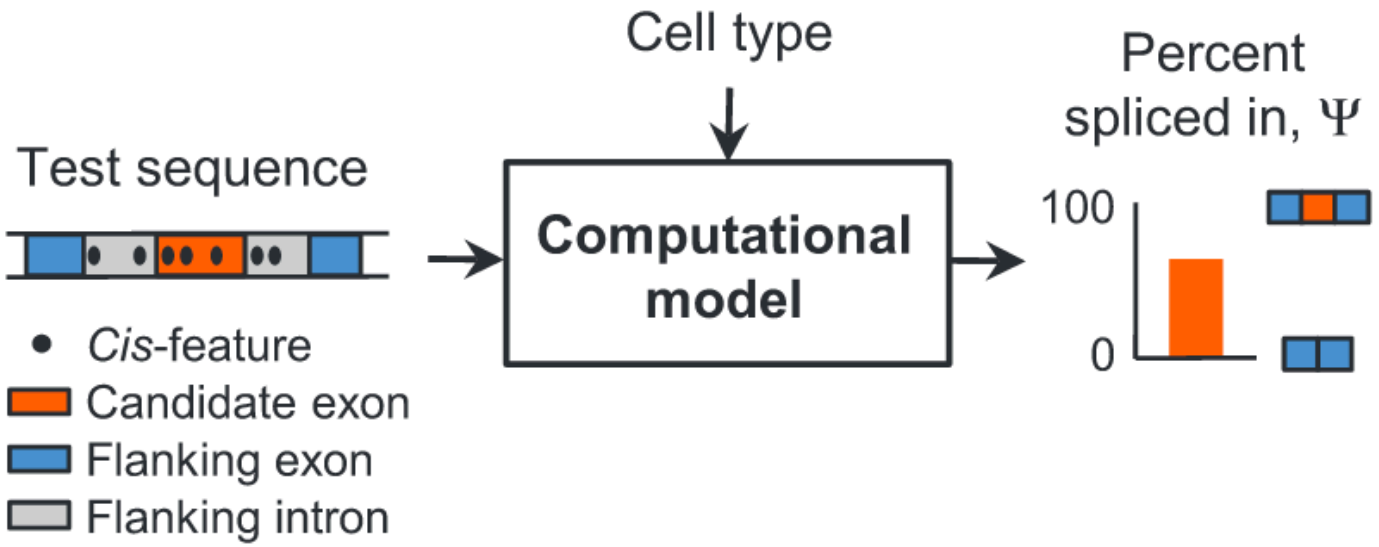
Overview

- Key Claim
  - ***Presents a deep learning algorithm to predict RNA splicing locations from genomic DNA sequence and RNA-seq from 16 healthy human tissues***
- Importance
  - ***This method is important because it does not rely upon disease annotations and has predictive power***
- Issues
  - ***The paper is lopsided: it describes the drawbacks in one sentence buried in the discussion***

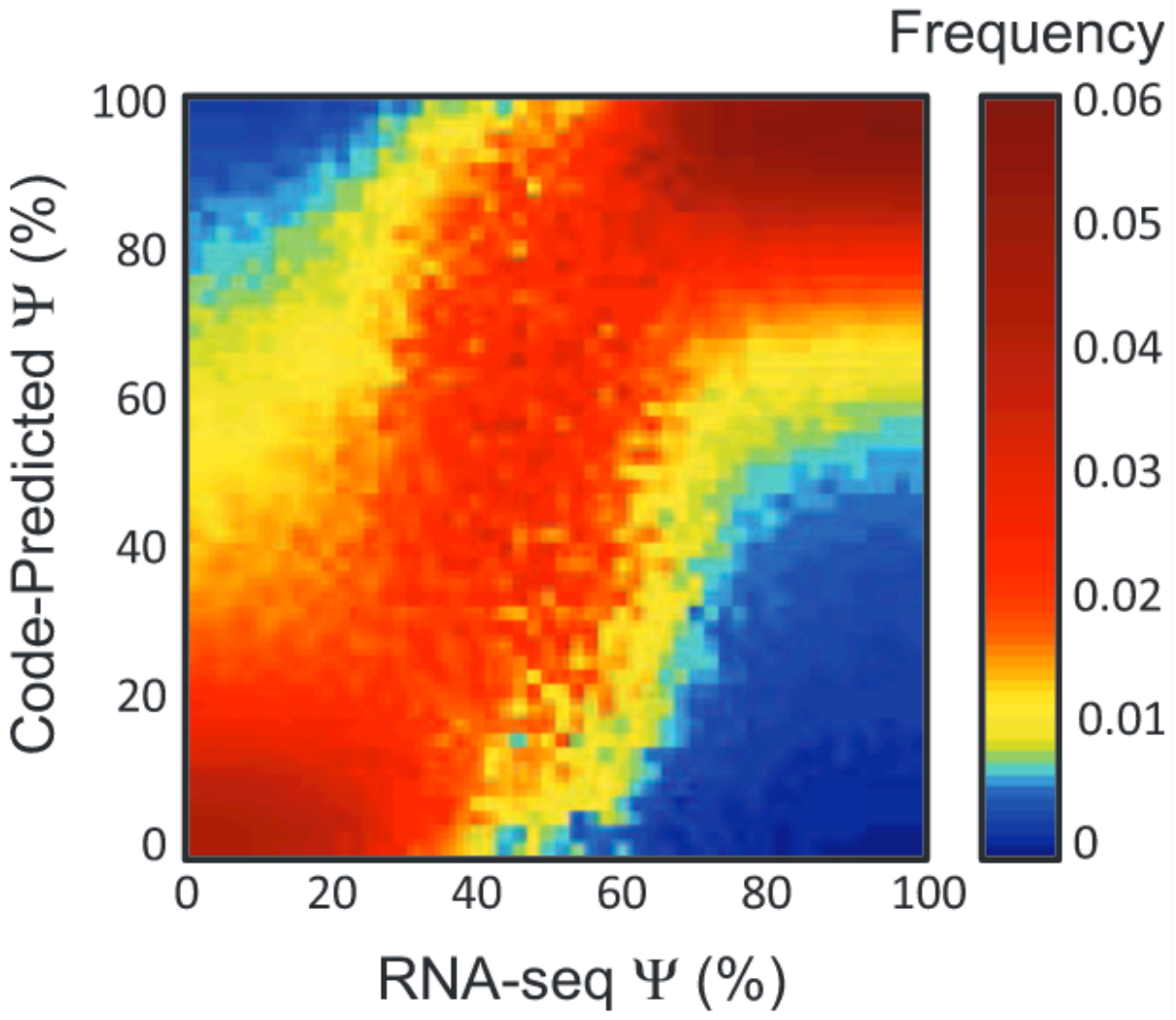
Data Sources, Method Overview



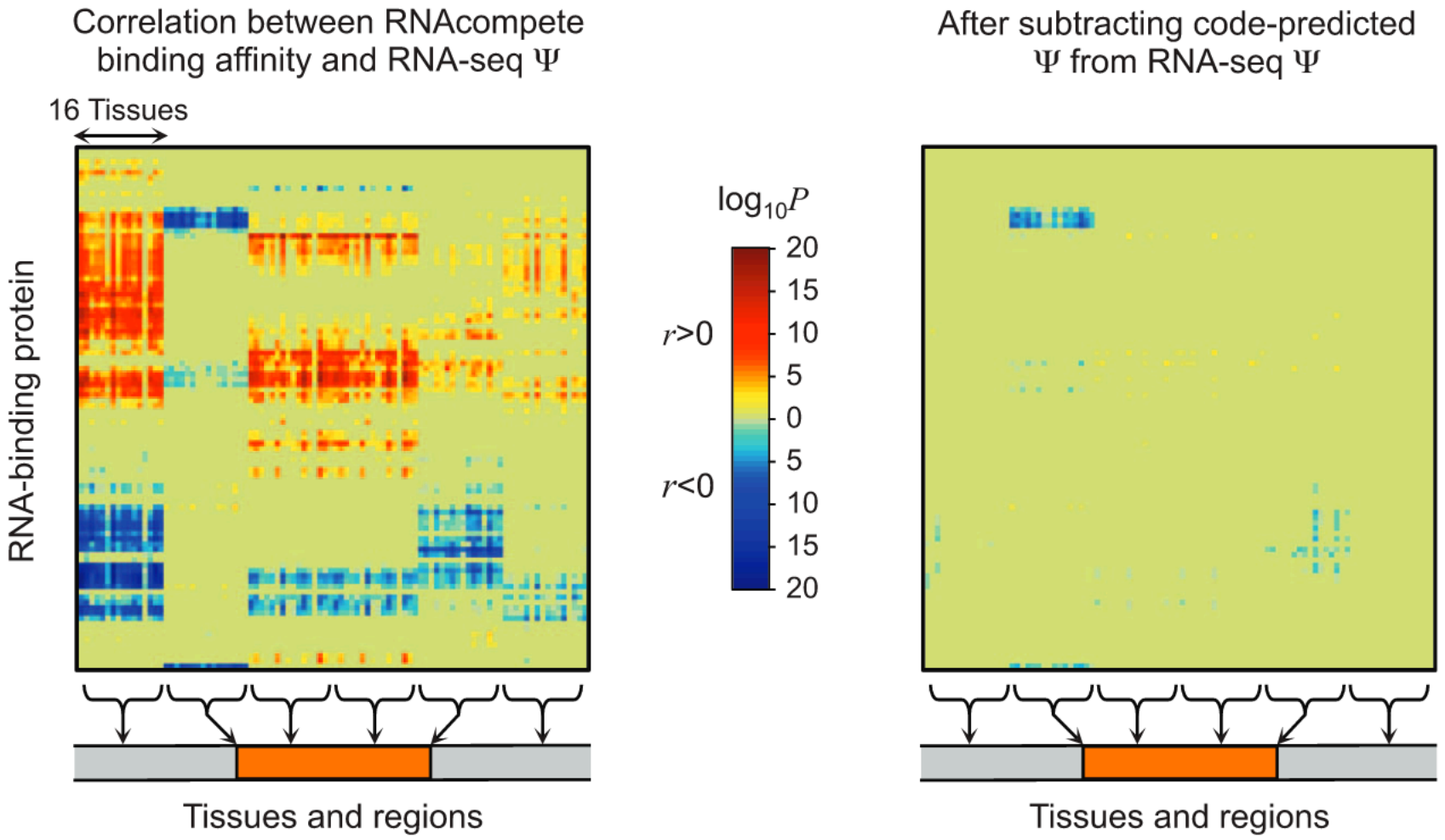
Method


Cell type
Test sequence
Computational model
Percent spliced in, Ψ
100
0
Cis-feature
Candidate exon
Flanking exon
Flanking intron

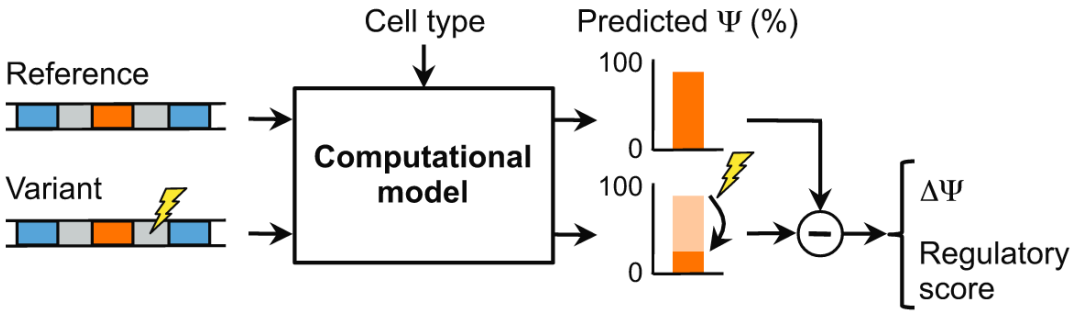
Validations

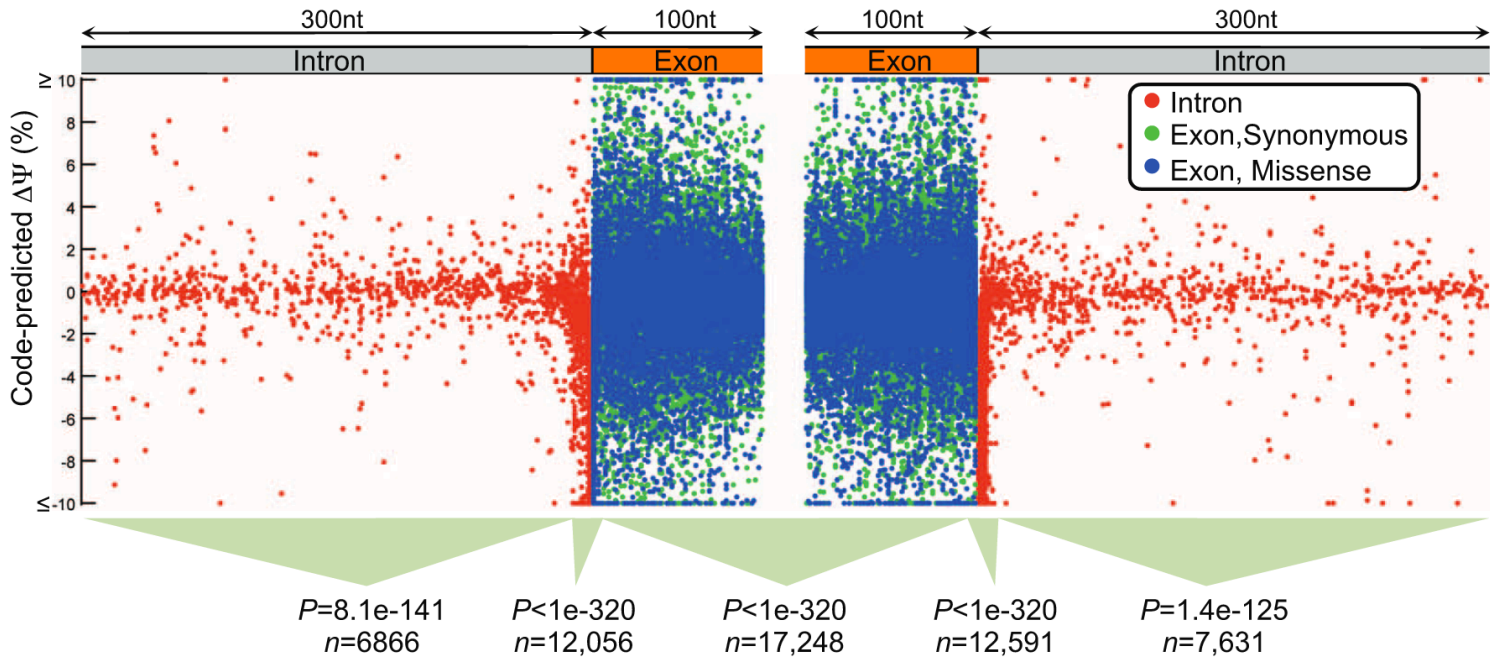


Validations

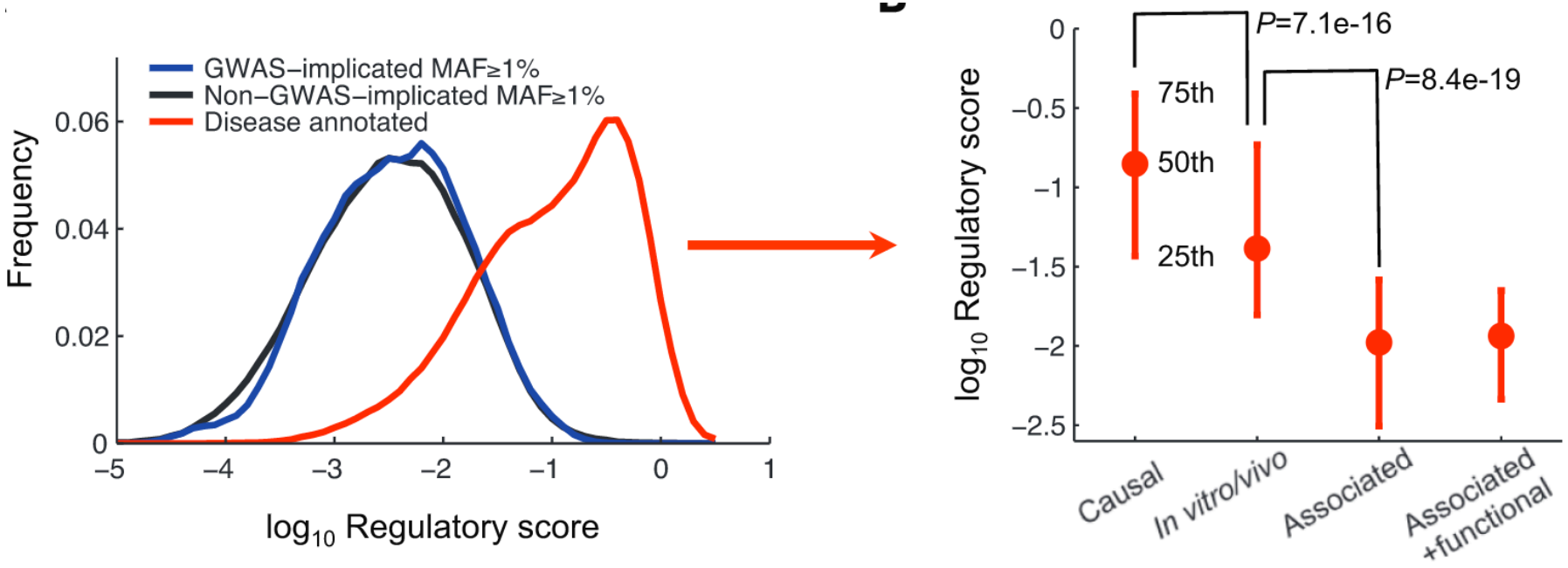


Results: SNV in exons, introns

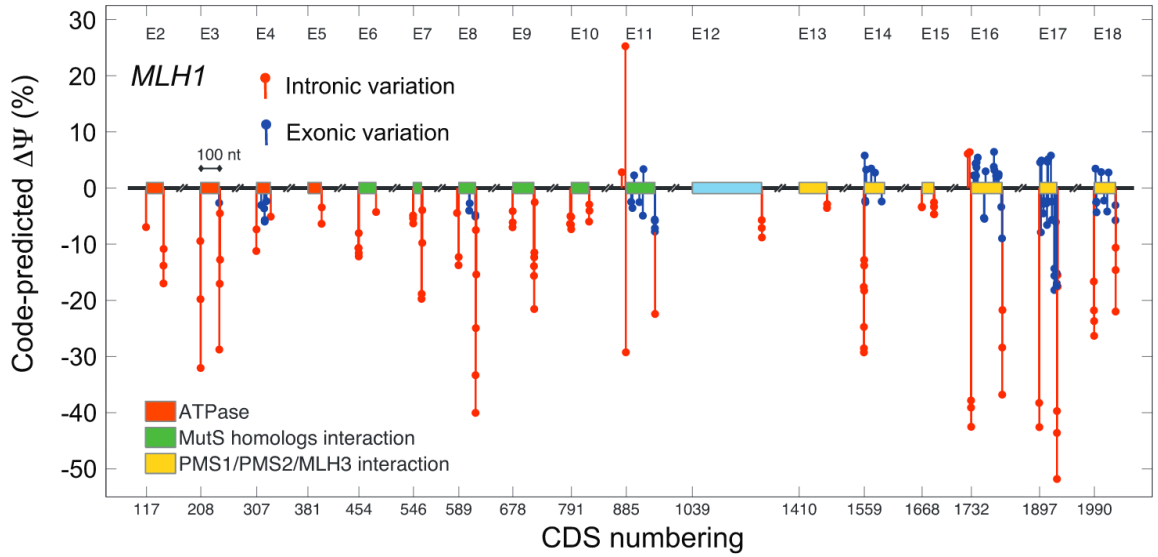


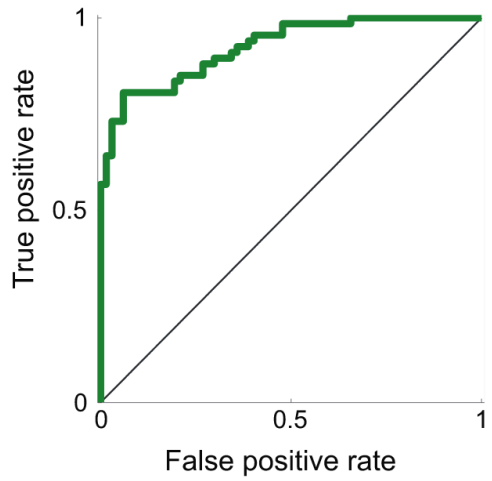


Results: Comparison with GWAS

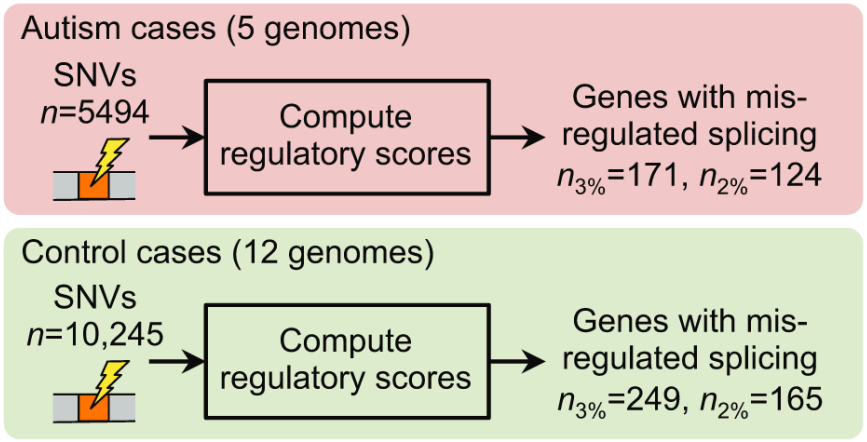


Results: Nonpolyposis colorectal cancer



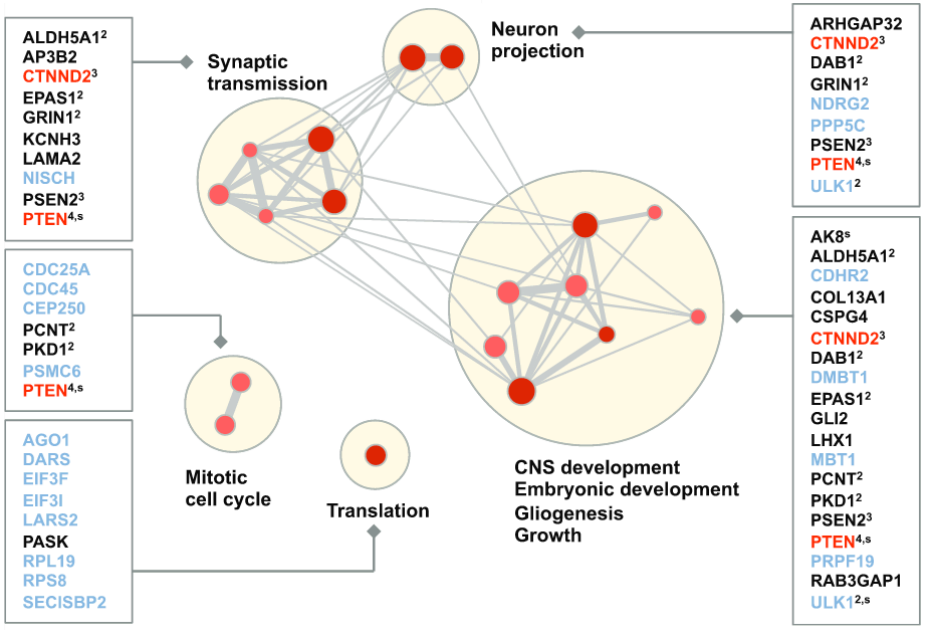


Results: Autism Spectrum Disorder



	Brain genes	Other genes	Fraction
Autism	46	125	27%
Controls	33	216	13%

P=3.8e-4 (Fisher's exact test)



Key Claims

- ***Method provides useful splicing scores trained on genomic data. Experimental validation.***
- ***It compares favorably to previous disease-based methods***
- ***SVNs score better than SNPs***
- ***SNVs score better than GWAS-implicated SNPs***
- ***It has predictive power for spinal muscular atrophy, nonpolyposis colorectal cancer, and Autism Spectrum Disorder.***

Analysis

- ***The paper uses statistical tests, multiple hypothesis correction, and error estimates***
- ***The results should be reproducible, using their parameters for training the network***
- ***The paper fails to provide a comprehensive set of biological drawbacks for their method***
- ***The method compares favorably to the prior method of GWAS, which uses disease data***

Impact

- ***This paper proves that machine learning starting from the genomic sequence (without disease and non-disease populations) can predict splicing by combining multiple features/context***
- ***It can be used for more and more diseases to find and classify potential mutants or provide a score for a specific mutation***

Summary and Conclusion

- Key Claim
  - ***Presents a deep learning algorithm to predict RNA splicing locations from genomic DNA sequence and RNA-seq from 16 healthy human tissues***
- Importance
  - ***This method is important because it does not rely upon disease annotations and has predictive power***
- Issues
  - ***The paper is lopsided: it describes the drawbacks in one sentence buried in the discussion***

FIN - Thank You