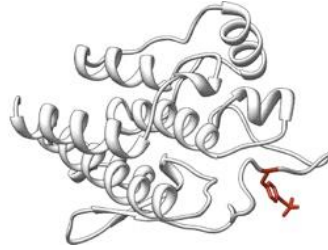


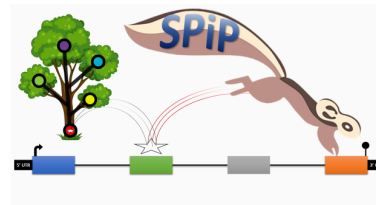
CAGI6 Challenge



SPIP

Splicing Prediction Pipeline

A machine learning tool for massive detection of exonic and intronic variant effect on mRNA splicing.



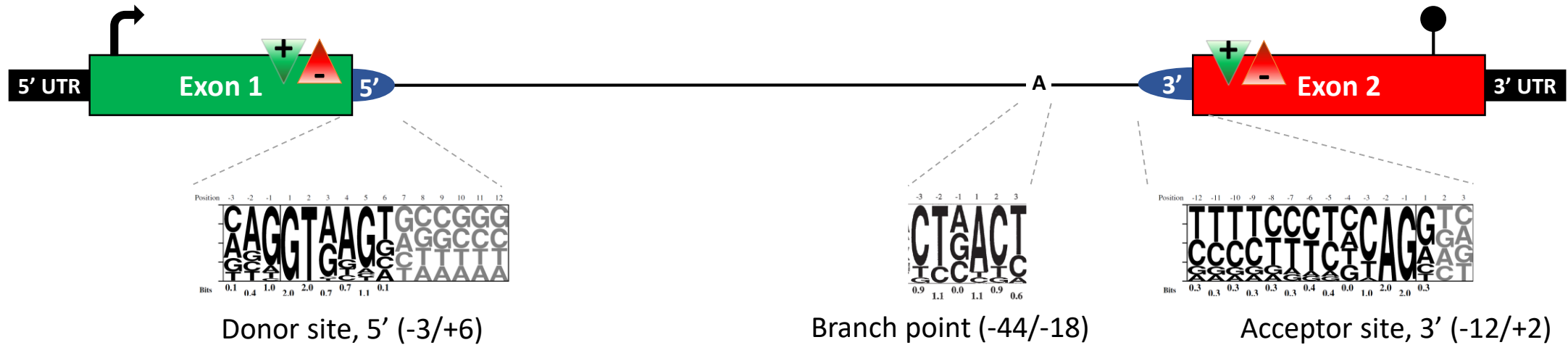
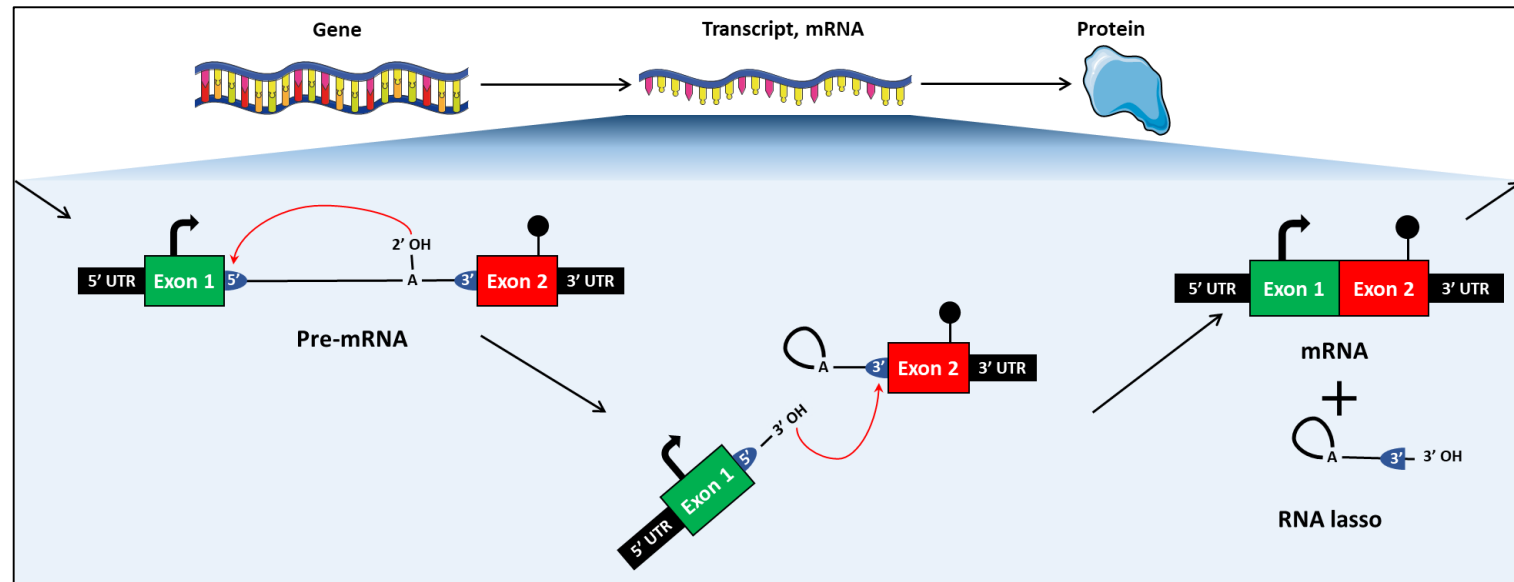
Raphaël Leman PhD

r.leman@baclesse.unicancer.fr

Sophie Krieger and Claude Houdayer

Inserm U1245, CLCC Fr. Baclesse, CAEN

Splicing motives



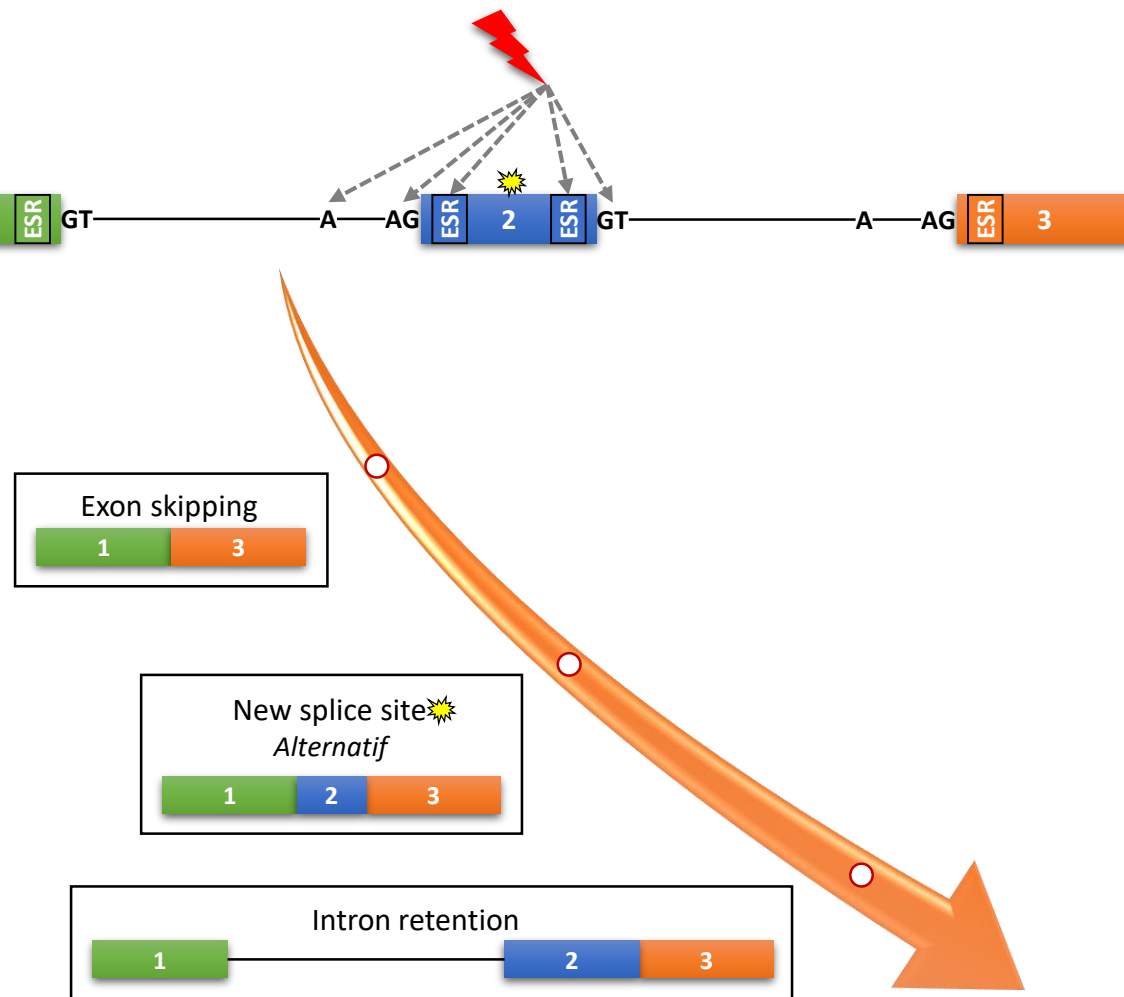
Exonic Splicing Regulators (ESRs): Enhancers/Silencers

May 2022

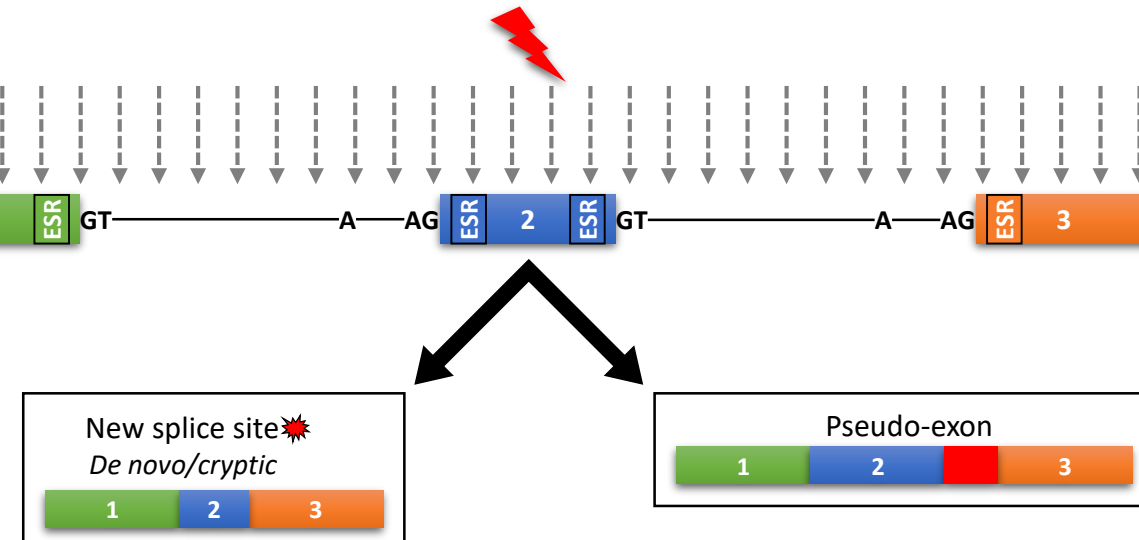
Raphaël Leman, CAGI6 challenge

Splicing alteration mechanisms

Disturbing natural splice site motives

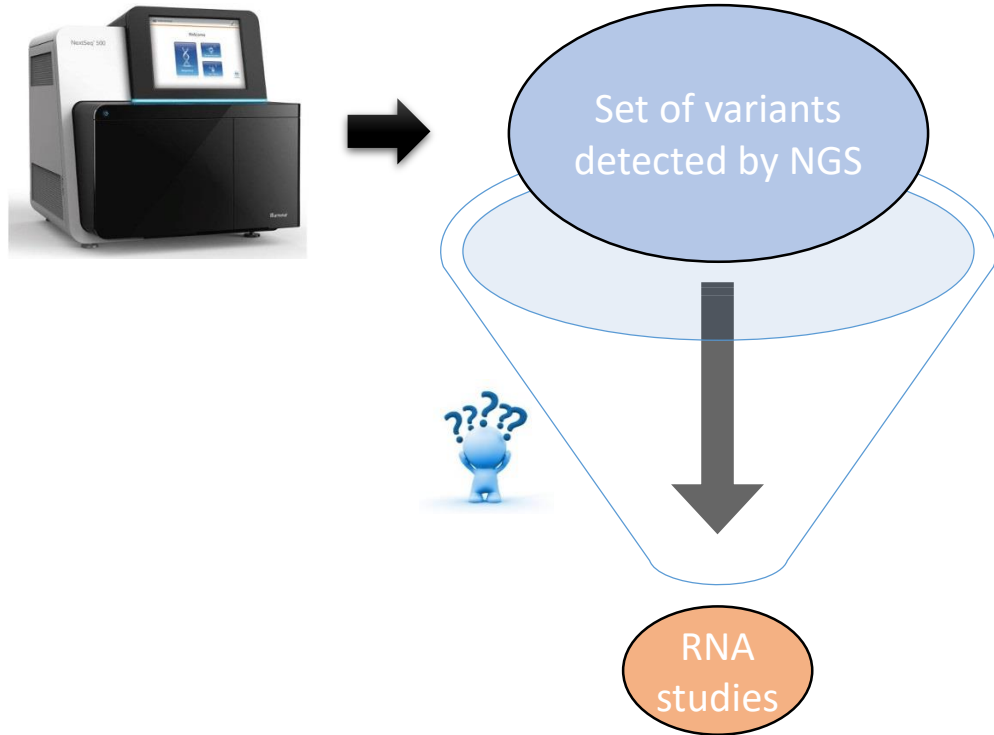


Creating new splice site motives



3,8 % of exonic and intronic genetic variations near exons **impact splicing** without hereditary context

SPiP, goals



SPiP

Combination of optimal tools for each splicing motif



- ☐ Predict splicing defect whatever the variant position
- ☐ Give an explanation of why a variant is spliceogenic
- ☐ Suitable for all genes
- ☐ Easy-to-use
- ☐ Evaluate the performance of SPiP against the state of the art

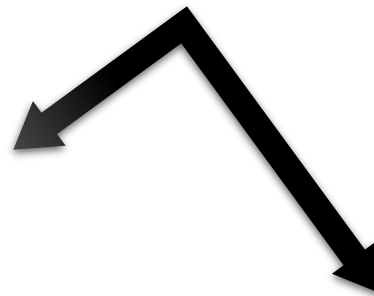
Variants source

- **4 616 variants** with RNA *in vitro* data
 - Study for research and diagnostic purpose in genes involved in genetic disorders



Publish data, n = 4 114 variants

Review of > 350 published articles from 1983 to 2019



Unpublished data, n = 502 variants

Splicing network of Unicancer genetic group, n = 166

NF1 Diag. Paris Lab. (Cochin) , France, n = 110

Inserm U1078, Brest, France, n = 63

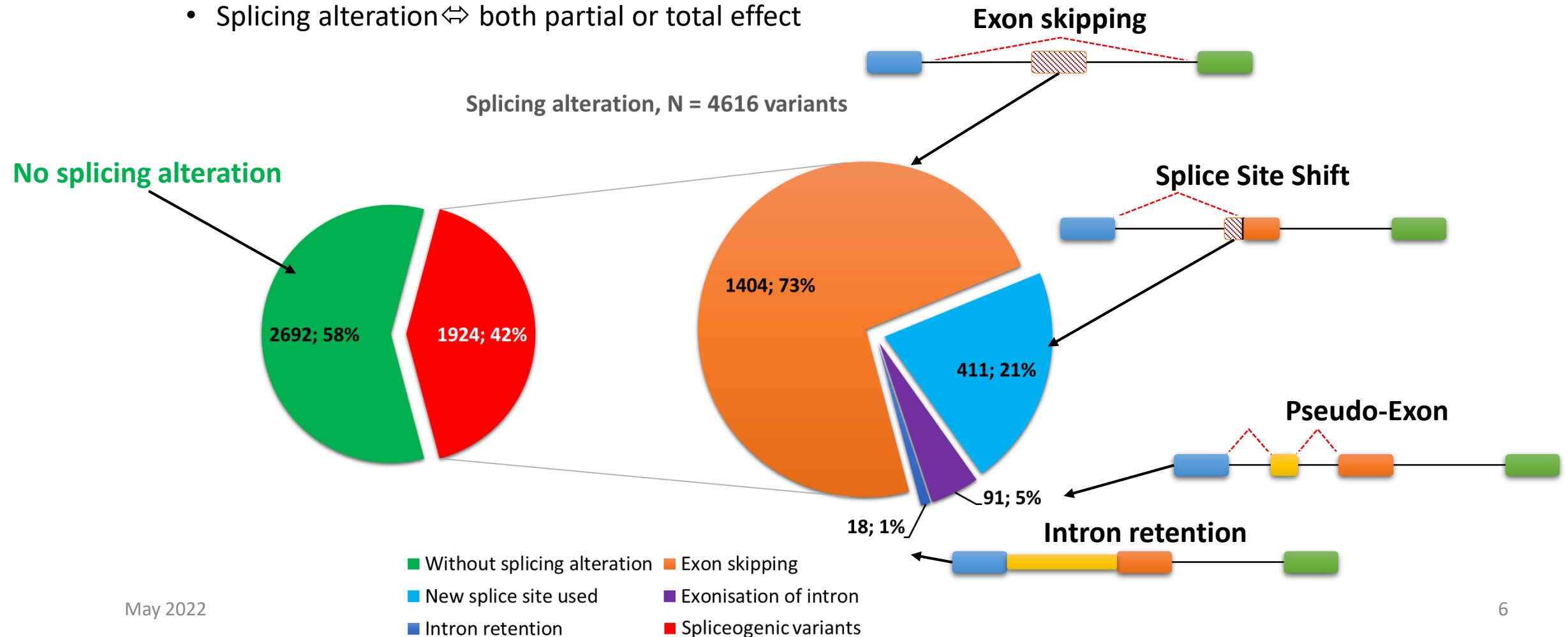
Inserm U1245, Rouen, France, n = 22

Lab. Genet., AP-HP, Paris, France, n = 17

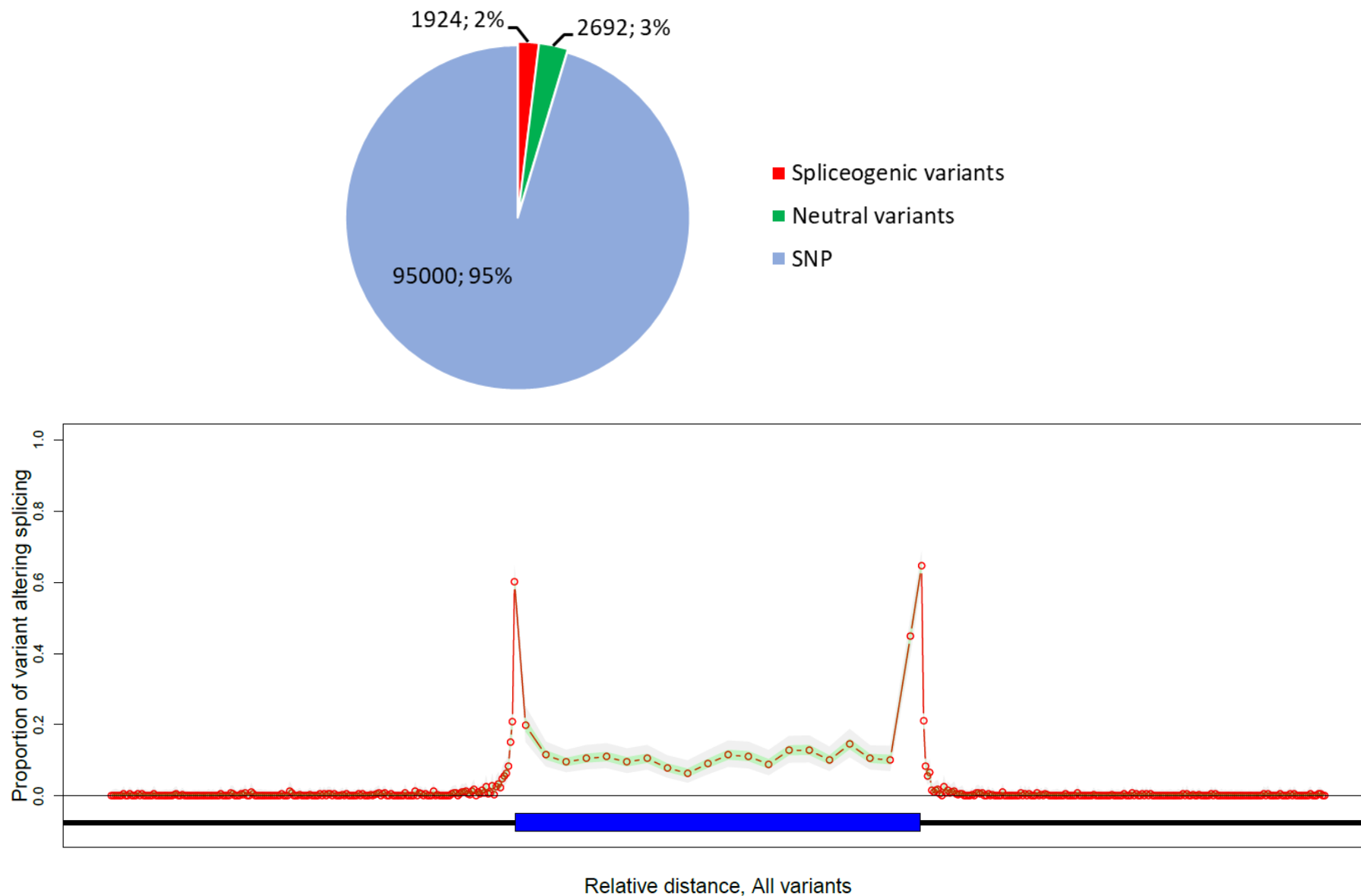
ENIGMA coll: Cent. for Geno. Med., Copenhagen, Denmark, n = 3

Distribution of splicing impact

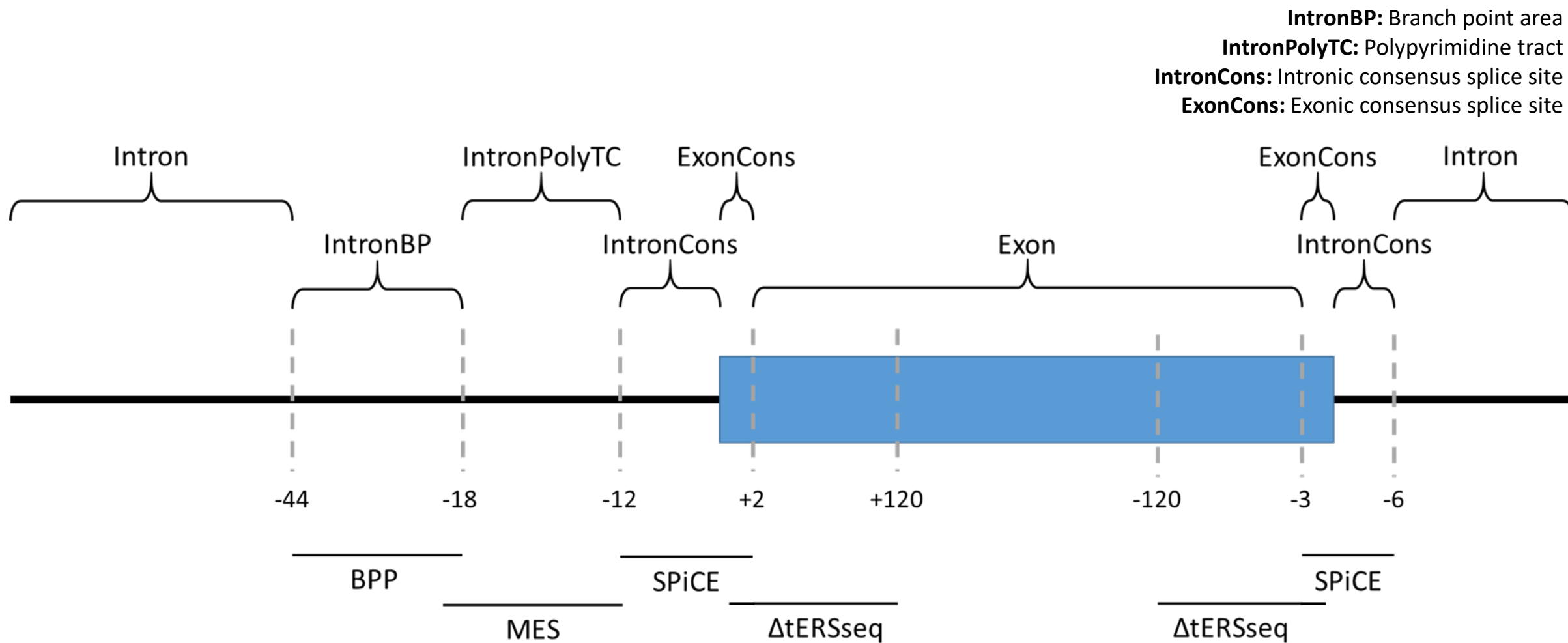
- Variant description:
 - 227 genes, mainly *BRCA1*, *BRCA2* et *NF1*
 - Splicing alteration $\Leftrightarrow$ both partial or total effect



Distribution of splicing impact with polymorphisms



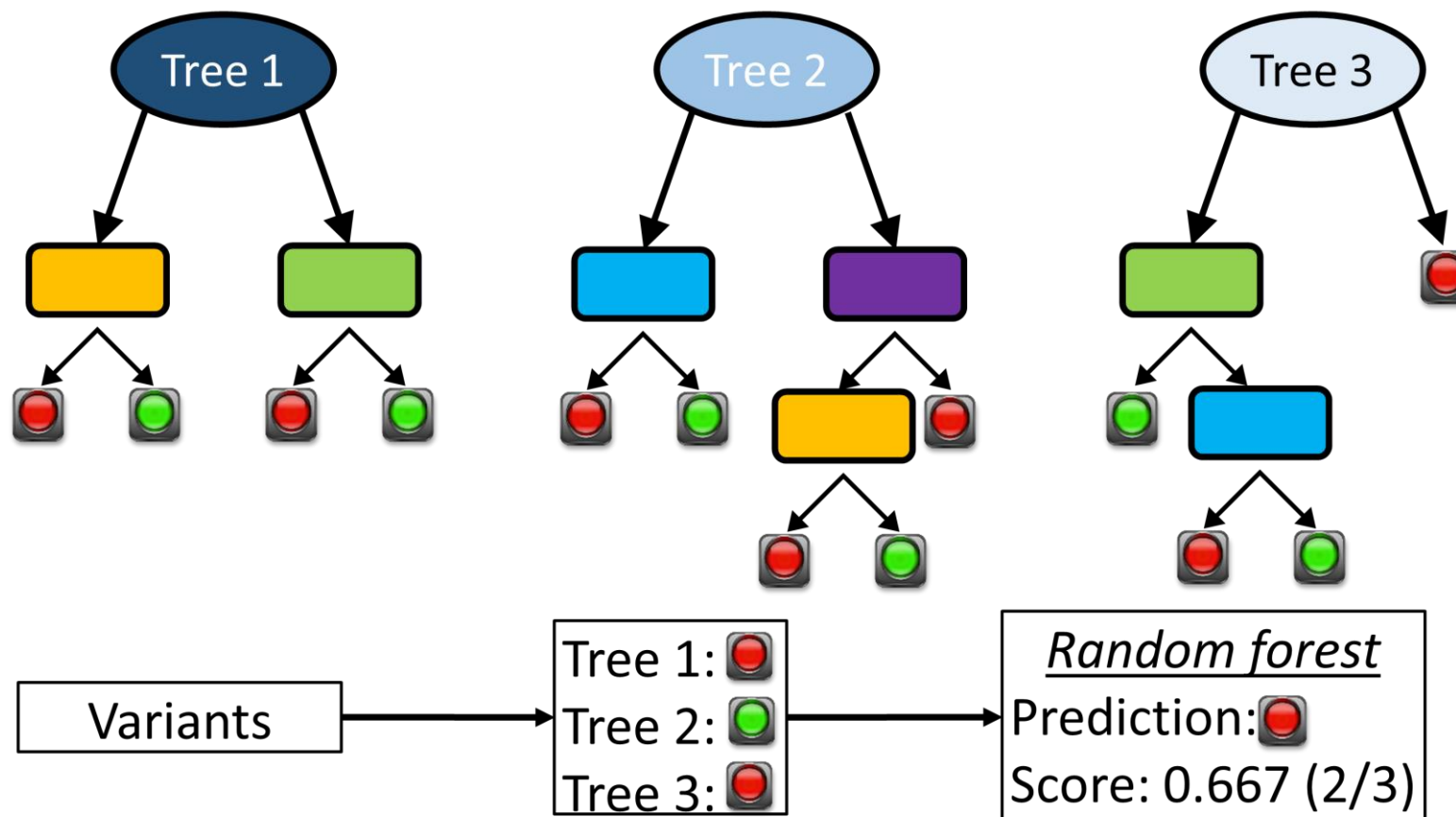
Set of scores used by SPiP



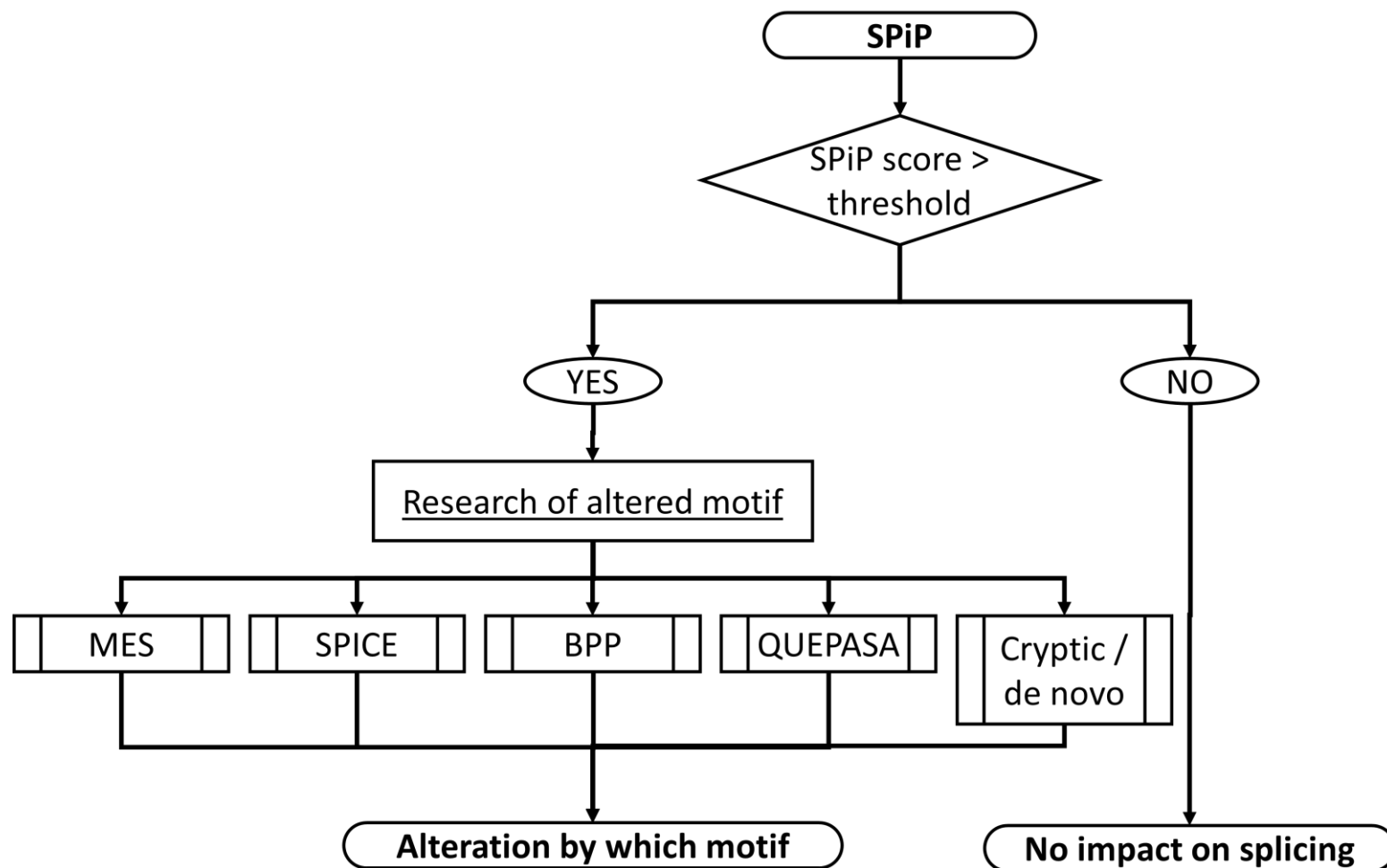
Scan for cryptic/*de novo* splice site

Score SPiP

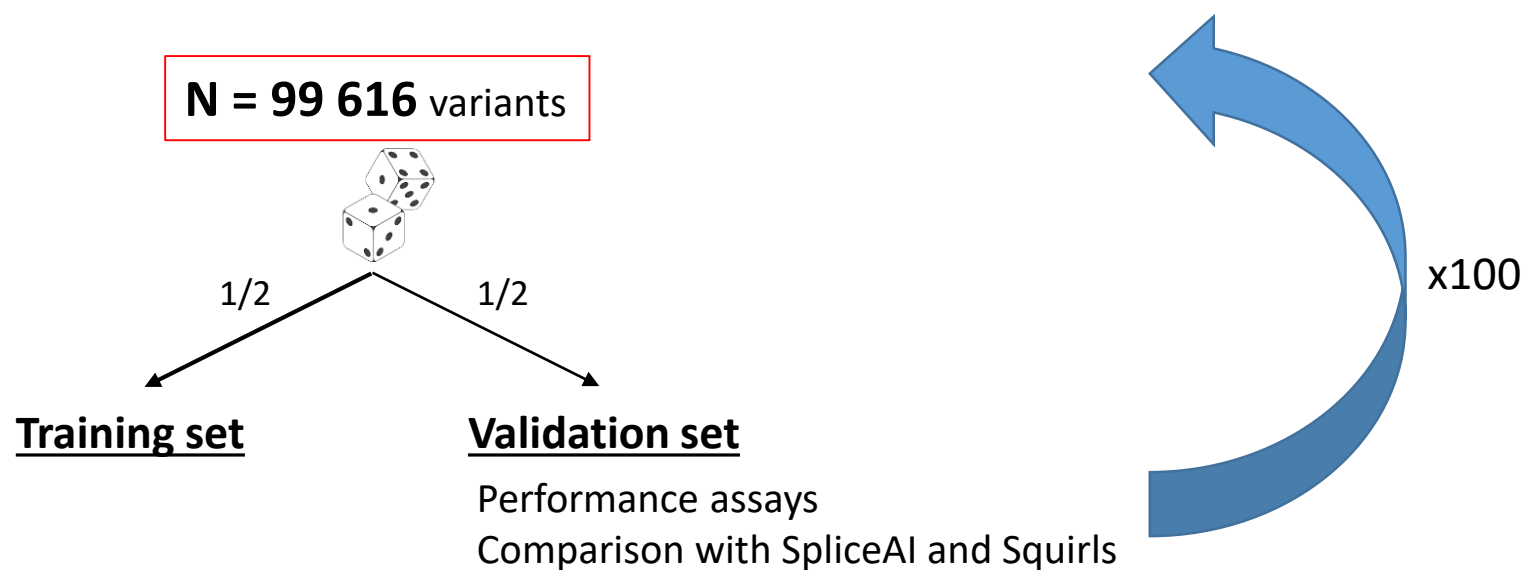
Random Forest



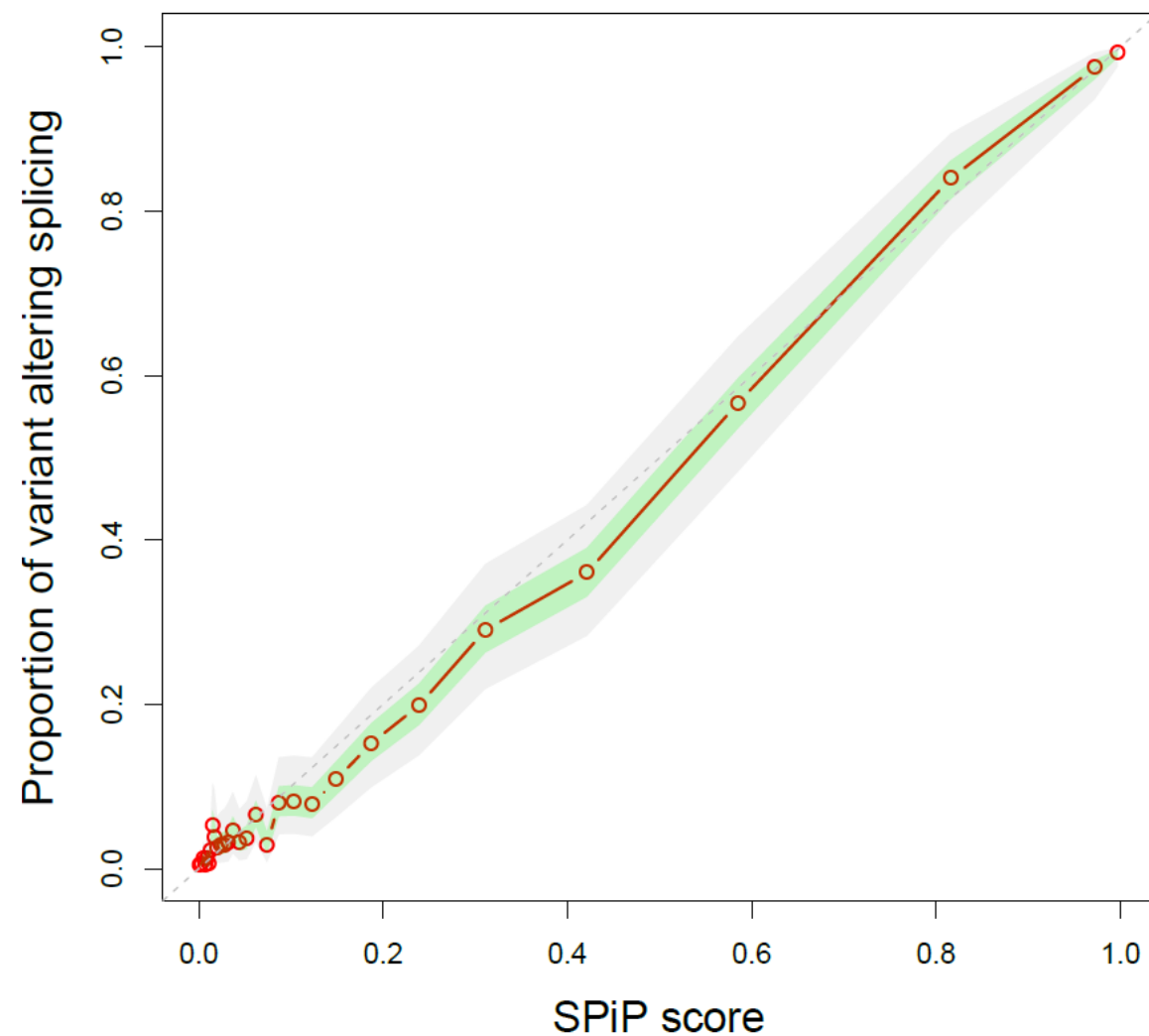
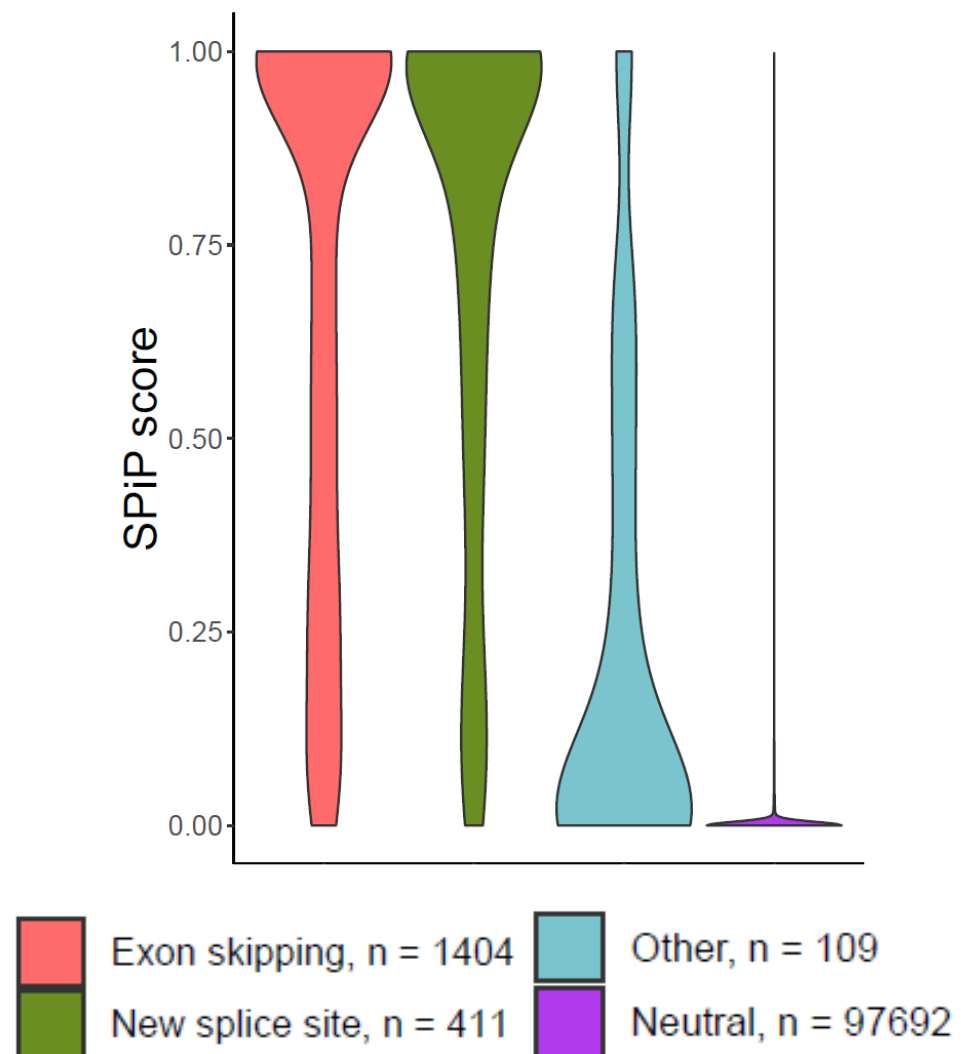
Annotations SPiP



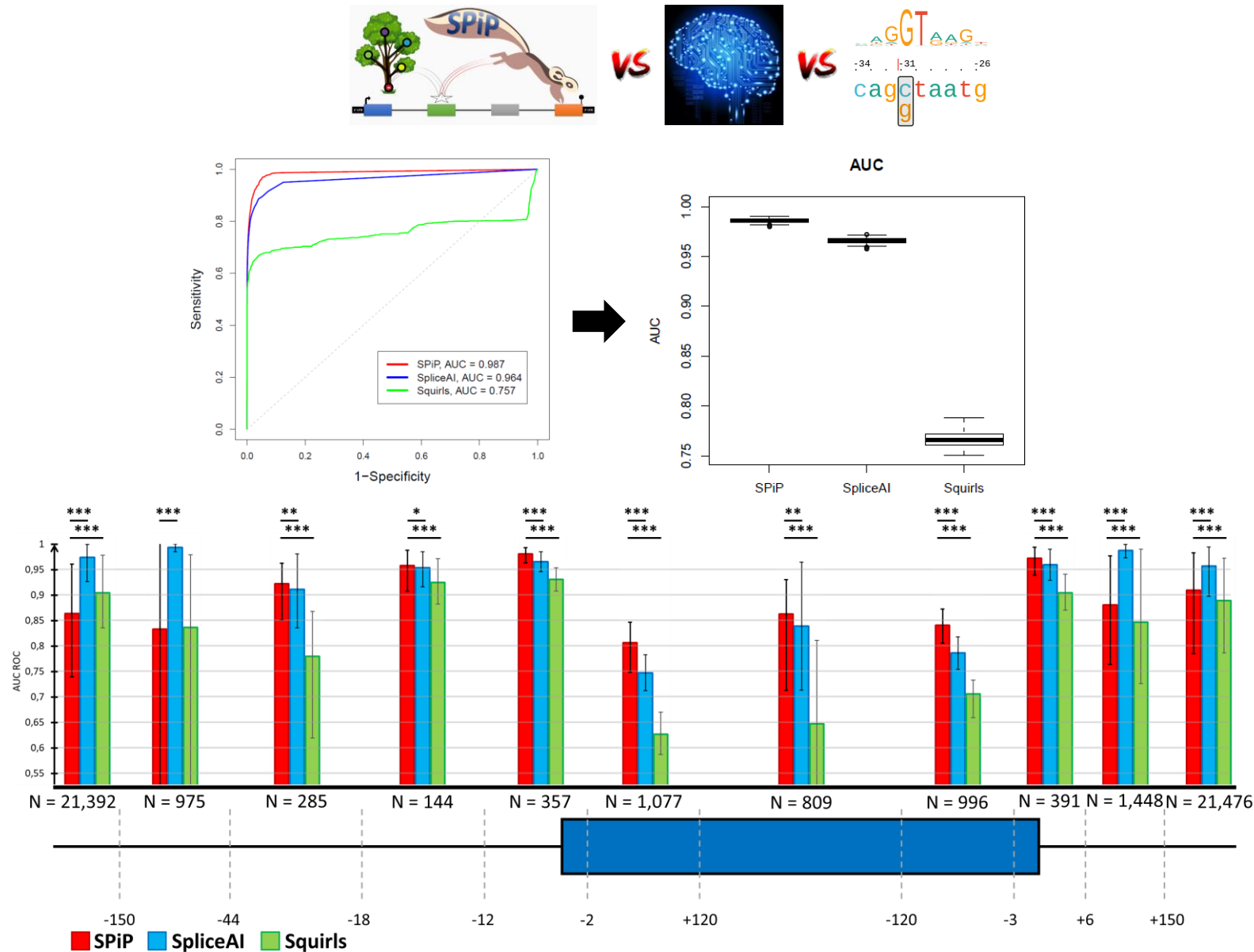
Training and validation sets



Performances SPiP

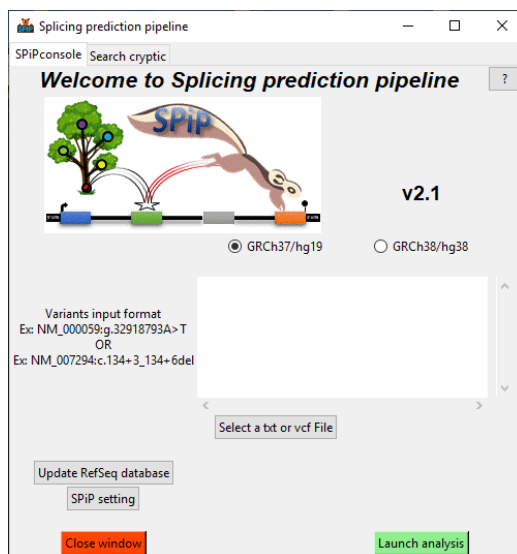


SPiP compared to machine learning tools



Using SPiP

Windows



- Mutation ID: NM_007294:c.4096+3A>T
- Stand alone app



<https://sourceforge.net/projects/splicing-prediction-pipeline/>

Linux



```
$ Rscript /path/to/SPiPv2,1.r \
-I /path/to/inputFile.vcf \
-O /path/to/outputFile.txt \
```

- Batch analyses
- Input: .txt or .vcf files
- Parallelizable



<https://github.com/raphaelleman/SPiP>

Internet

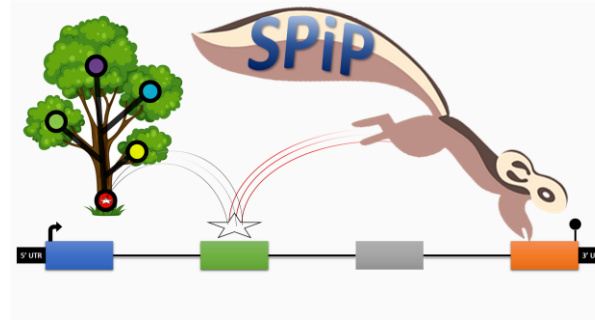


Baux, D., Van Goethem, C., Ardouin, O. et al.
MobiDetails: online DNA variants interpretation.
Eur J Hum Genet (2020).

- Implementation of last version of SPiP

<https://mobidetails.iurc.montp.inserm.fr/MD/>

Conclusion



- ✓ SPiP allows the extension of splicing predictions wherever the variant position in gene
- ✓ Helpful to prioritize RNA *in vitro* studies
- ✓ Outperforms SpliceAI and Squirrels
 - ✓ For consensus splice sites, polypyrimidine tracts, branch points and exonic regions

Thanks for your attention

A great thank you to all our collaborators:

Béatrice PARFAIT
Dominique VIDAUD
Emmanuelle GIRODON
Laurence PACOT
Gérald LE GAC
Chandran KA
Claude FEREC
Yann FICHO
Céline QUESNELLE
Etienne MULLER

Dominique VAUR
Laurent CASTERA
Agathe RICOU
Hélène TUBEUF
Omar SOUKARIEH
Pascaline GAILDRAT
Florence Riant
Marine GUILLAUD-BATAILLE
Sandrine CAPUTO
Virginie CAUX-MONCOUTIER

Nadia BOUTRY-KRYZA
Françoise BONNET-DORION
Ines SCHULTZ
Maria ROSSING
Michael T. PARSONS
Amanda B. SPURDLE
Thierry FRÉBOURG
Alexandra MARTINS
Claude HOUDAYER
Sophie KRIEGER

CAGI6 Challenge

