



GGATGAGATTCT
TCGAAAATCTGC



GGATGAGATTCT
TCGAAAATCTGC



01001011010010
1011001010110
0101110101011
1001001010010
1101101011011



01001011010010
1011001010110
0101110101011
1001001010010
1101101011011



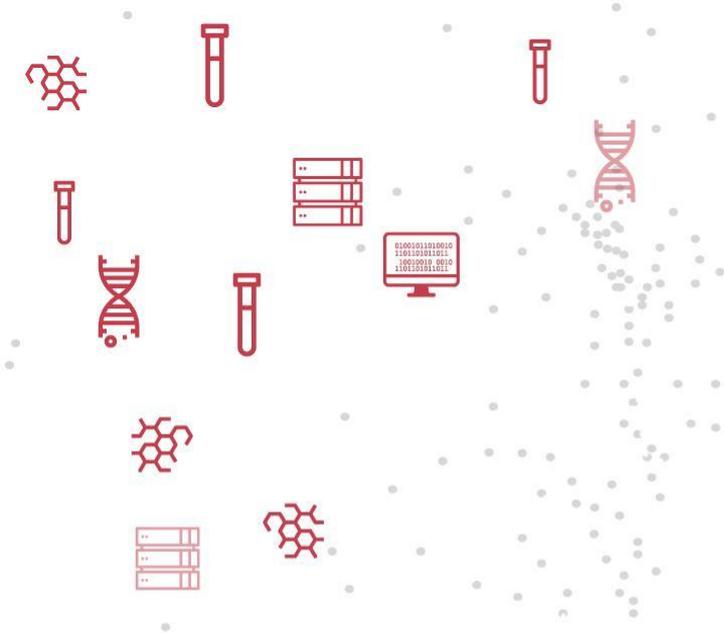
BITGENIA

—
TRANSCENDING
GENOMICS

Impacto de la Inteligencia Artificial en Genómica Clínica

Bioq. NICOLÁS M. AGUIRRE

Analista de Datos Genómicos | Bitgenia
Contacto: nicolas.aguirre@bitgenia.com



Introducción

Inteligencia Artificial



Subdisciplina del campo de la informática que busca la creación de máquinas que puedan imitar comportamientos inteligentes.

- Vehículos Autónomos
- Generación de Imágenes
- Generación de Texto
- Imitación de voz

BITGENIA

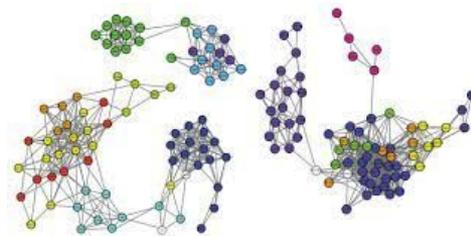


Aprendizaje Automático (Machine Learning)

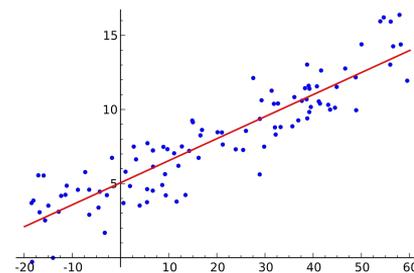
Rama del campo de la IA, que busca como dotar a las máquinas de aprendizaje.

Métodos:

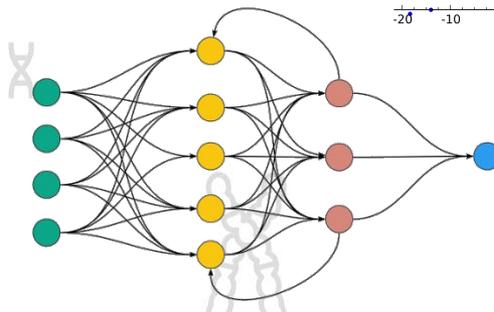
- Árboles de decisión
- Modelos de regresión
- Modelos de clasificación
- Técnicas de clusterización
- Redes Neuronales:
 - Información Jerarquizada
 - Aumento de capas de abstracción
 - Deep Learning



Clusterización



Regresión Lineal

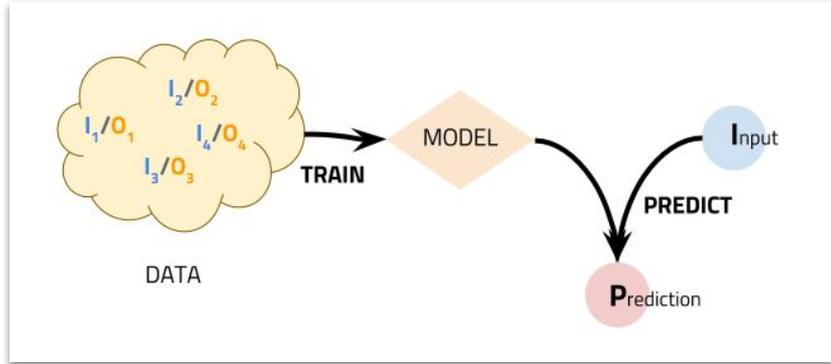


Redes Neuronales

B

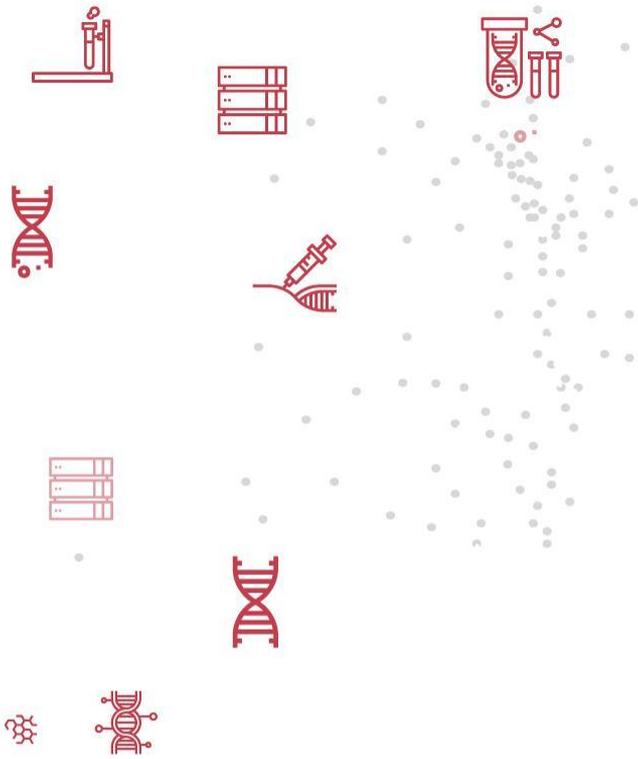
—

Cambio de Paradigma



Anteriormente se programaba un algoritmo que seguía paso a paso una determinada tarea. **Había que contemplar cada variable posible.**

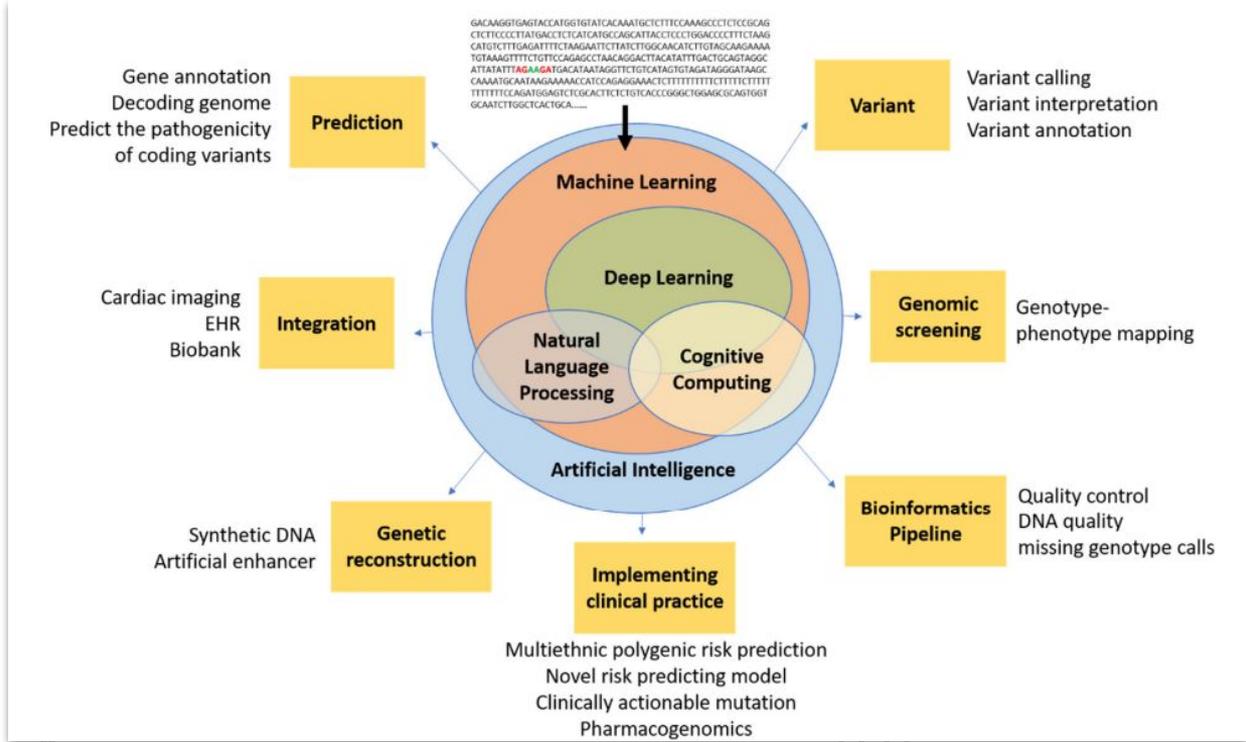
Con IA/ML se le da un objetivo y el programa va **aprendiendo hasta alcanzarlo**. Esto le otorga una capacidad disruptiva.



Actualidad



Inteligencia Artificial en Genómica



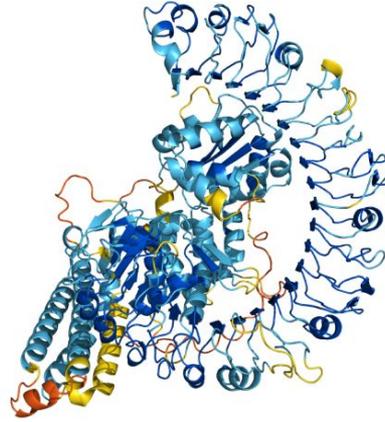
A

B

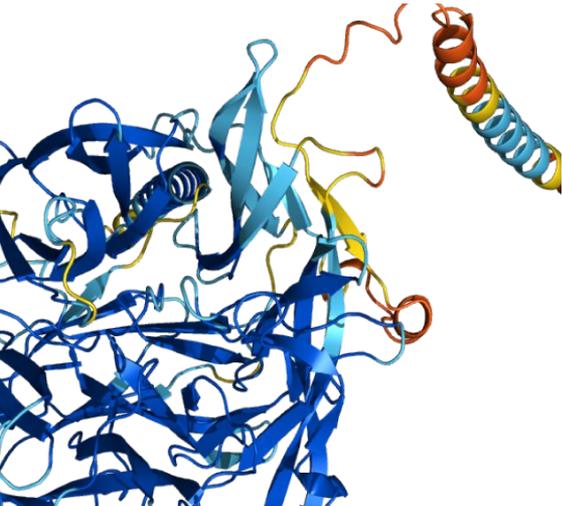


AlphaFold

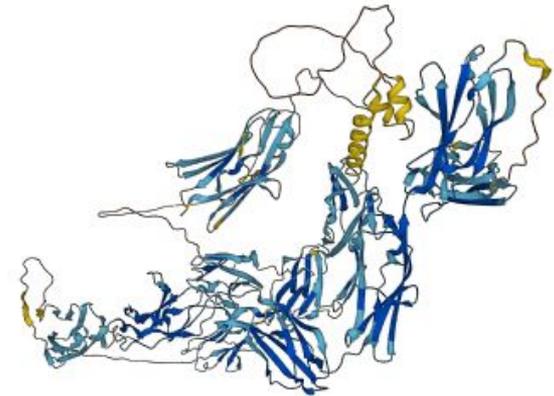
Es un Sistema de Inteligencia artificial que **predice** la estructura **3D** de las proteínas, **superando** en algunos casos la exactitud de los experimentos.



200 millones de predicciones,
50.000 estructuras proteicas
humanas



Utiliza nuevas **redes neuronales** y entrenamientos basados en **restricciones evolutivas, físicas y geométricas.**



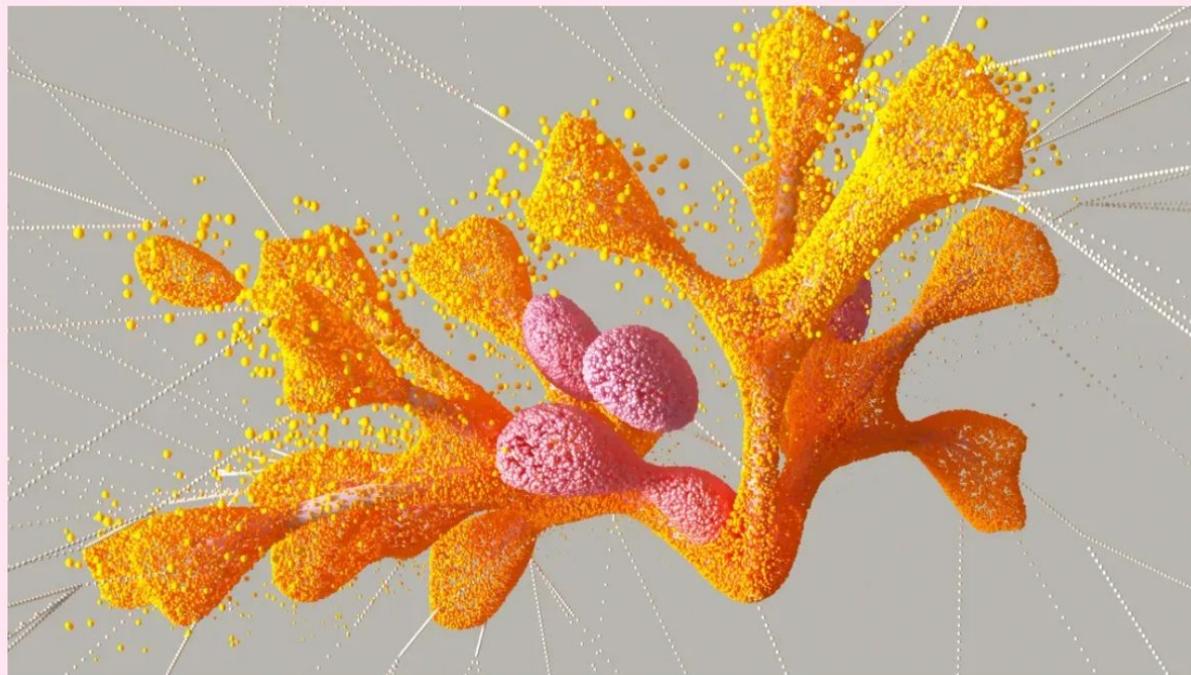


Alpha Missense



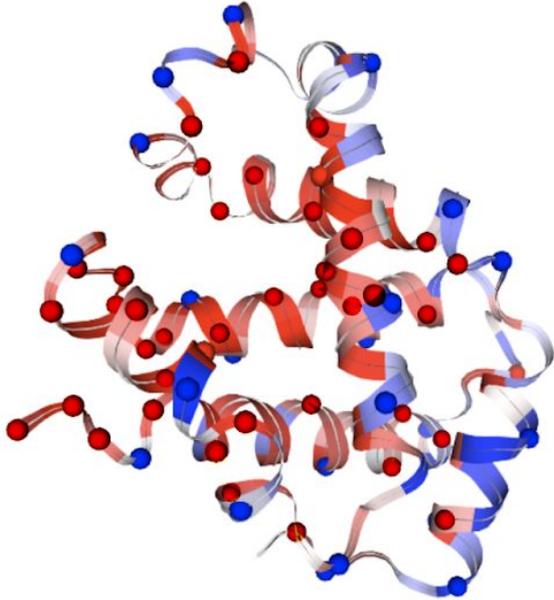
A catalogue of genetic mutations to help pinpoint the cause of diseases

September 19, 2023





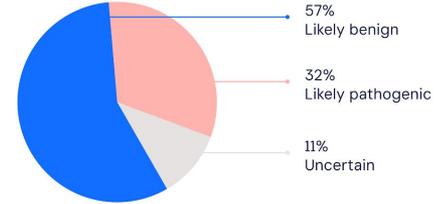
Alpha Missense



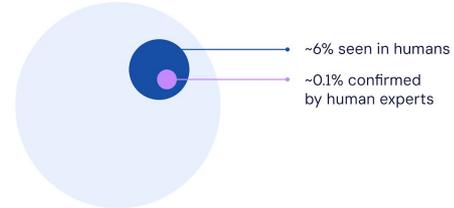
Hemaglobin subunit beta (HBB)

All possible 71 million human missense variants

AlphaMissense predictions:



Human annotations:



AlphaMissense predijo la patogenicidad de **71 millones de posibles variantes Missense**.

Clasificó el 89%, prediciendo el 57% como benignos y el 32% patógenos.

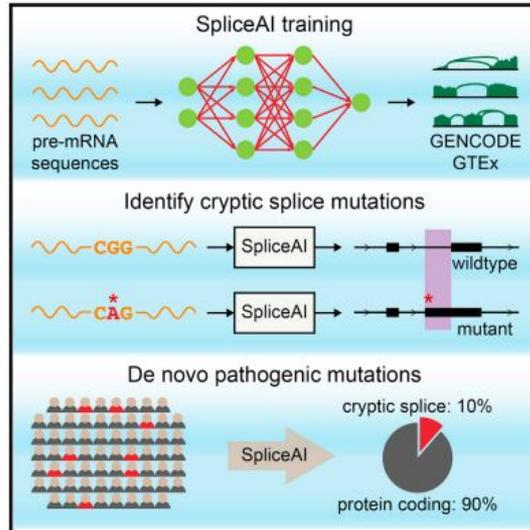
B



SpliceAI

Predicting Splicing from Primary Sequence with Deep Learning

Graphical Abstract



Authors

Kishore Jaganathan,
Sofia Kyriazopoulou Panagiotopoulou,
Jeremy F. McRae, ..., Serafim Batzoglou,
Stephan J. Sanders, Kyle Kai-How Farh

Correspondence

kfarh@illumina.com

In Brief

A deep neural network precisely models mRNA splicing from a genomic sequence and accurately predicts noncoding cryptic splice mutations in patients with rare genetic diseases.

- Es una **red neuronal profunda** de 32 capas, que **predice el splicing** a partir de una secuencia de pre-ARNm
- **75%** de las variantes de splicing crípticas predichas **se confirman** en RNA-seq
- Las **variantes de splicing críptico** dan lugar con frecuencia a splicing alternativo, que puede derivar en **variantes truncantes**.



BITGENIA

— TRANSCENDING
GENOMICS

En Bitgenia

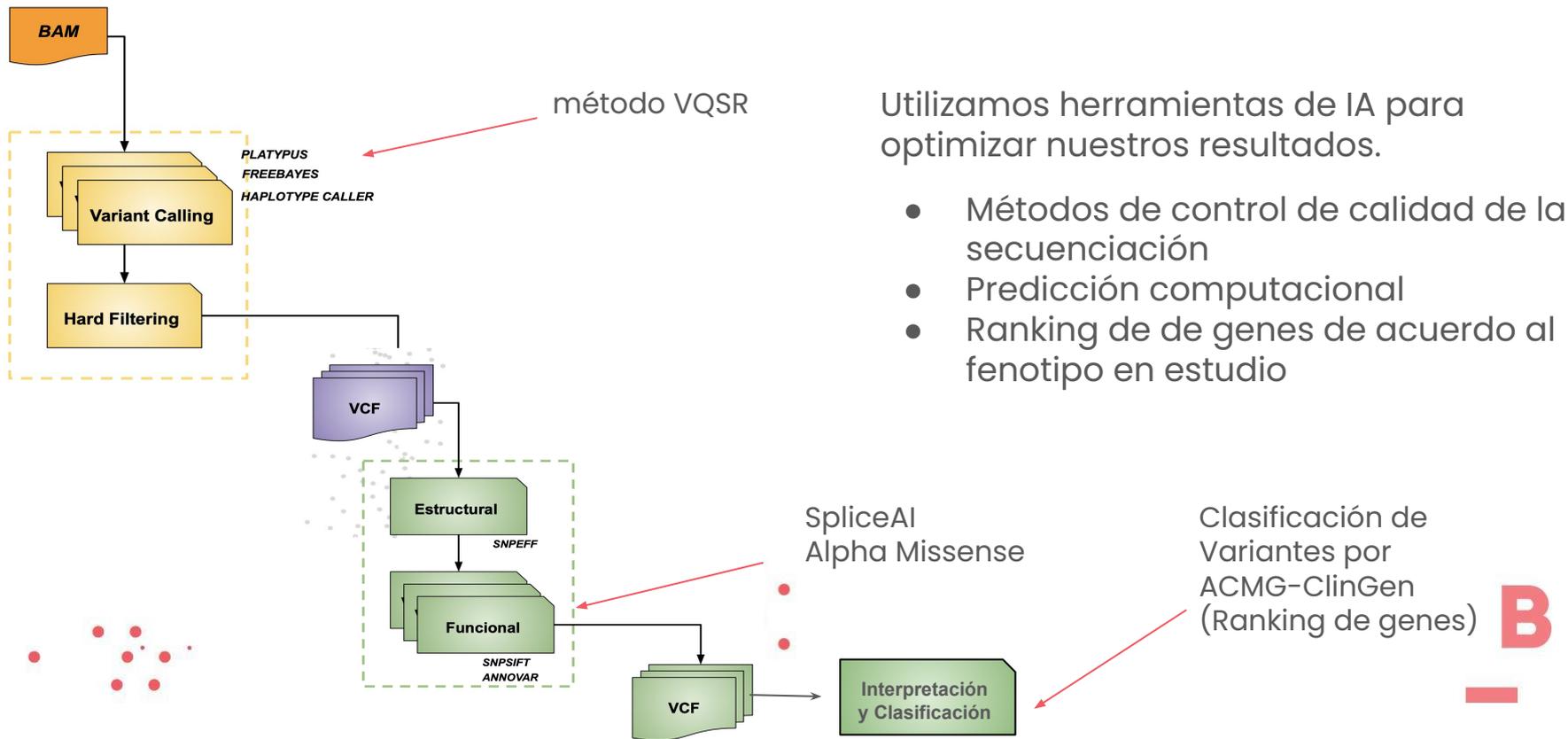
Venimos desarrollando e implementando herramientas que faciliten y mejoren el análisis de variantes genómicas



010001 1 1 0
1 1 011 01 0
01011001 1

GA TT CT
A AA TC TGC
G A GA TT CT
G T A AA TC TG

Incorporación de IA en Análisis Genómico



Modelado bayesiano de la clasificación de variantes

Modeling the ACMG/AMP Variant Classification Guidelines as a Bayesian Classification Framework

$$Post_P = \frac{OddsPath * Prior_P}{((OddsPath - 1) * Prior_P + 1)}$$

Sean V. Tavtigian^{1,†}, Marc S. Greenblatt², Steven M. Harrison³, Robert L. Nussbaum⁴, Snehit A. Prabhu⁵, Kenneth M. Boucher⁶, and Leslie G. Biesecker⁷ on behalf of the ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI)[‡]

Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines

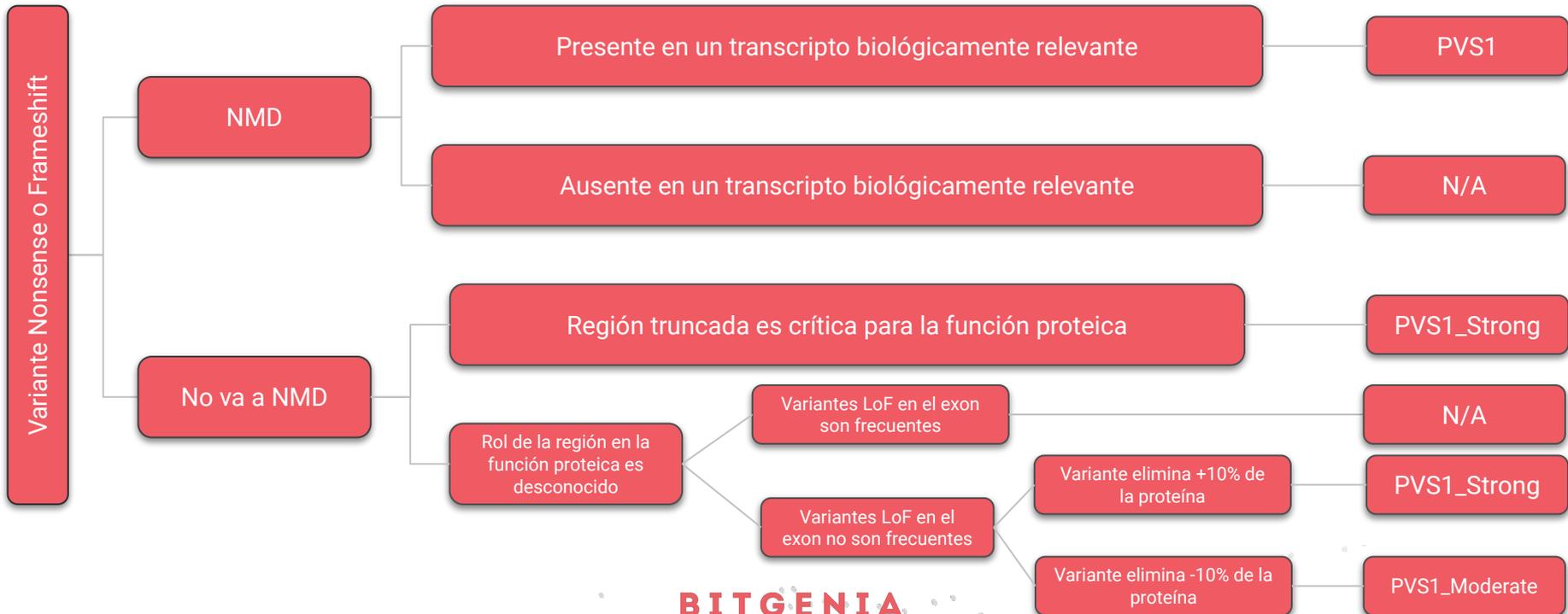
[Sean V. Tavtigian](#),^{1,2} [Steven M. Harrison](#),³ [Kenneth M. Boucher](#),^{2,4} and [Leslie G. Biesecker](#)⁵

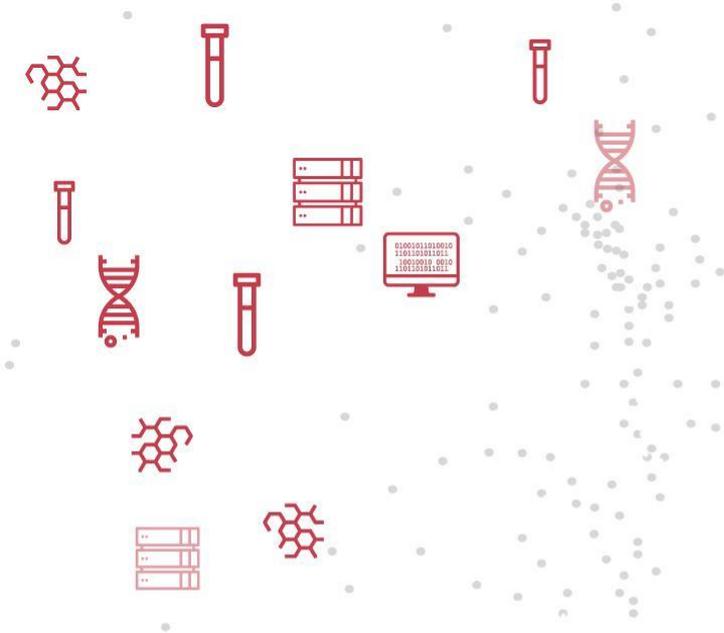
B

—



PVS1 (Criterio para variantes LoF)





Futuro



Árboles de Decisión



Variants located outside of donor/acceptor $\pm 1,2$ dinucleotide positions

SpliceAI Δ score ≤ 0.1

BP4

Synonymous (silent) variants and intronic variants outside donor and acceptor splice regions

Yes

No

BP7

BP7 N/A

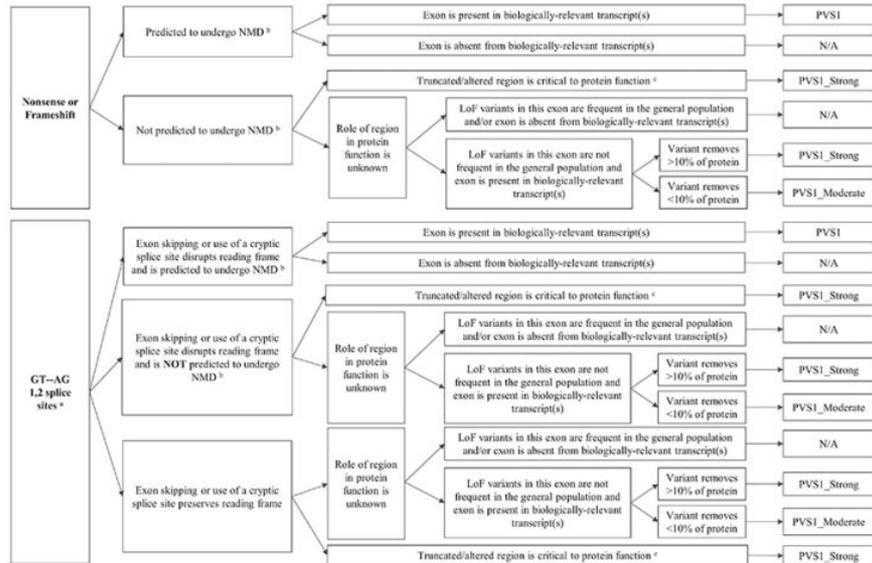
SpliceAI Δ score > 0.1 and < 0.2

PP3 N/A (Splicing)

Consider missense/indel predictions for exonic variants

SpliceAI Δ score ≥ 0.2

PP3

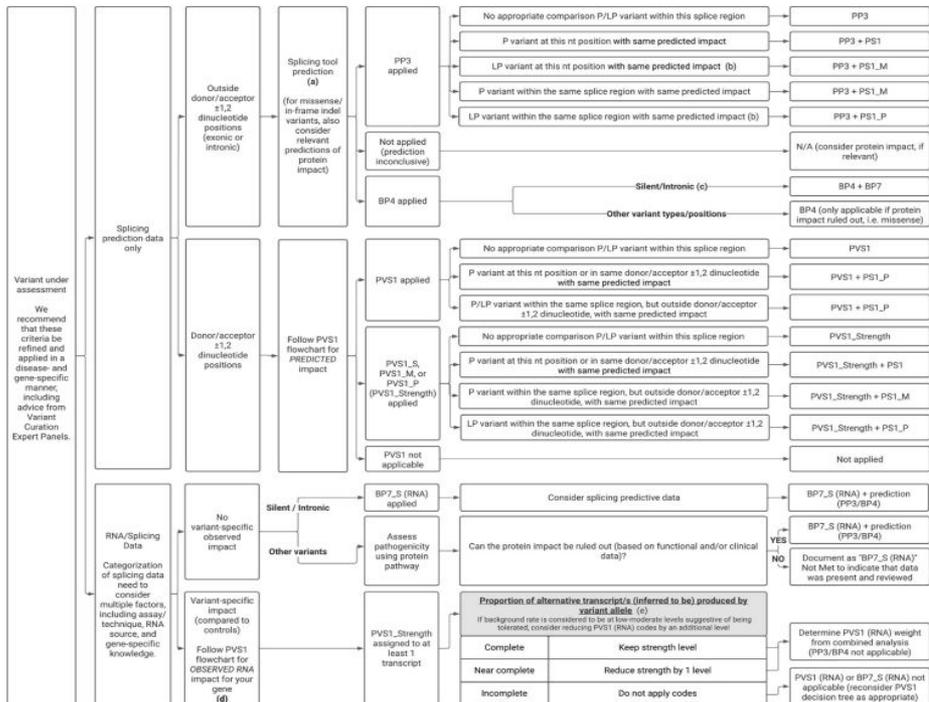
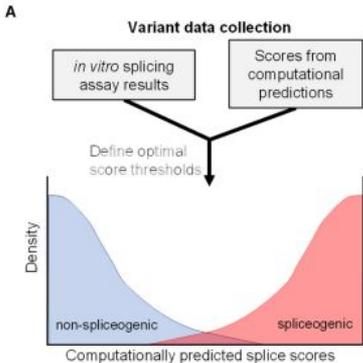


B



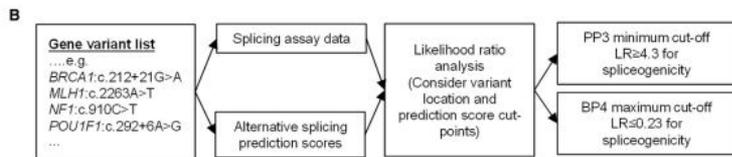
Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: Recommendations from the ClinGen SVI Splicing Subgroup

Logan C. Walker,¹ Miguel de la Hoya,² George A.R. Wiggins,¹ Amanda Lindy,³ Lisa M. Vincent,⁴ Michael T. Parsons,⁵ Daffodil M. Canson,⁵ Dana Bis-Brewer,³ Ashley Cass,⁶ Alexander Tchourbanov,⁶ Heather Zimmermann,⁶ Alicia B. Byrne,⁷ Tina Pesaran,⁶ Rachid Karam,⁶ Steven M. Harrison,^{6,7,*} Amanda B. Spurdle,^{5,8} and ClinGen Sequence Variant Interpretation Working Group



	Non-spliceogenic	Proportion	Spliceogenic	Proportion	Likelihood of event
Low prediction score	n1	$P1 = \frac{n1}{n1+n2+n3}$	n4	$P4 = \frac{n4}{n4+n5+n6}$	$LR1 = P4/P1$
Intermediate	n2	$P2 = \frac{n2}{n1+n2+n3}$	n5	$P5 = \frac{n5}{n4+n5+n6}$	$LR2 = P5/P1$
High prediction score	n3	$P3 = \frac{n3}{n1+n2+n3}$	n6	$P6 = \frac{n6}{n4+n5+n6}$	$LR3 = P6/P3$

Likelihood ratio for spliceogenic: 350 (Very strong), 17.7 (Strong), 4.3 (Moderate), 2.08 (Supporting), Likelihood ratio for non-spliceogenic: 0.481 (Supporting), 0.233 (Moderate), 0.053 (Strong), 0.003 (Very strong)



B

—



Proximos avances

Assess Variant Effect

*****CHOOSE a single predictor for use*****

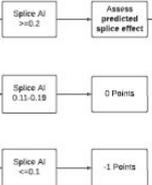
	-4	-3	-2	-1	0	+1	+2	+3	+4
BayesDel	≤ -0.52	-0.021 to 0.34	-0.021 to 0.16	0.35 to 0.16	0.17 to 0.12	0.13 to 0.26	+0.27 to +0.40	+0.41 to +0.50	≥ 0.50
MuPred2	≤ -0.10	0.011 to 0.0316	0.0316 to 0.191	0.191 to 0.392	0.392 to 0.727	0.727 to 0.829	0.829 to 0.931	0.931 to 0.992	≥ 0.992
REVEL	≤ 0.016	0.017 to 0.052	0.052 to 0.184	0.184 to 0.290	0.291 to 0.643	0.644 to 0.772	0.773 to 0.878	0.879 to 0.931	≥ 0.932
VEST	≤ 0.077	0.078 to 0.302	0.303 to 0.449	0.450 to 0.763	0.764 to 0.880	0.881 to 0.908	0.909 to 0.964	0.965 to 0.995	≥ 0.995

Exon/Transcript Relevance	
Residue present in all transcripts	Full score
Residue is present in most disease-specific or highly expressed transcripts	Half score
Residue is not expressed in clinically relevant transcripts	No score

Missense Variant

Assess AA Change Prediction

Assess Splice Change Prediction

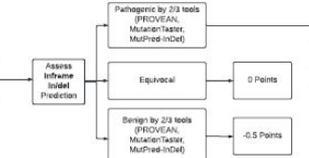


Disrupted reading frame leading to NMD

Disrupted reading frame or preserved reading frame NOT leading to NMD

Small Inframe InDel (<10% of protein) NOT leading to NMD

Exon/Transcript Relevance	
Residue present in all transcripts	+1.5
Residue absent in some transcripts but present in most disease-specific or highly expressed transcripts	+0.5
Residue is not expressed in clinically relevant tissue(s)	No score

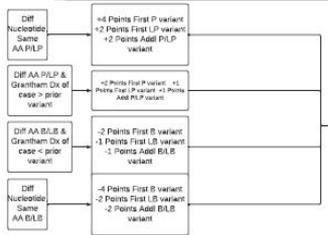


Exon/Transcript Relevance	
Residue present in all transcripts	+3.5
Residue is present in most disease-specific or highly expressed transcripts	+2.5
Residue is not expressed in clinically relevant tissue(s)	No score

Region Information	
Role of region in protein function is unknown	+0
Removes/alters entire critical functional domain relevant to disease mechanism OR removes/alters >50% of protein	+2
Partially removes/alters critical functional domain relevant to disease mechanism OR removes/alters >25% of protein	+1
Removes/alters region with plausible role in relevant protein function OR removes/alters >10% of protein	+0.5
Removes/alters <10% of protein	+0

Exon/Transcript Relevance	
Residue present in all transcripts	+1.5
Residue absent in some transcripts but present in most disease-specific or highly expressed transcripts	+0.5
Residue is not expressed in clinically relevant tissue(s)	No score

Assess Comparison Variants



Informative Variants - NMD effect	
PLP variant in the same donor or acceptor with same predicted splicing impact	+2 for first P variant +1 for first L/P variant +1 per additional
No informative variants in exon	+0
B/E variant in the same donor or acceptor with same predicted splicing impact	-2 for first B variant -1 for first L/B variant -1 per additional

Informative Variants - non-NMD effect	
PLP variant in the same donor or acceptor with same predicted splicing impact OR PLP variant that removes the same sequence as VUA	+2 for first variant +1 per additional
No informative variants in exon	+0
B/E variant in the same donor or acceptor with same predicted splicing impact OR B/E variant that removes the same sequence as VUA	-2 per variant

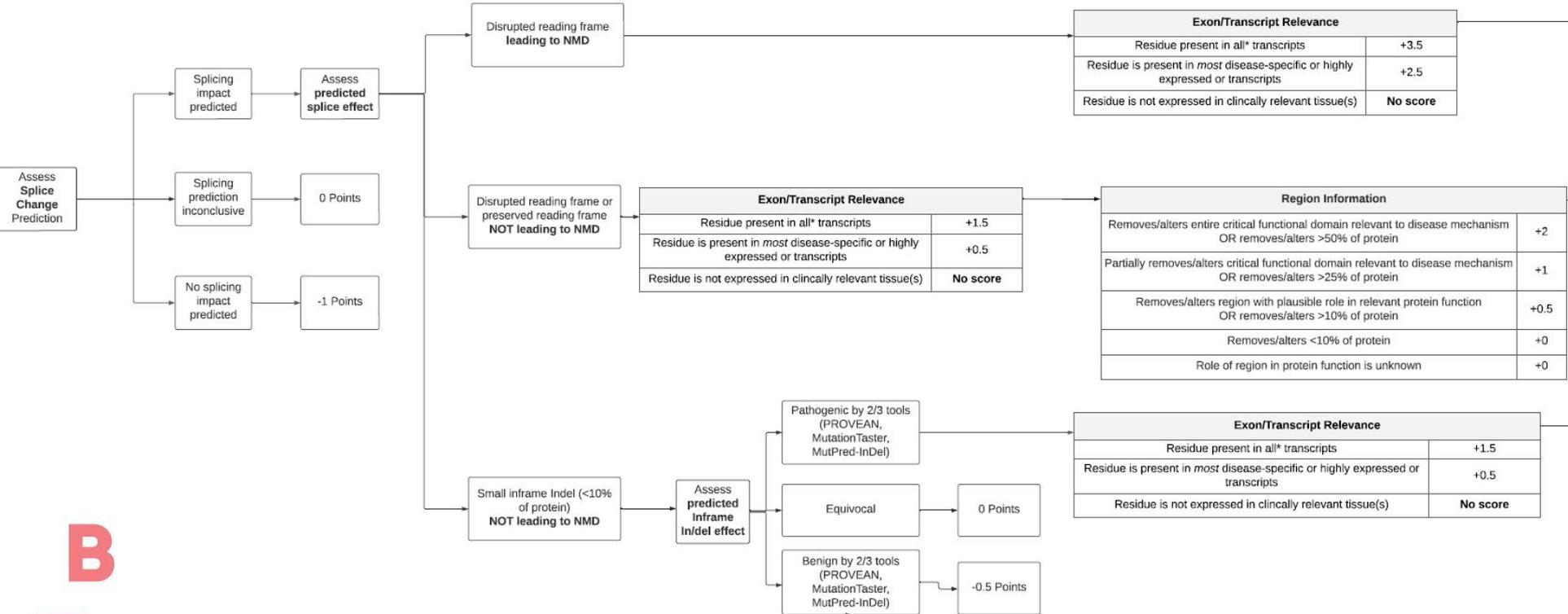
Select the more positive value and proceed with that path only
*** Compare the total score not capped score ***

IMP_MSS -4 to +9 (cap at these levels)

IMP_SPL -4 to +8 (cap at these levels)

B

Proximos avances

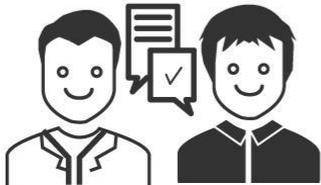


B

—



Conclusiones



B





Conclusión

La IA/ML está aportando un **nuevo modo de explotación** de la información brindándonos la capacidad de hacer análisis **n-dimensionales**.

La información está en proceso de revisión para **eliminar los sesgos humanos** ya que estos van a estar incluidos en cualquier modelo de aprendizaje automatizado.

La potencia de la inteligencia artificial está a la vista, es nuestro deber formar parte de los grupos de decisión para hacer un **uso ético de los mismos**.



Referencias

¿Qué es el Machine Learning? ¿Y Deep Learning? Un mapa conceptual | DotCSV

Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>

Tunyasuvunakool, K., Adler, J., Wu, Z. et al. Highly accurate protein structure prediction for the human proteome. *Nature* 596, 590–596 (2021). <https://doi.org/10.1038/s41586-021-03828-1>

Overview of ACMG/AMP v4 Sequence Variant Guidelines – ClinGen | Clinical Genome Resource

Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, Biesecker LG; ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med*. 2018 Sep;20(9):1054–1060. doi: 10.1038/gim.2017.210. Epub 2018 Jan 4. PMID: 29300386; PMCID: PMC6336098.

Tavtigian SV, Harrison SM, Boucher KM, Biesecker LG. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat*. 2020 Oct;41(10):1734–1737. doi: 10.1002/humu.24088. Epub 2020 Aug 30. PMID: 32720330; PMCID: PMC8011844.

Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglu S, Sanders SJ, Farh KK. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019 Jan 24;176(3):535–548.e24. doi: 10.1016/j.cell.2018.12.015. Epub 2019 Jan 17. PMID: 30661751.

<https://alphafold.ebi.ac.uk/>

<https://www.deepmind.com/blog/alphamissense-catalogue-of-genetic-mutations-to-help-pinpoint-the-cause-of-diseases>

BITGENIA

BITGENIA

— TRANSCENDING
GENOMICS

Bioq. NICOLÁS M. AGUIRRE

Analista de Datos Genómicos | Bitgenia

Contacto: nicolas.aguirre@bitgenia.com

INFO@BITGENIA.COM -  54 9 11 6724 2109 - BUENOS AIRES, ARGENTINA

WWW.BITGENIA.COM