VaCS: variant prioritization using RNA sequencing in cell therapies

Rafaella Buzatu

Prof. Micha Drukker
Asst. Prof. Christian Schröter
Tineke van den Hoorn
Marcel Hoefnagel

Universiteit Leiden

CBG MEB

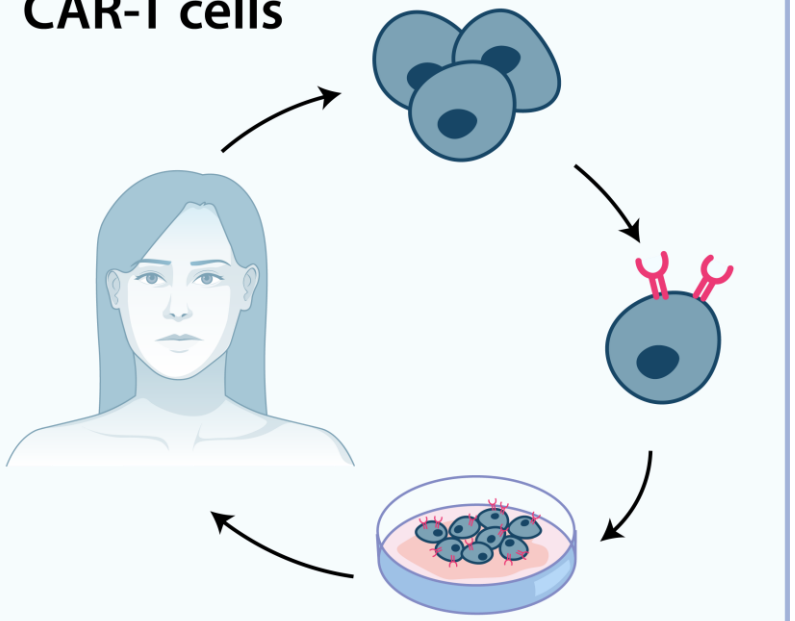
Advanced Therapeutic Medicinal Products (ATMPs)

- Cell Therapies 
- Gene Editing Products 
- Tissue Engineered Medicines 
- Combined ATMPs

Advanced Therapeutic Medicinal Products (ATMPs)

- Cell Therapies
- Gene Editing Products
- Tissue Engineered Medicines
- Combined ATMPs

CAR-T cells



Advanced Therapeutic Medicinal Products (ATMPs)

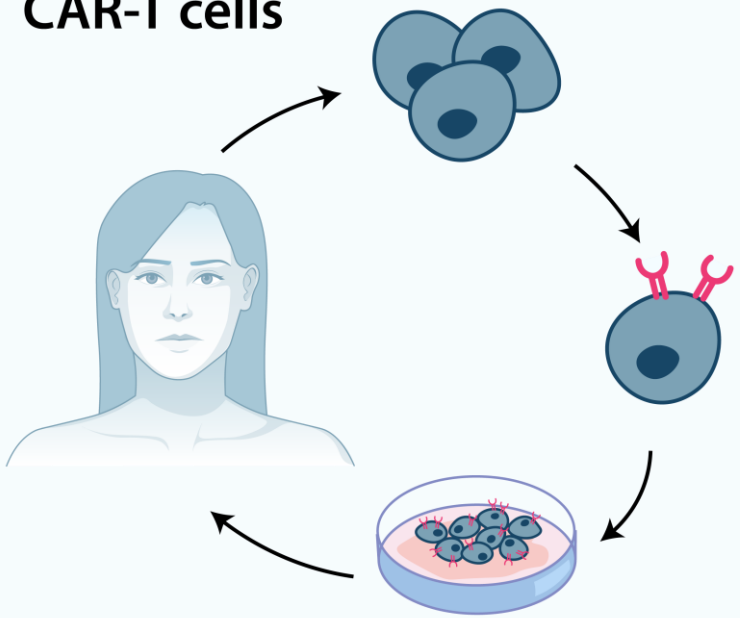
Cell Therapies

Gene Editing Products

Tissue Engineered Medicines

Combined ATMPs

CAR-T cells



iPSC-derived cell products



iPSCs

Advanced Therapeutic Medicinal Products (ATMPs)

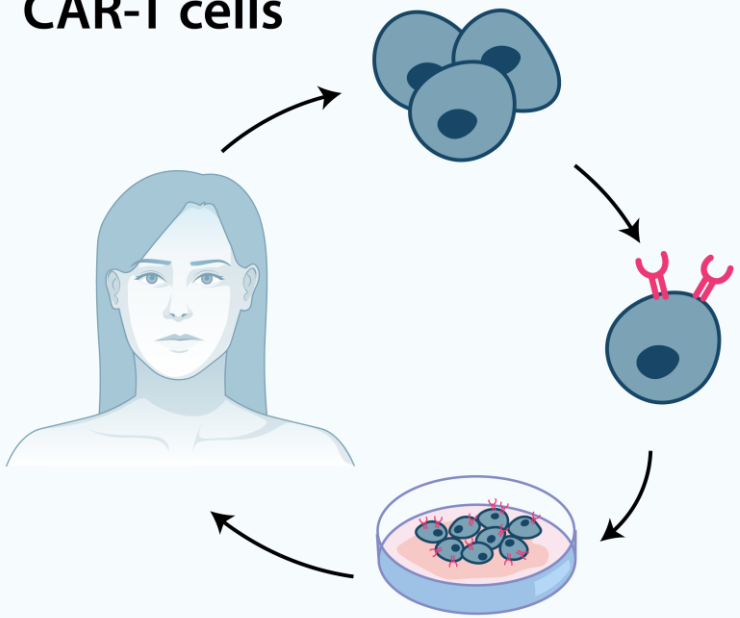
Cell Therapies

Gene Editing Products

Tissue Engineered Medicines

Combined ATMPs

CAR-T cells



iPSC-derived cell products

Cardio-myocytes

Neuro-mesoderm

Mesoderm

Fibroblasts

Primitive streak

iPSCs

Pancreas

Endoderm

Lung

Intestinal cells

Neuro-ectoderm

Motor neurons

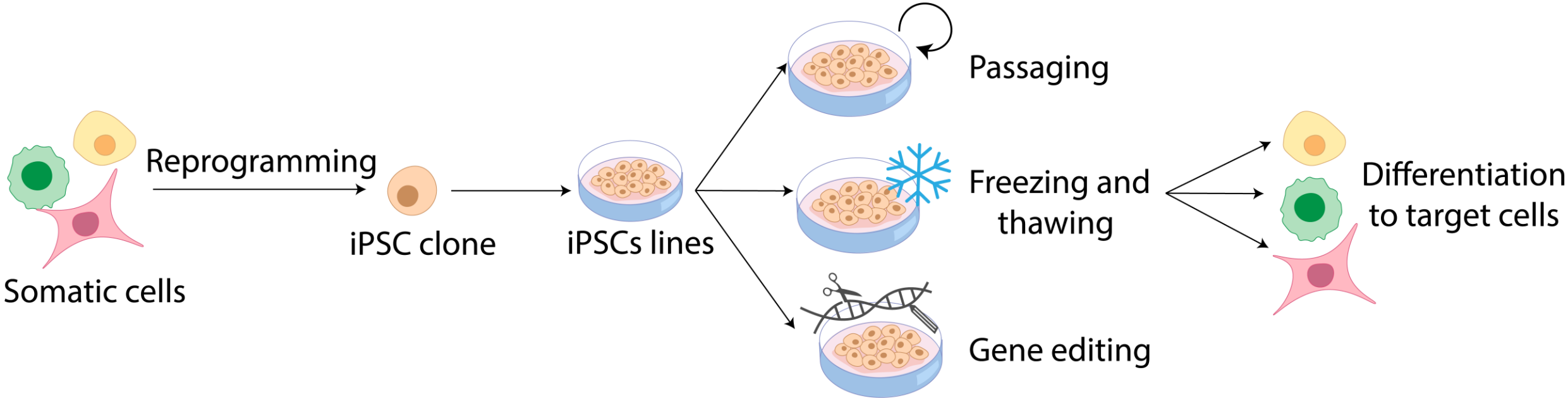
Sensory neurons

Ectoderm

Surface ectoderm

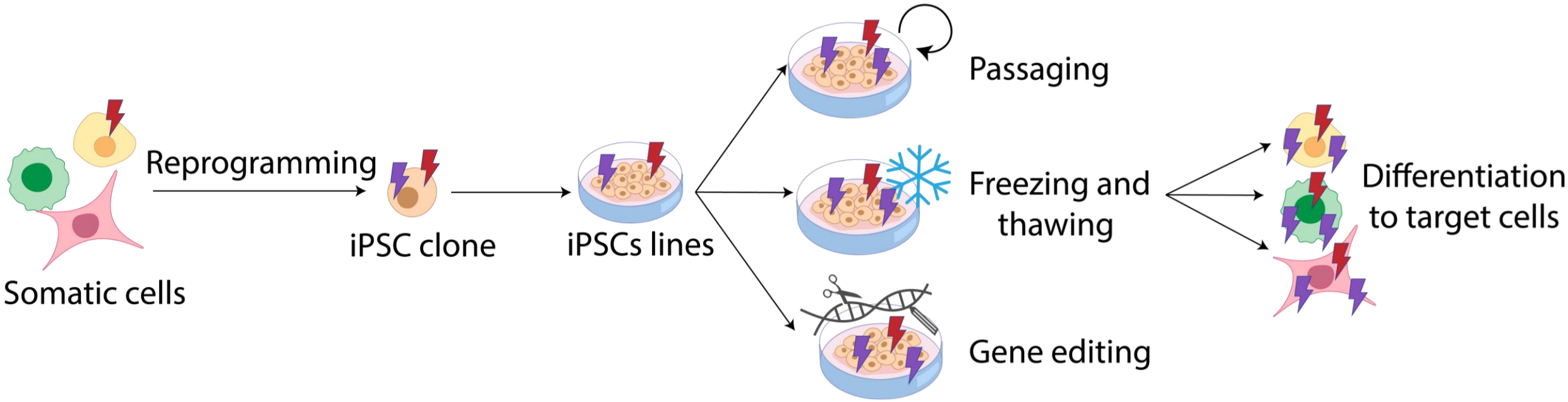
Epithelial cells

DNA variation can accumulate and persist during manufacturing

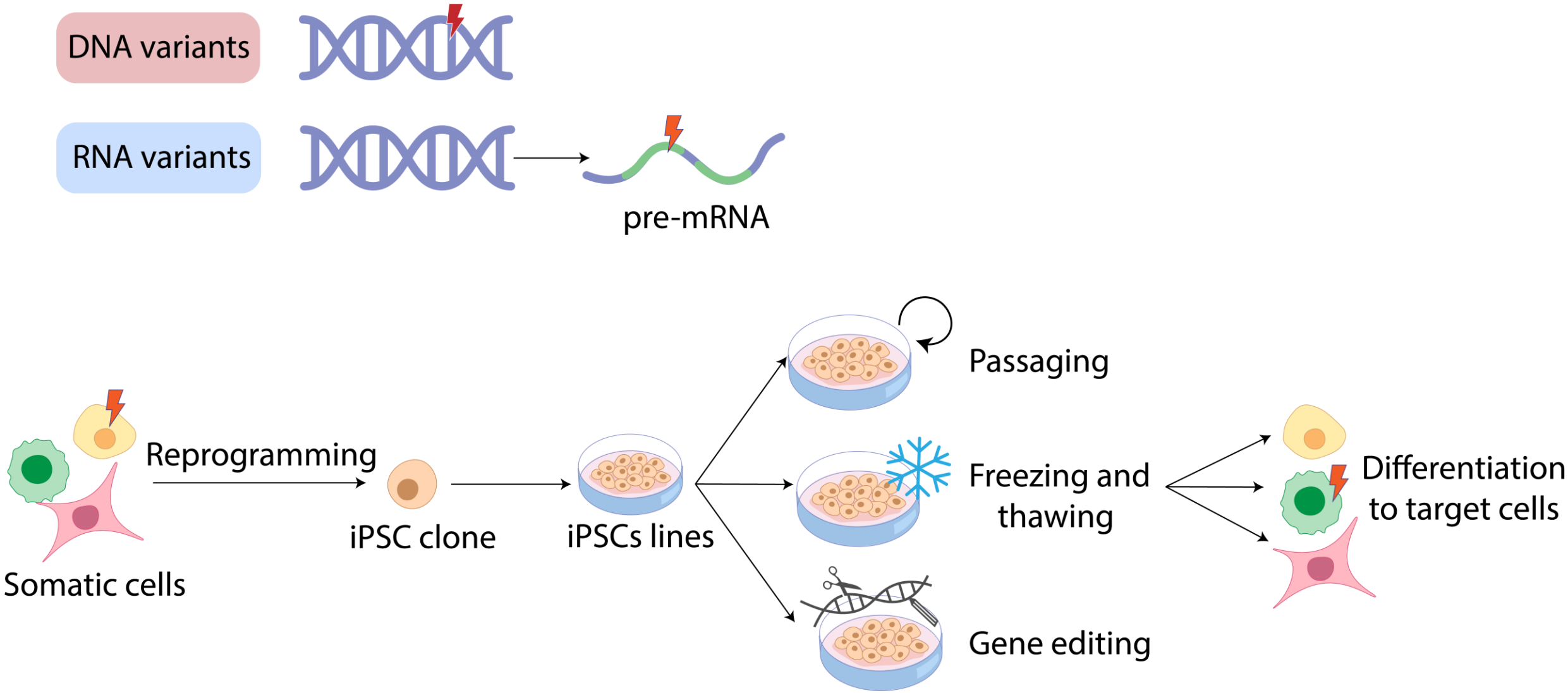


DNA variation can accumulate and persist during manufacturing

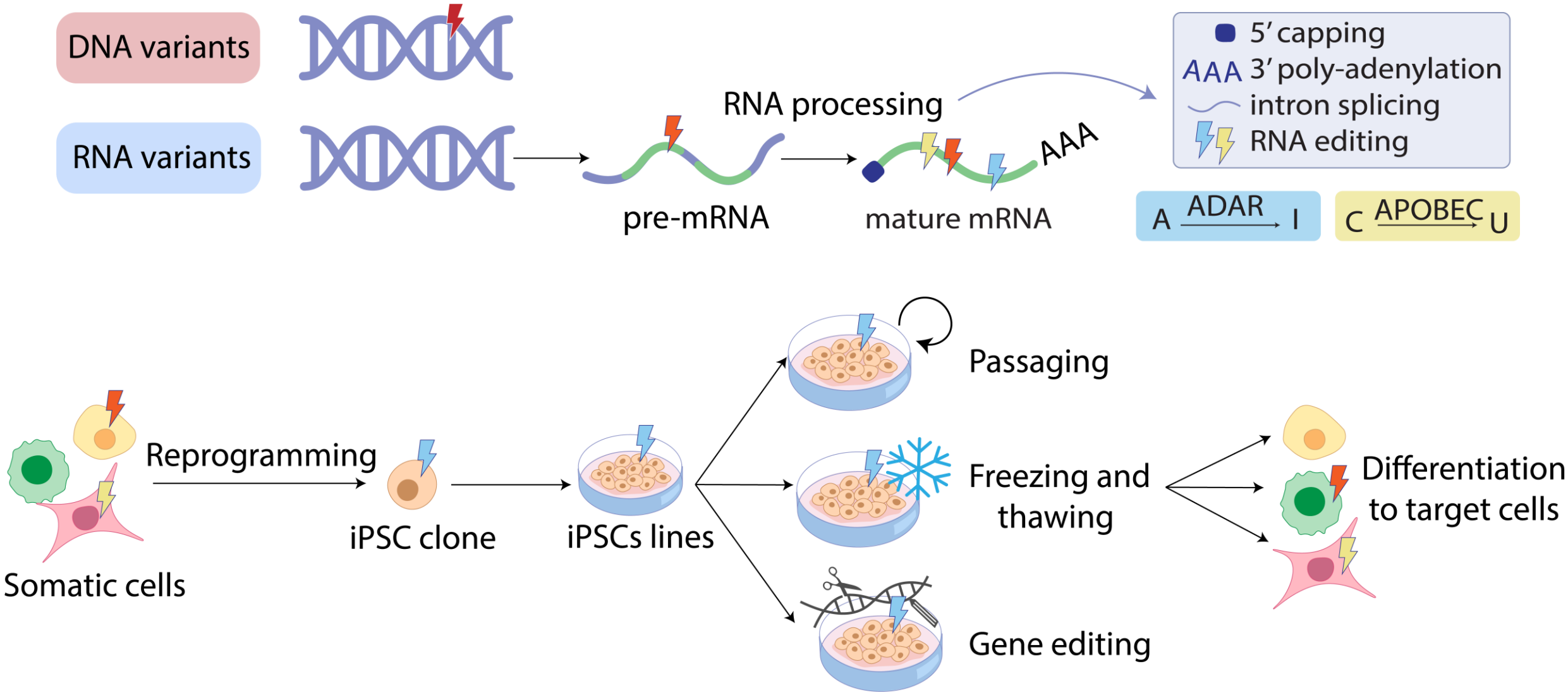
DNA variants



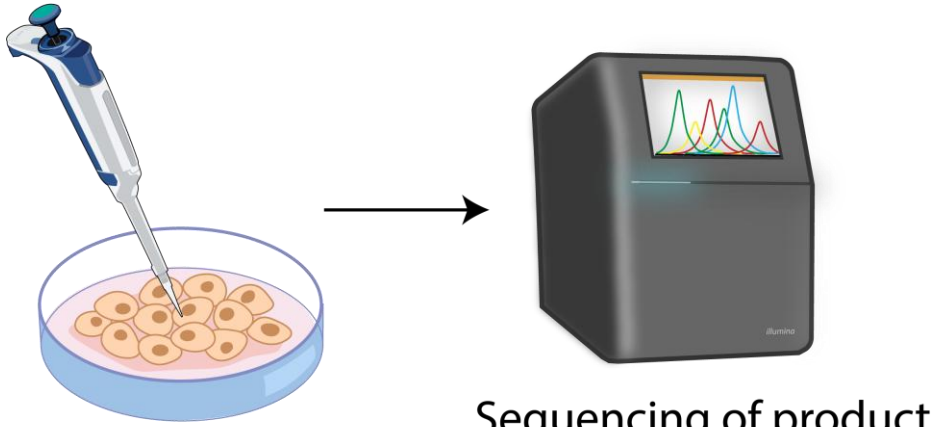
Transcript-level variants are dynamic and reflect the cellular state



Transcript-level variants are dynamic and reflect the cellular state

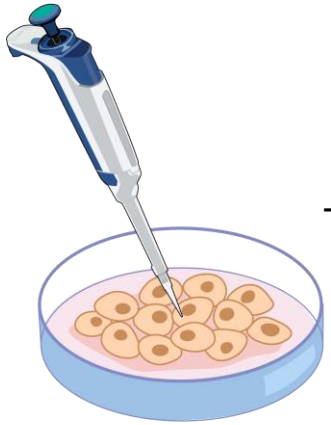


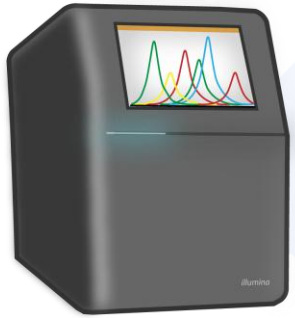
Next Generation Sequencing (NGS) forms the basis of modern variant identification pipelines



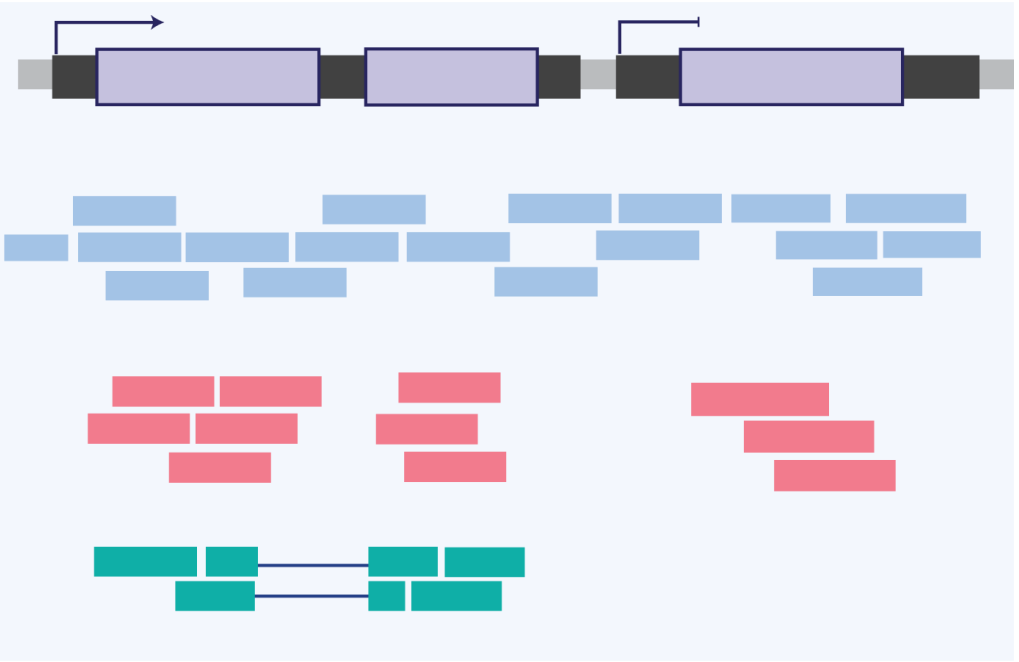
Sequencing of product

Next Generation Sequencing (NGS) forms the basis of modern variant identification pipelines

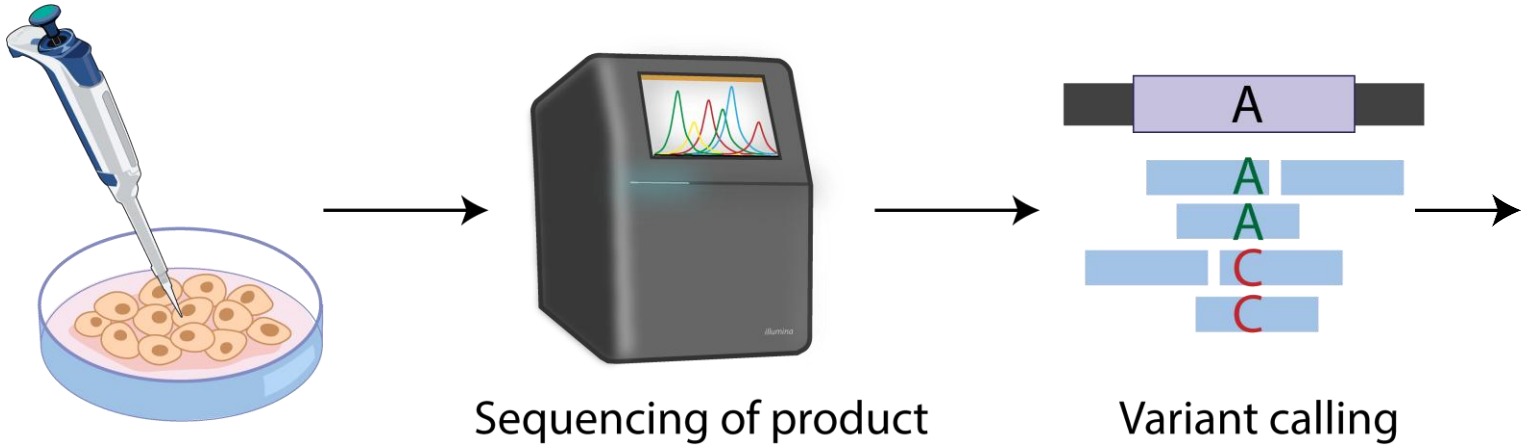




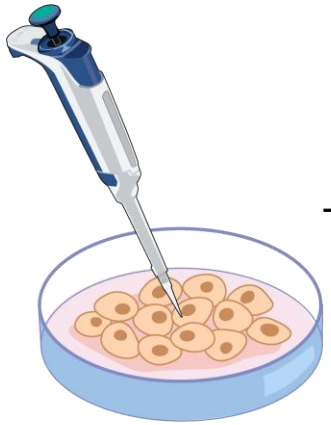
Sequencing of product

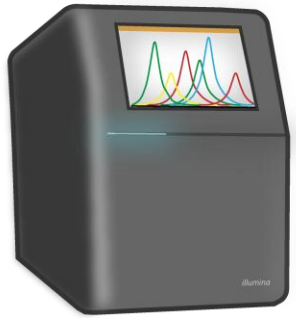


Annotation sources differ in their coverage and assessment of variant risk

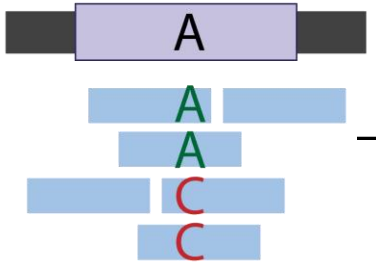


Annotation sources differ in their coverage and assessment of variant risk

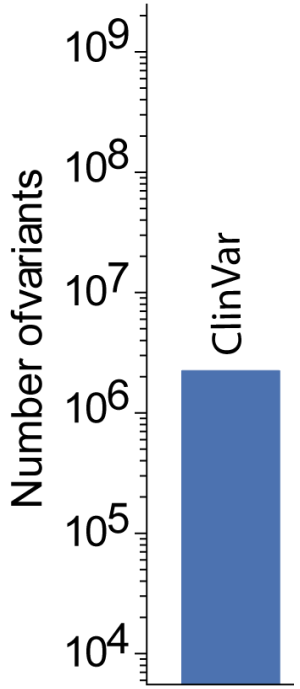




Sequencing of product



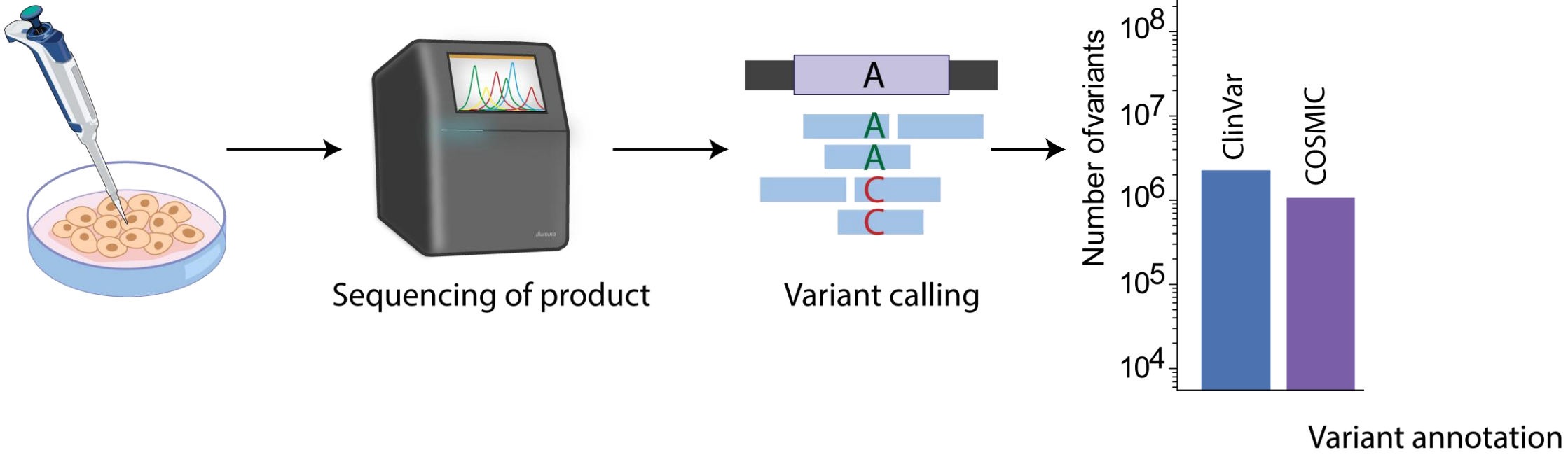
Variant calling



Variant annotation

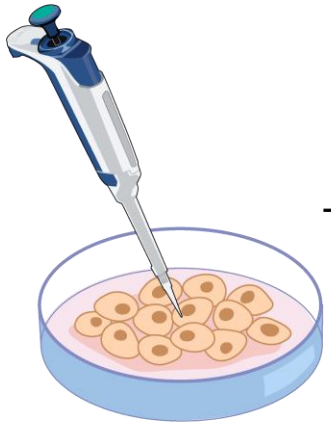
Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research*, *42*(Database issue), D980–D985. https://doi.org/10.1093/nar/gkt1113

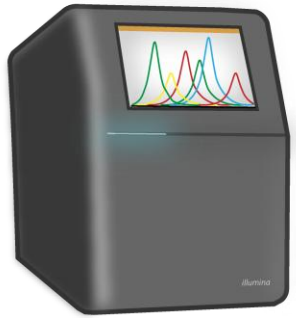
Annotation sources differ in their coverage and assessment of variant risk



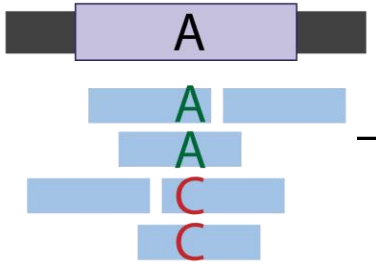
Sondka, Z., Dhir, N. B., Carvalho-Silva, D., Jupe, S., Madhumita, McLaren, K., Starkey, M., Ward, S., Wilding, J., Ahmed, M., Argasinska, J., Beare, D., Chawla, M. S., Duke, S., Fasanella, I., Neogi, A. G., Haller, S., Hetenyi, B., Hodges, L., Holmes, A., … Teague, J. (2024). COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic acids research*, *52*(D1), D1210–D1217. https://doi.org/10.1093/nar/gkad986

Annotation sources differ in their coverage and assessment of variant risk

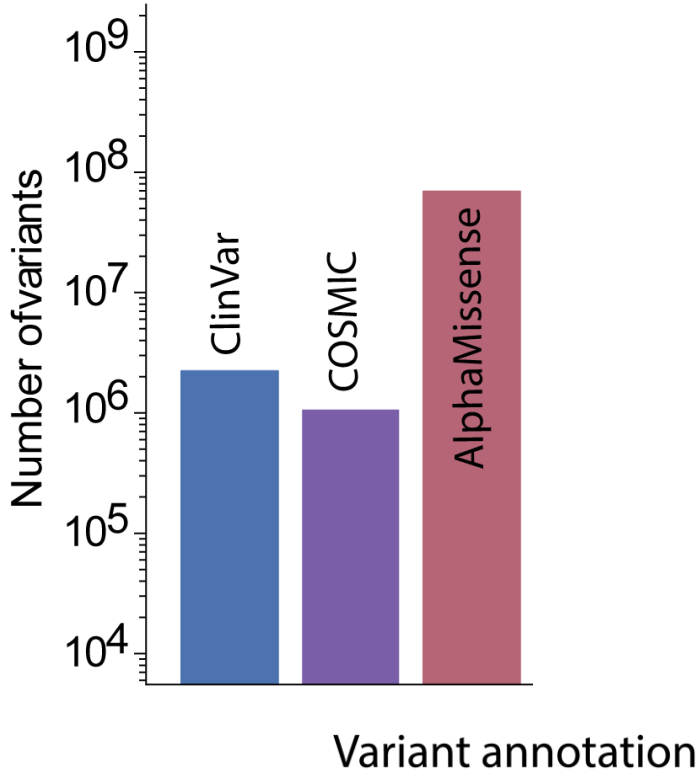




Sequencing of product

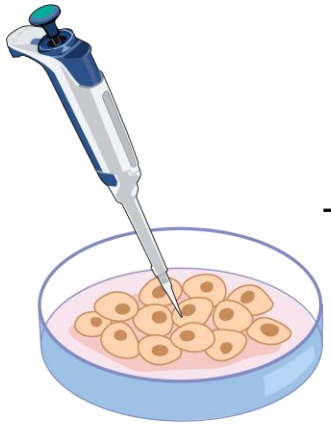


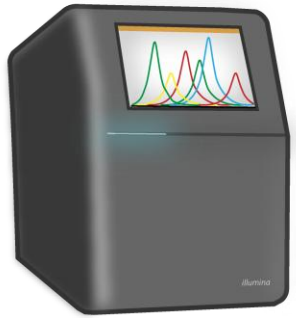
Variant calling



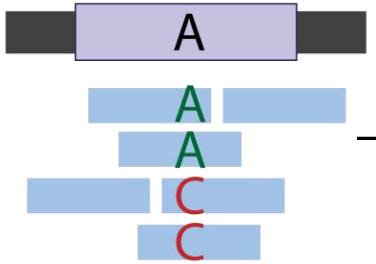
Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., Schneider, R. G., Senior, A. W., Jumper, J., Hassabis, D., Kohli, P., & Avsec, Ž. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science (New York, N.Y.), 381(6664), eadg7492. https://doi.org/10.1126/science.adg7492

Annotation sources differ in their coverage and assessment of variant risk

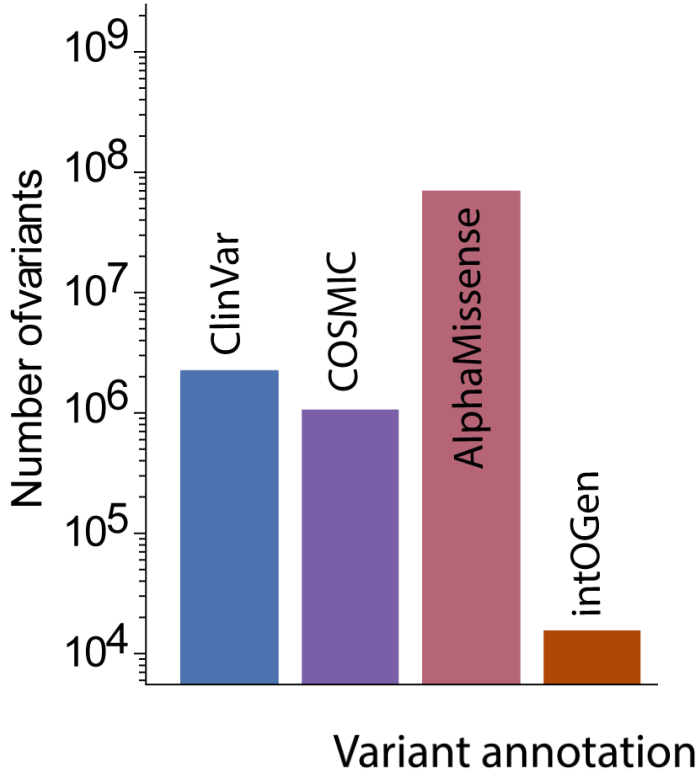




Sequencing of product

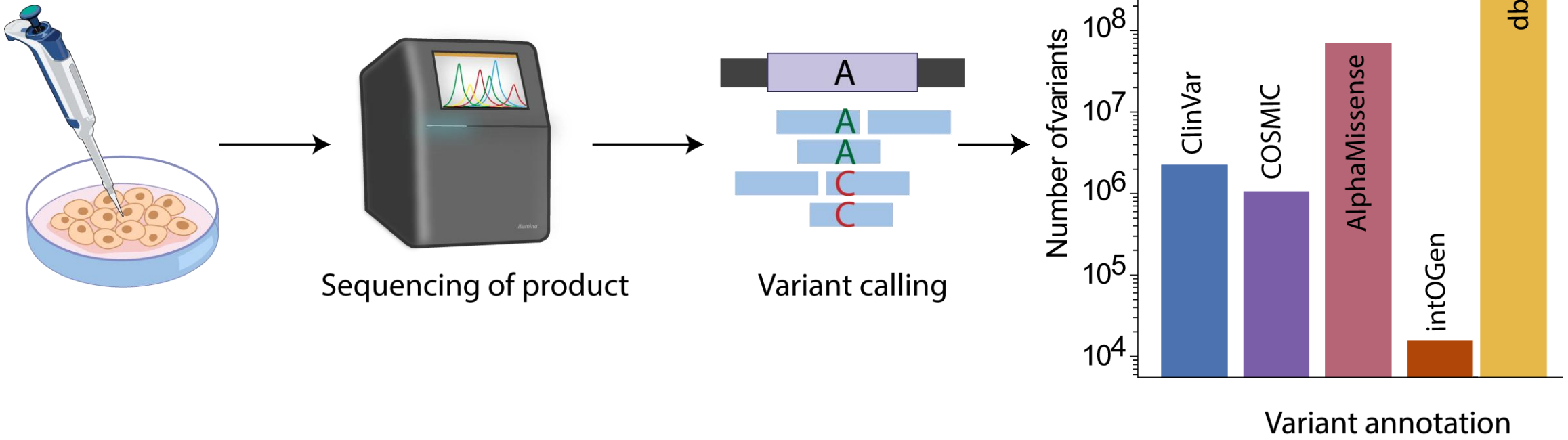


Variant calling



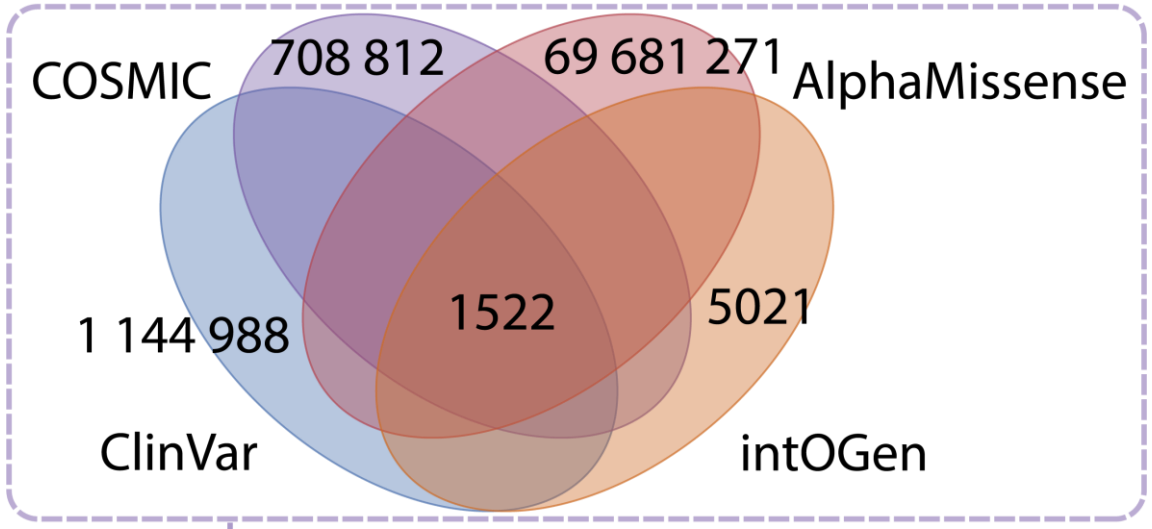
Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M. P., Jene-Sanz, A., Santos, A., & Lopez-Bigas, N. (2013). IntOGen-mutations identifies cancer drivers across tumor types. Nature methods, 10(11), 1081–1082. https://doi.org/10.1038/nmeth.2642

Annotation sources differ in their coverage and assessment of variant risk



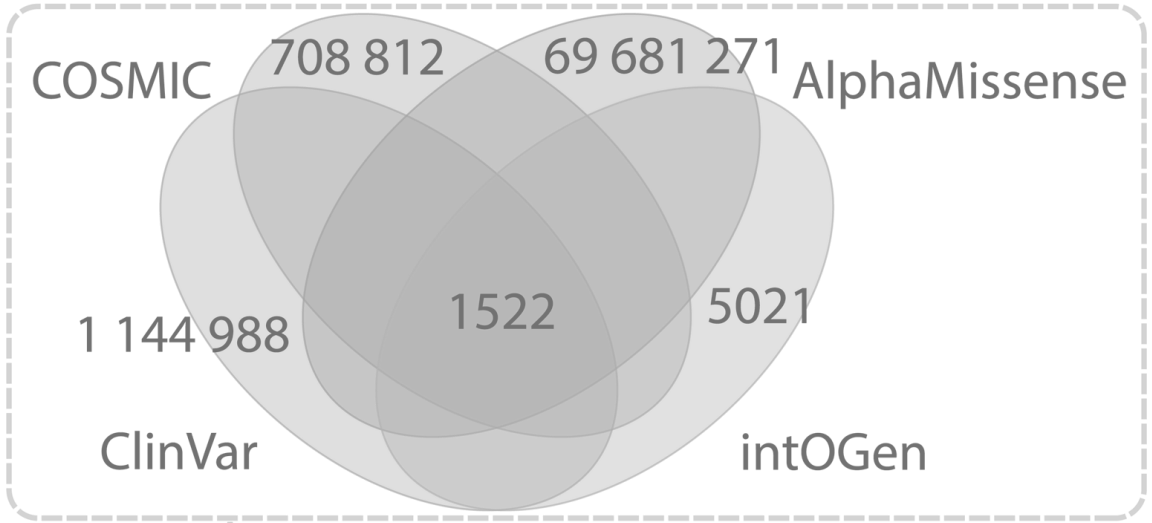
Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic acids research, 29(1), 308–311. https://doi.org/10.1093/nar/29.1.308

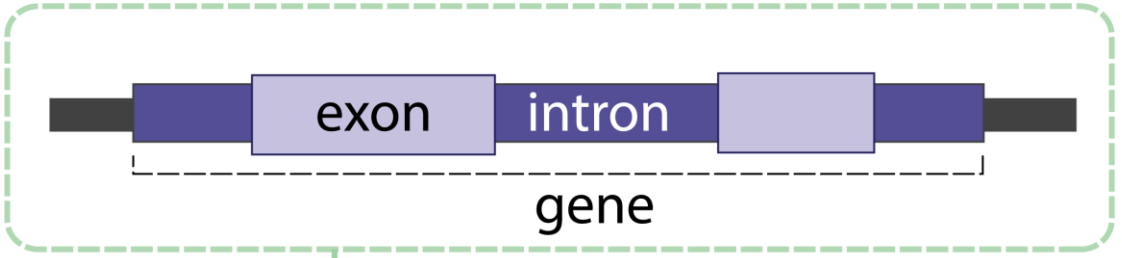
VaCS is a holistic weight-of-evidence score for genomic risk assessment



$$VaCS = \frac{(\,S_pathogenicity\,(var) + S_gene\,(var) \times S_location\,(var)\,) \times S_effect\,(var)}{max_risk_score} \in [0, 1]$$

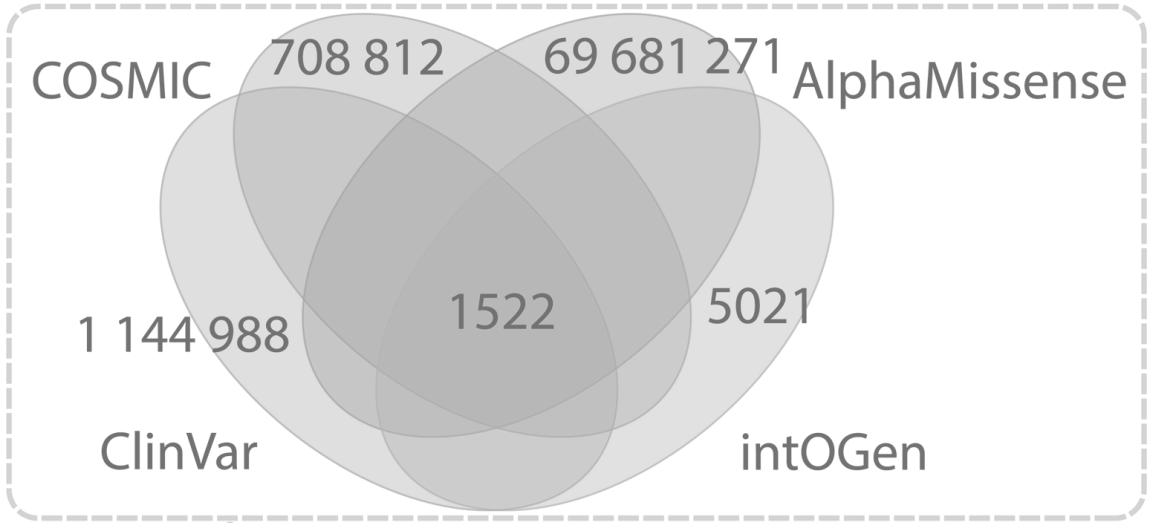
VaCS is a holistic weight-of-evidence score for genomic risk assessment

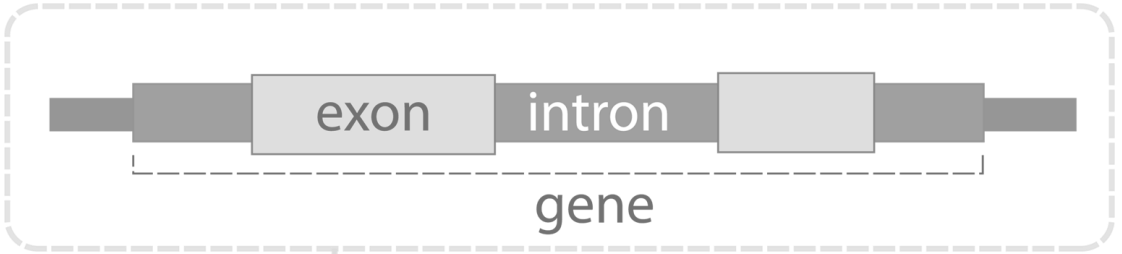




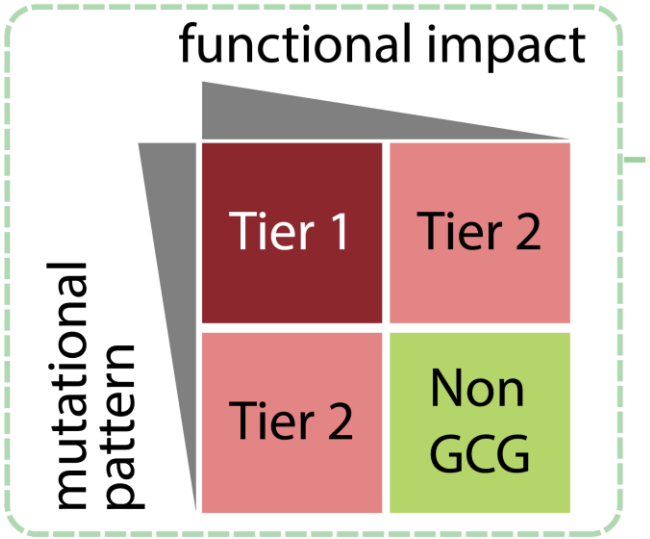
$$VaCS = \frac{(S_pathogenicity\,(var) + S_gene\,(var) \times S_location\,(var)) \times S_effect\,(var)}{max_risk_score} \in [0, 1]$$

VaCS is a holistic weight-of-evidence score for genomic risk assessment



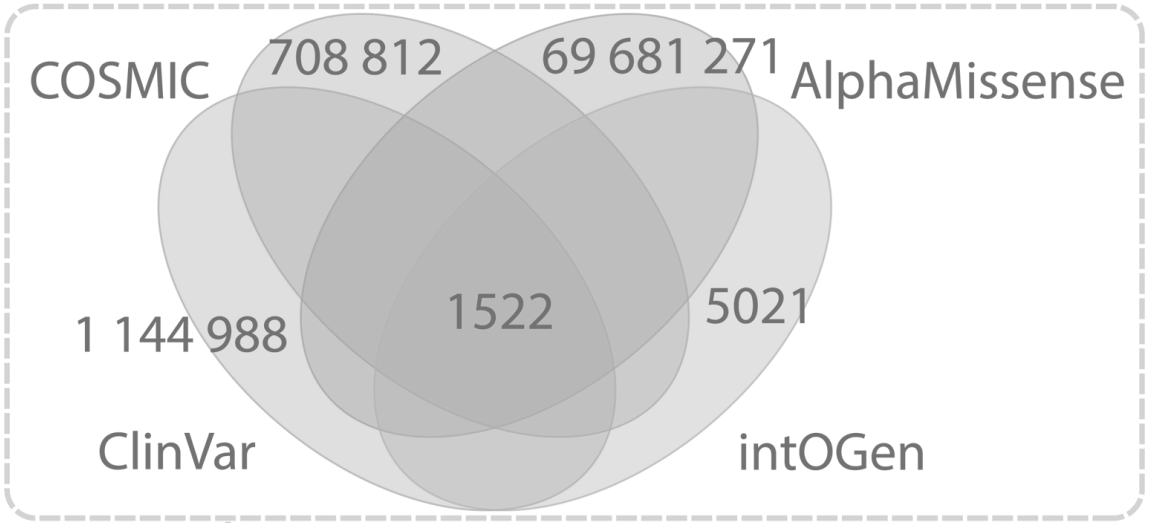


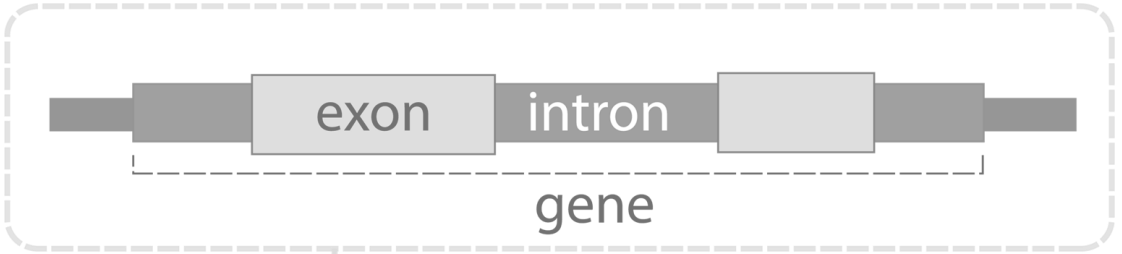
$$VaCS = \frac{(S_pathogenicity\,(var) + S_gene\,(var) \times S_location\,(var)) \times S_effect\,(var)}{max_risk_score} \in [0, 1]$$



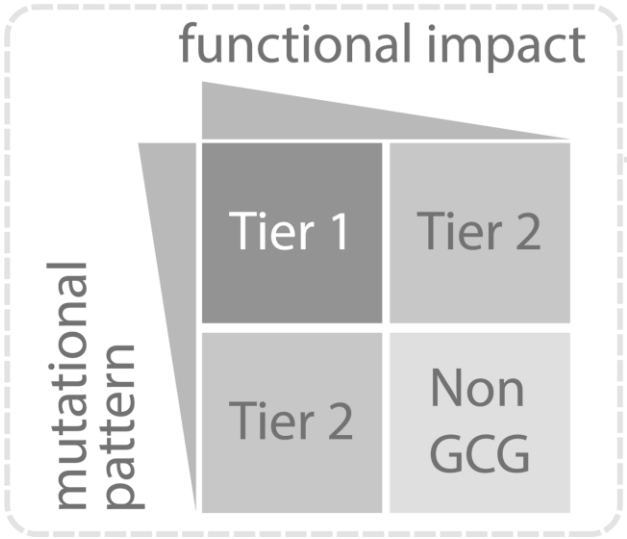
Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., & Forbes, S. A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nature reviews. Cancer, 18(11), 696–705. https://doi.org/10.1038/s41568-018-0060-1

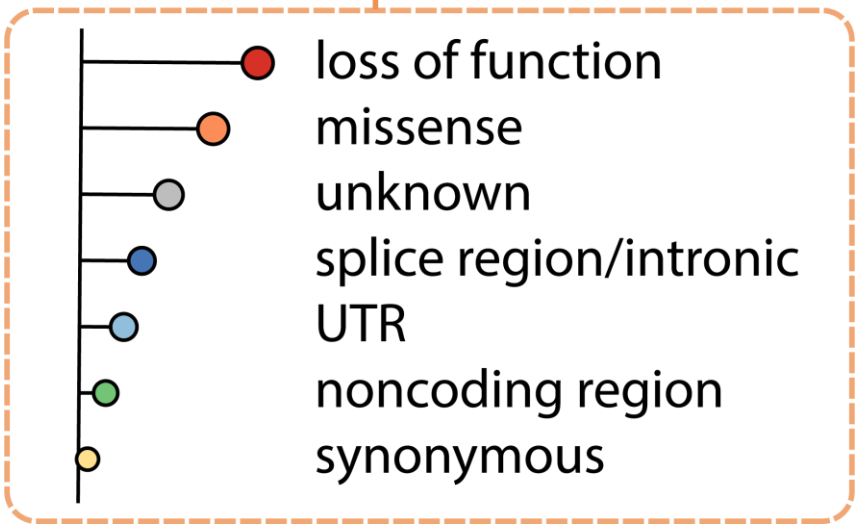
VaCS is a holistic weight-of-evidence score for genomic risk assessment





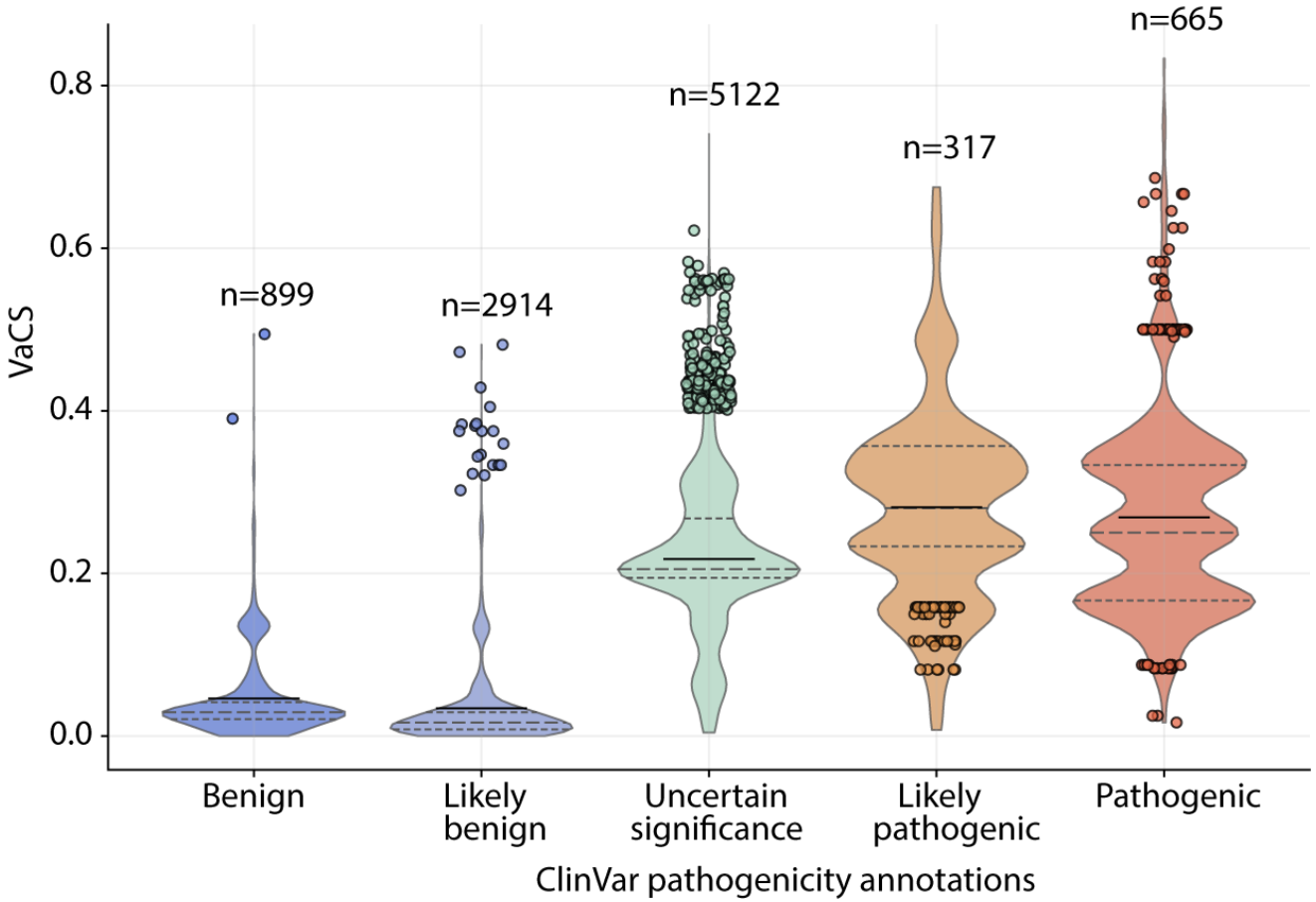
$$VaCS = \frac{(S_pathogenicity\,(var) + S_gene\,(var) \times S_location\,(var)) \times S_effect\,(var)}{max_risk_score} \in [0, 1]$$



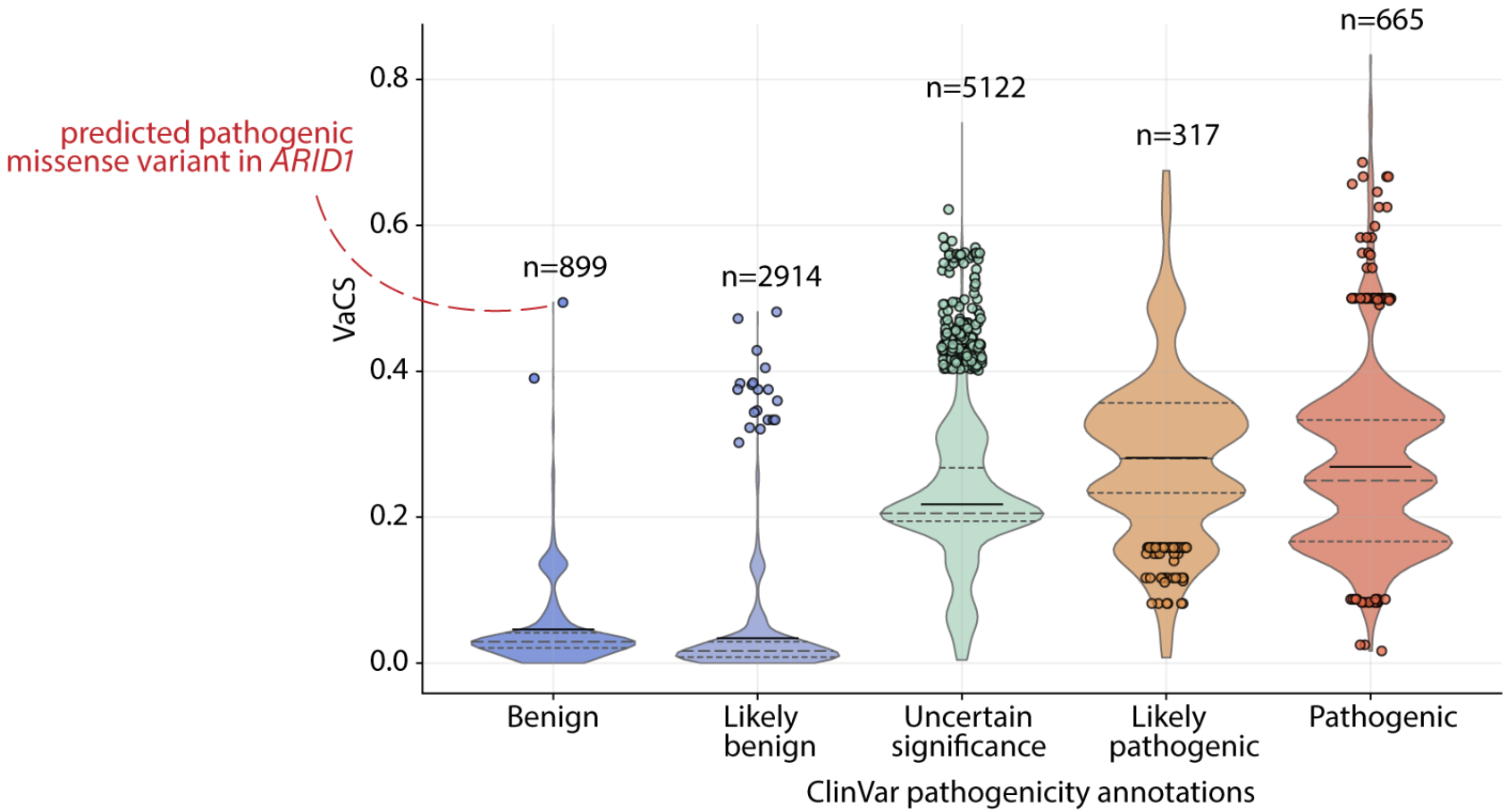


McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome biology, 17(1), 122. https://doi.org/10.1186/s13059-016-0974-4

VaCS resolves discordant annotations through evidence integration

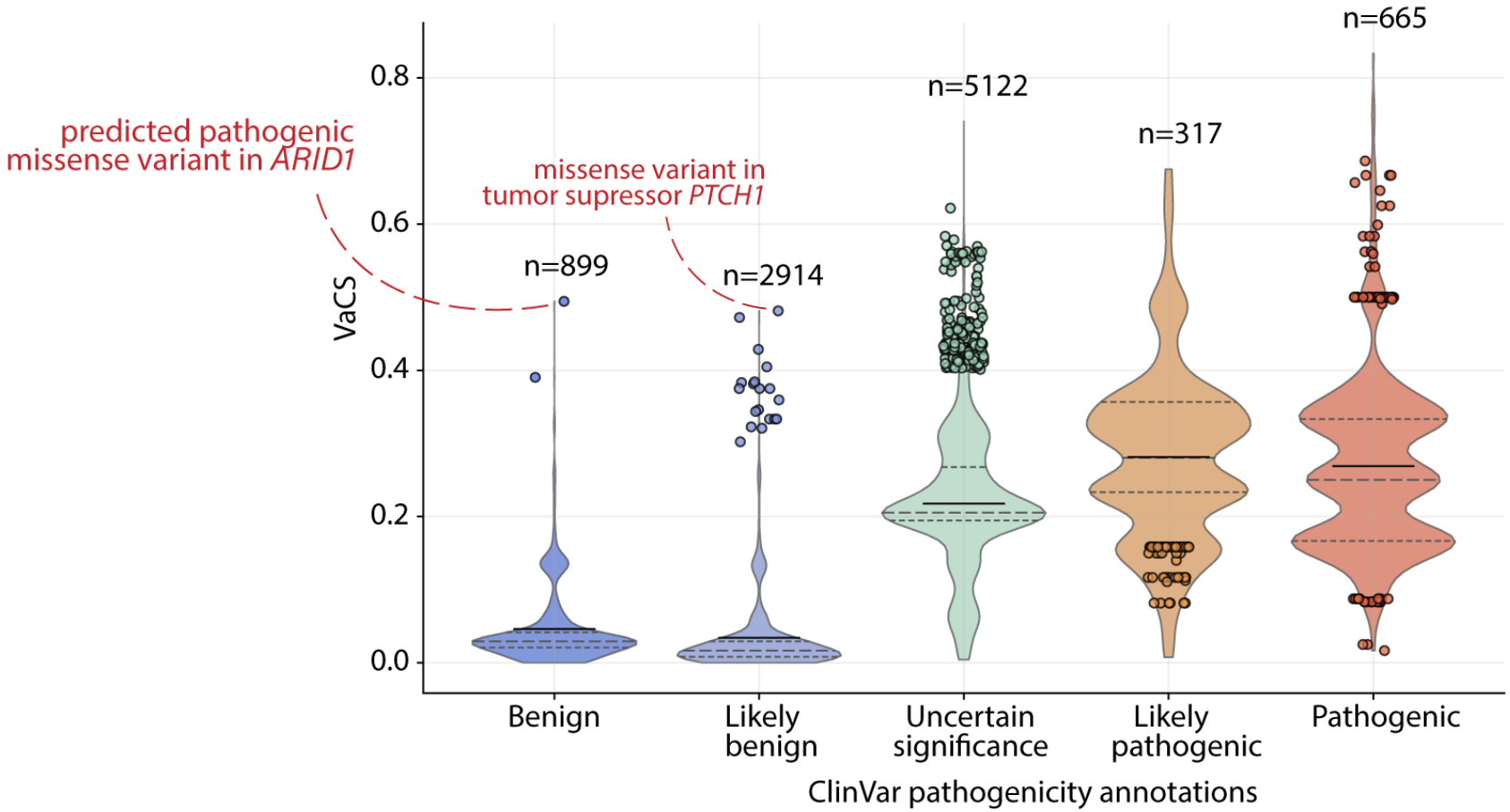


VaCS resolves discordant annotations through evidence integration

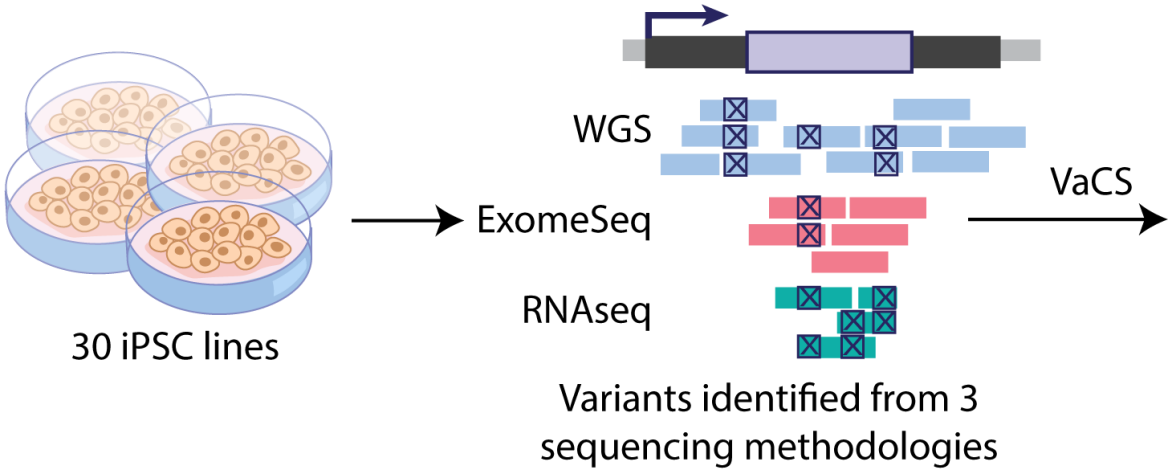


Wagenknecht, J. B., Haque, N., Gresser, J., Dong, X., & Zimmermann, M. T. (2026). ARID1B damaging variants from more than one million genomes, cause human diseases by impairing protein-protein interactions, stability, and regulation. bioRxiv : the preprint server for biology, 2026.06.19.733400. https://doi.org/10.64898/2026.06.19.733400

VaCS resolves discordant annotations through evidence integration

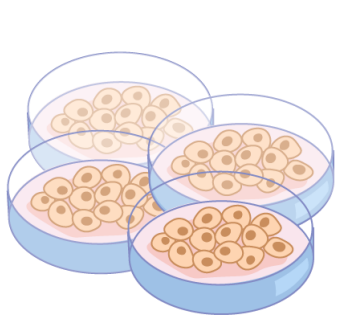


RNA-seq captures most high-risk variants identified across sequencing methods

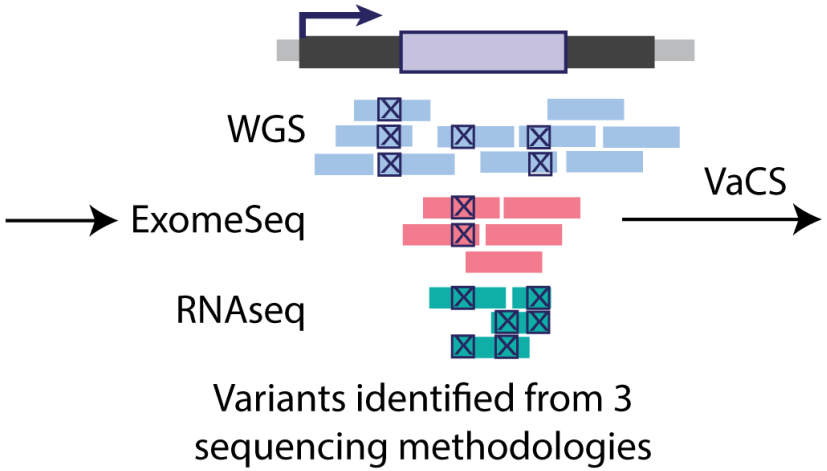


Streeter, I., Harrison, P. W., Faulconbridge, A., The HipSci Consortium, Flicek, P., Parkinson, H., & Clarke, L. (2017). The human-induced pluripotent stem cell initiative-data resources for cellular genetics. Nucleic acids research, 45(D1), D691–D697. https://doi.org/10.1093/nar/gkw928

RNA-seq captures most high-risk variants identified across sequencing methods

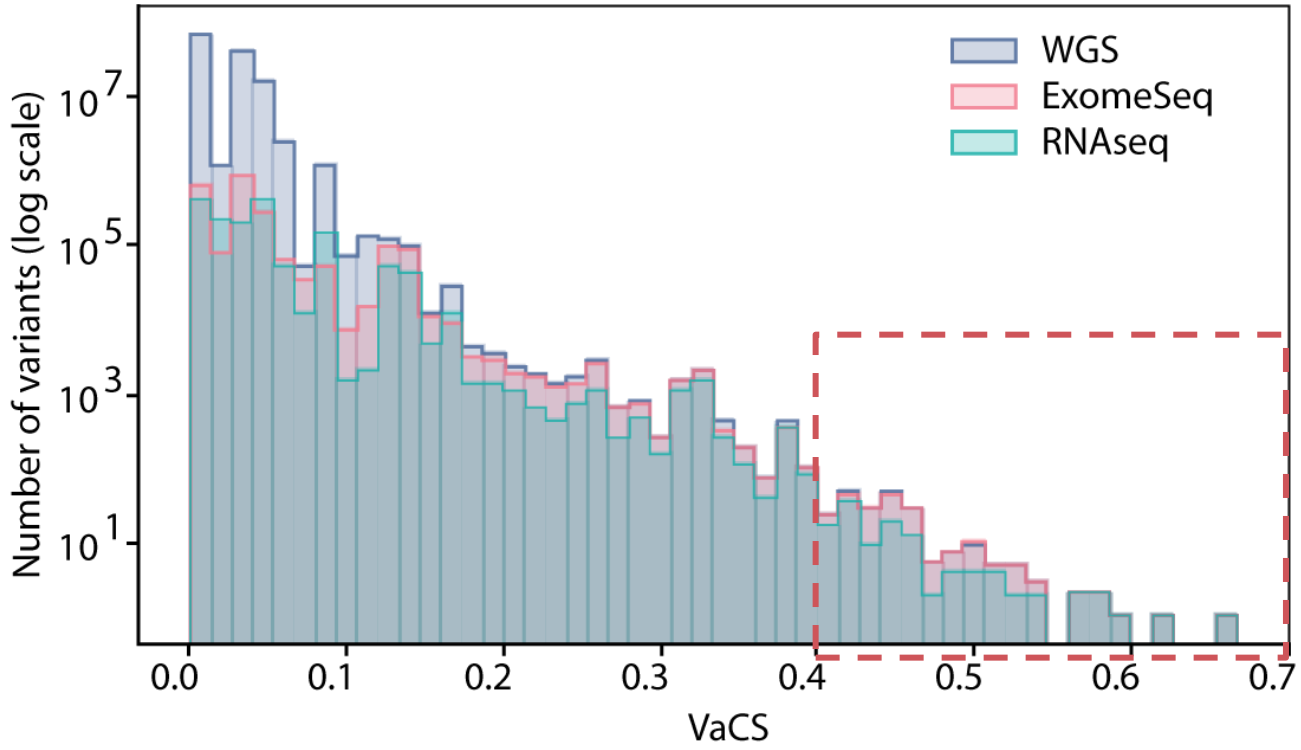


30 iPSC lines

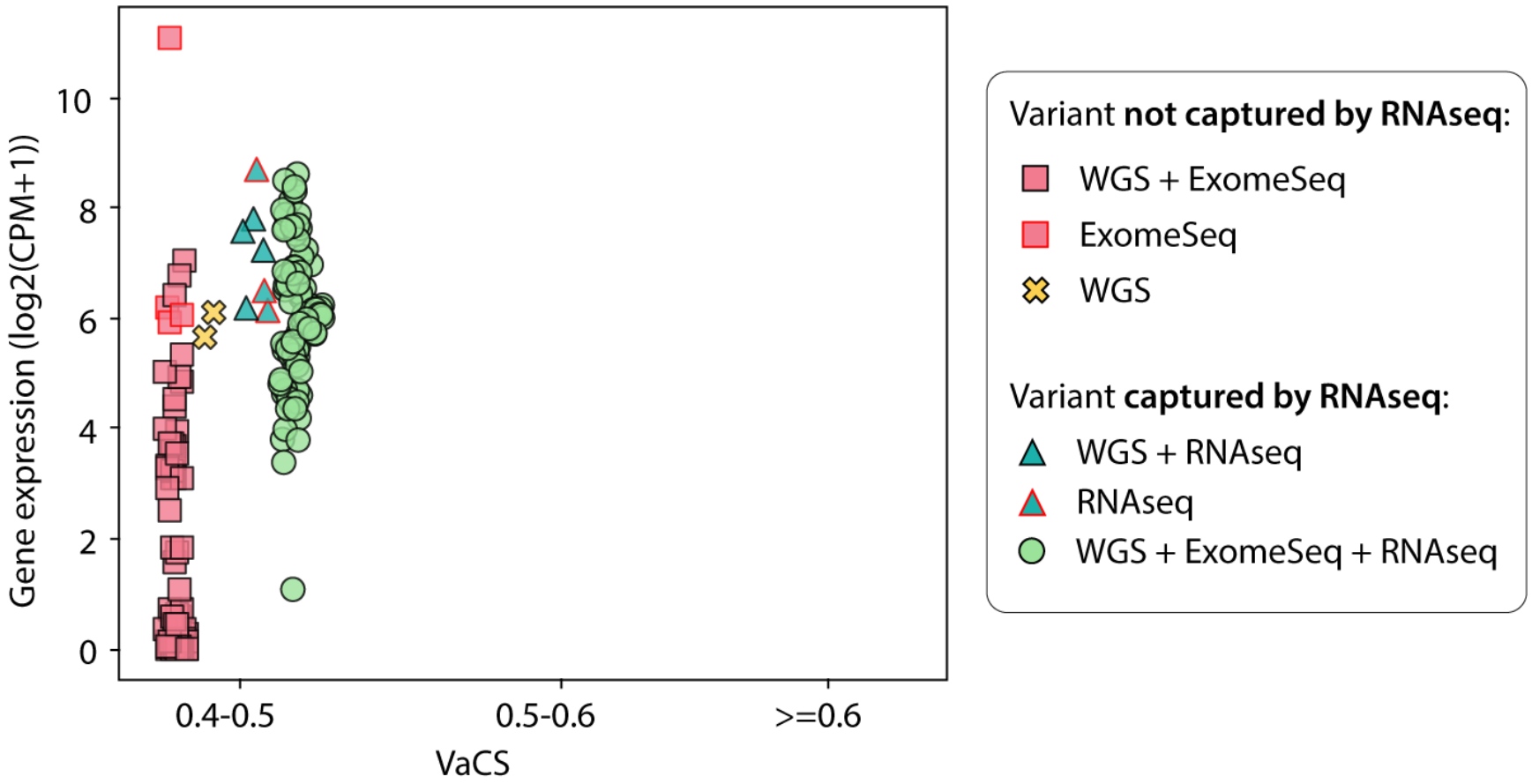


Variants identified from 3 sequencing methodologies

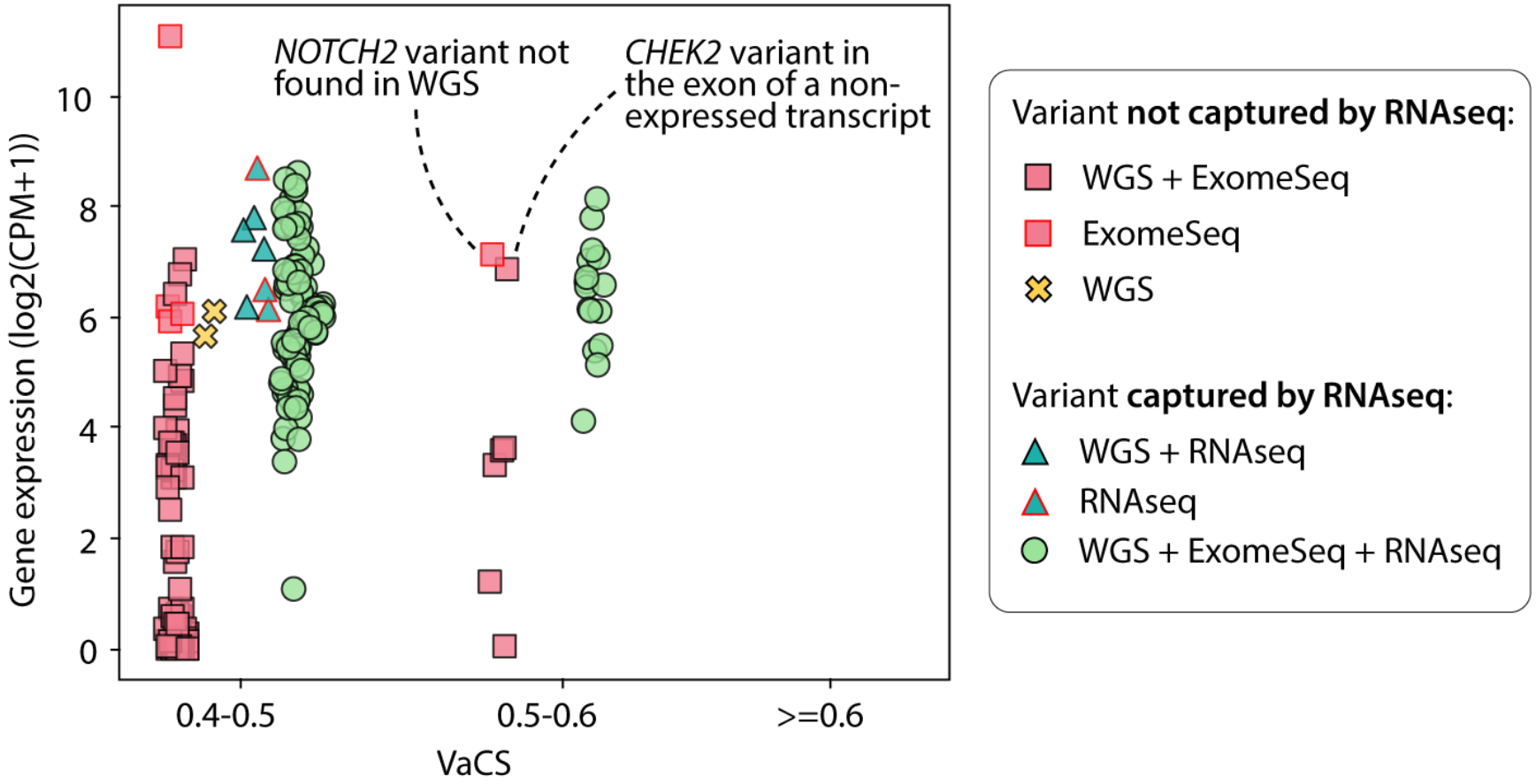
VaCS



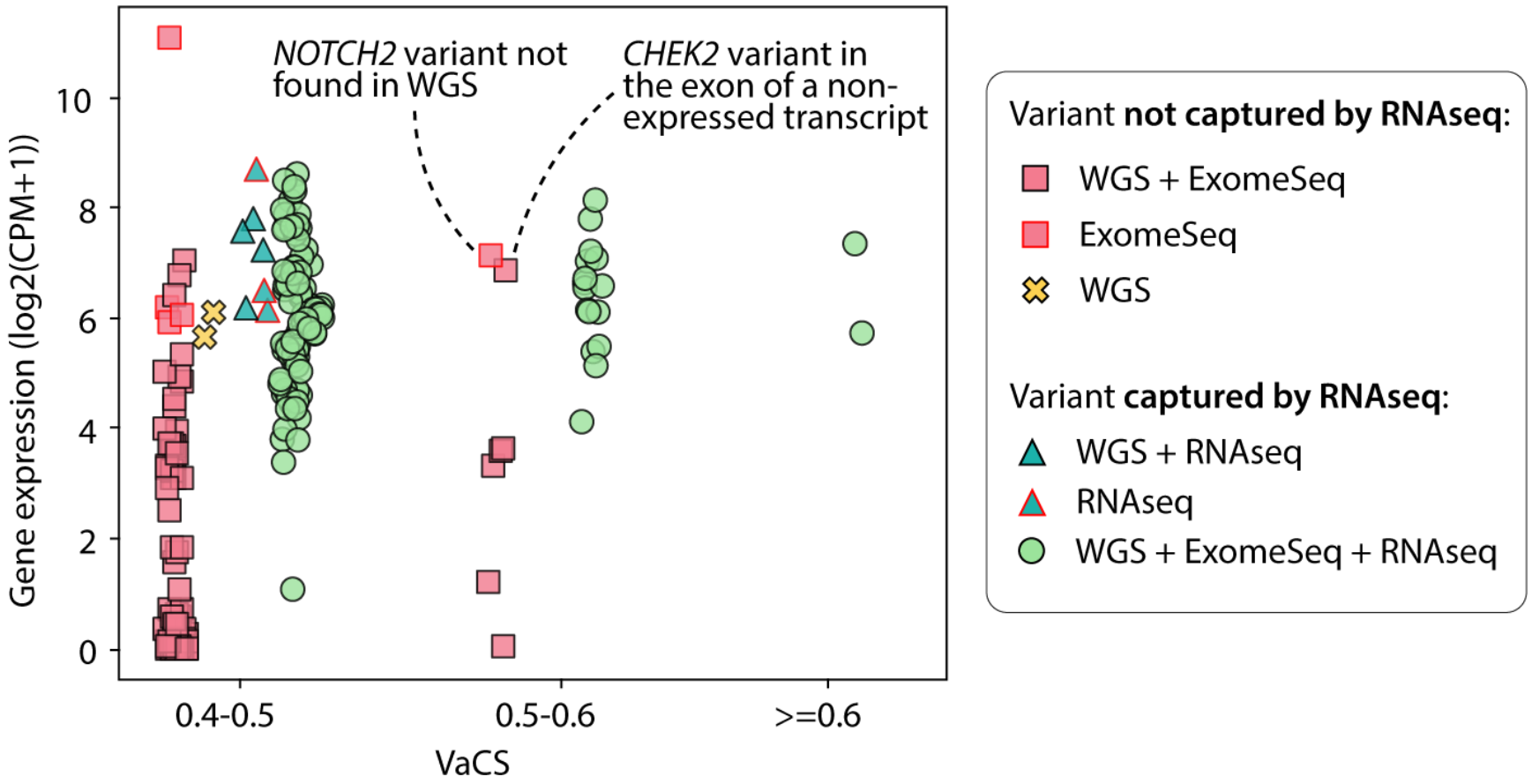
RNA-seq captures high-risk variants in expressed genes



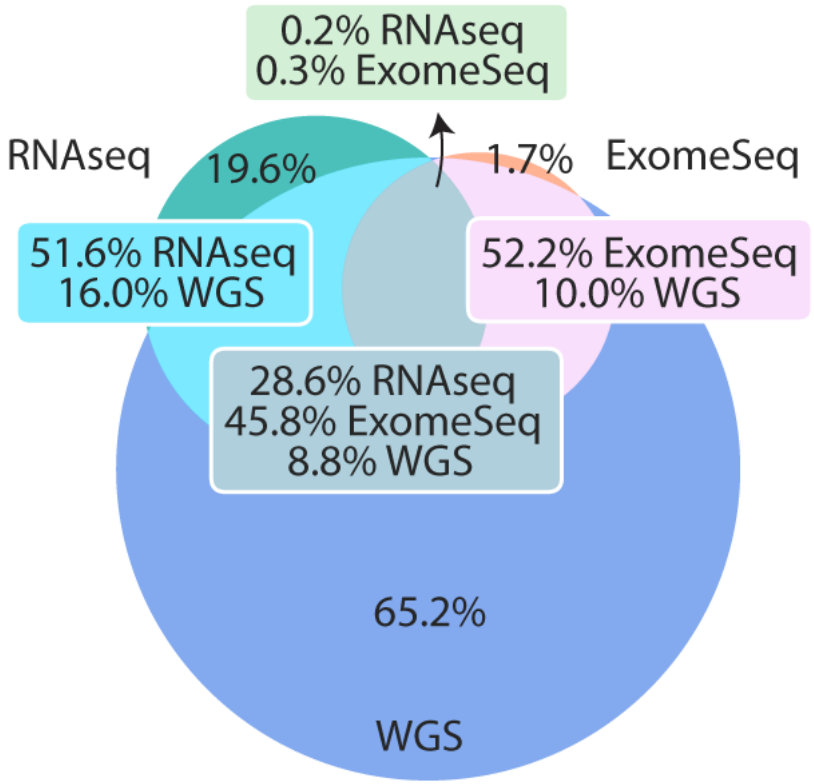
RNA-seq captures high-risk variants in expressed genes



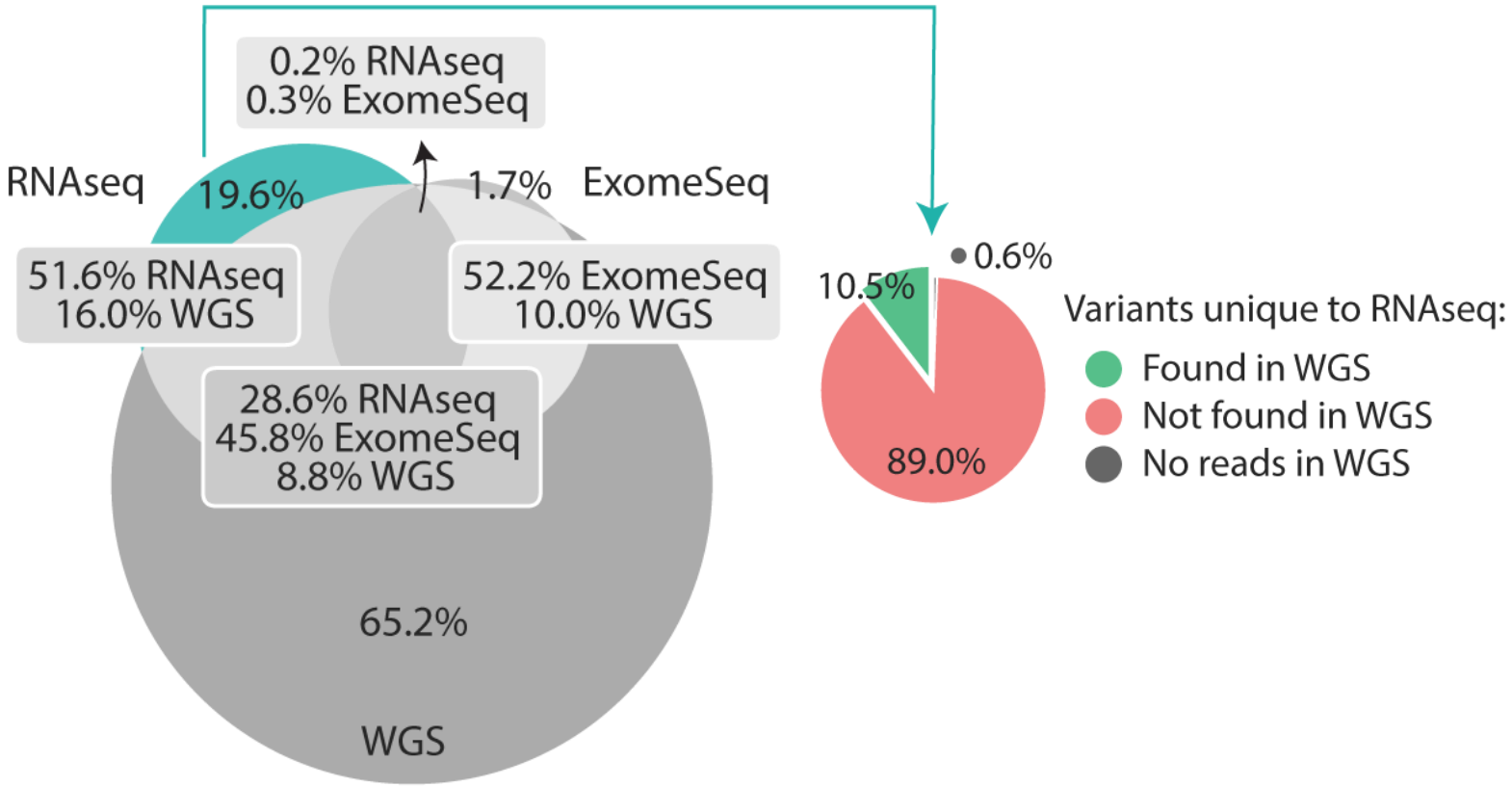
RNA-seq captures high-risk variants in expressed genes



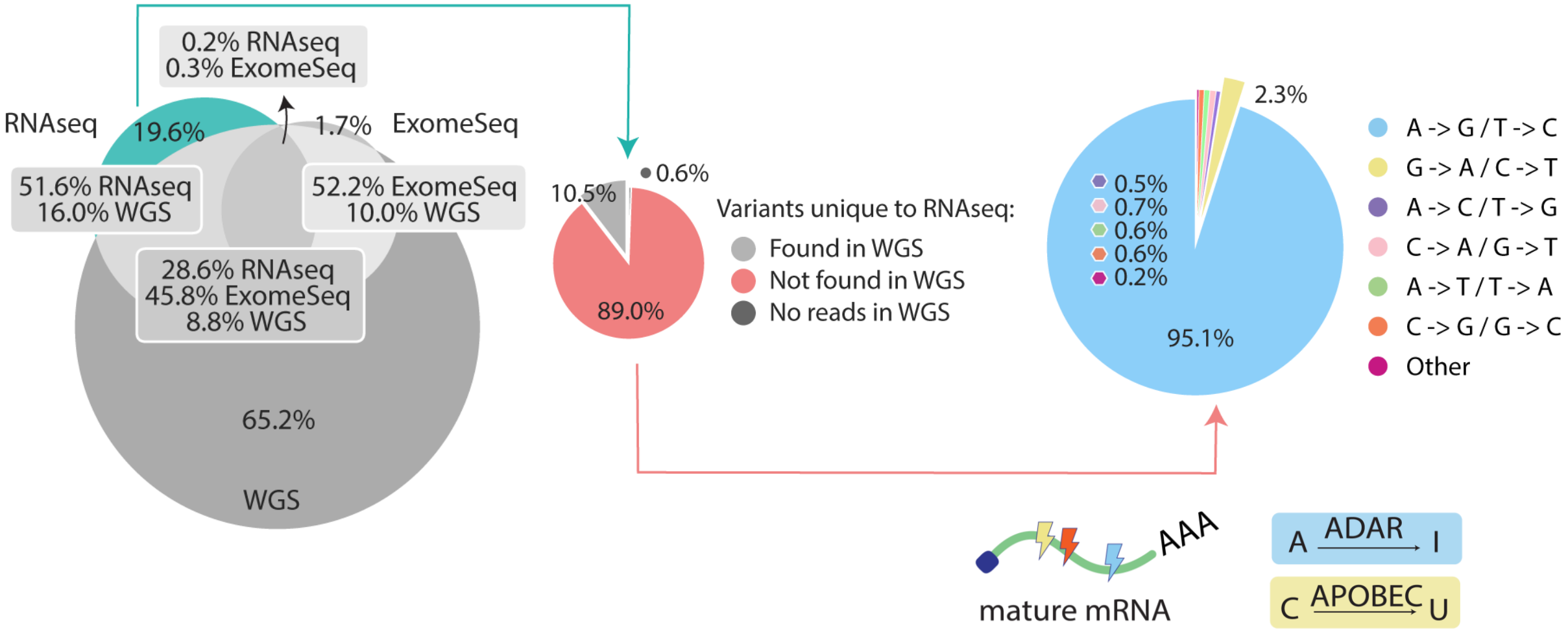
RNAseq-unique variants are potential RNA editing events



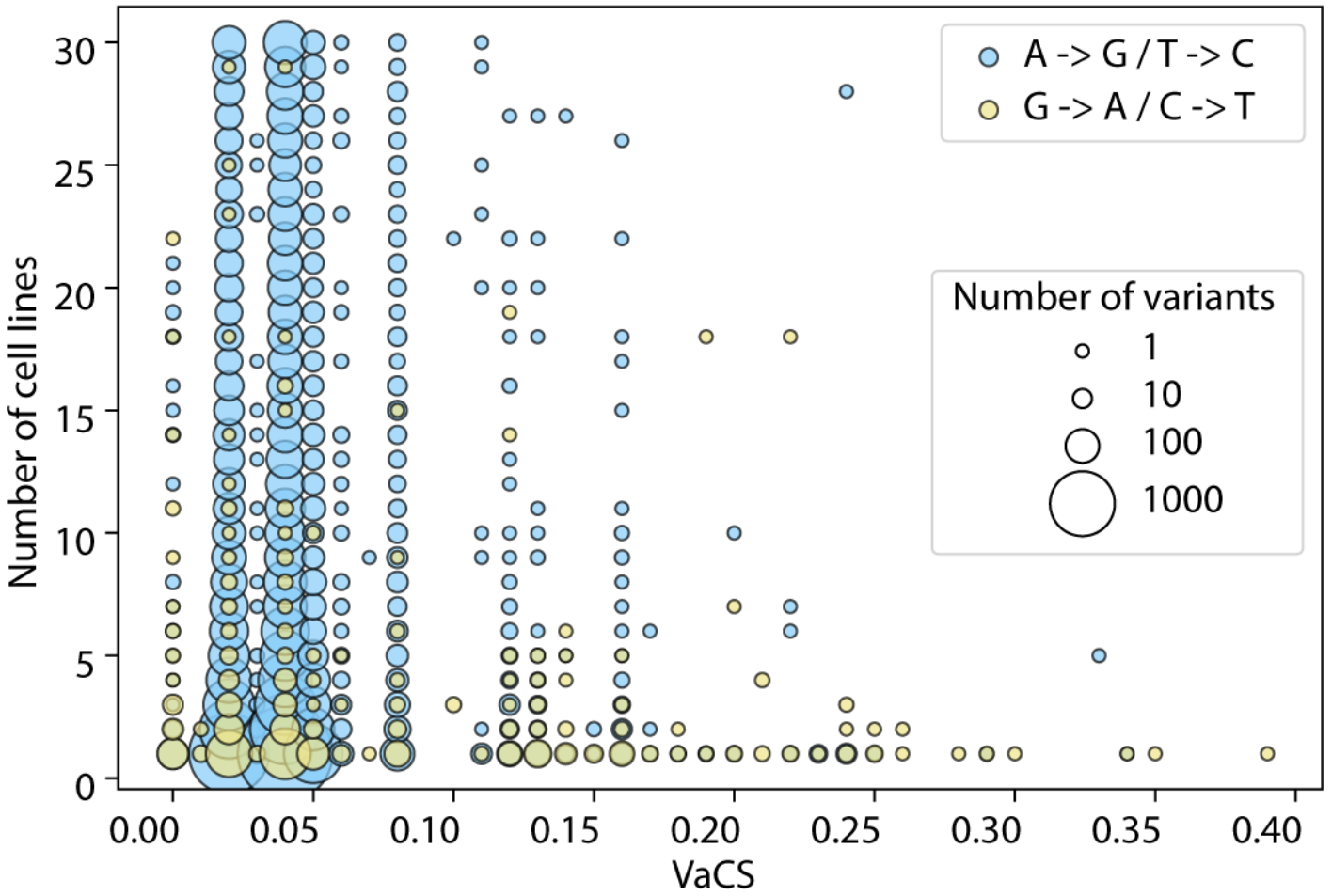
RNAseq-unique variants are potential RNA editing events



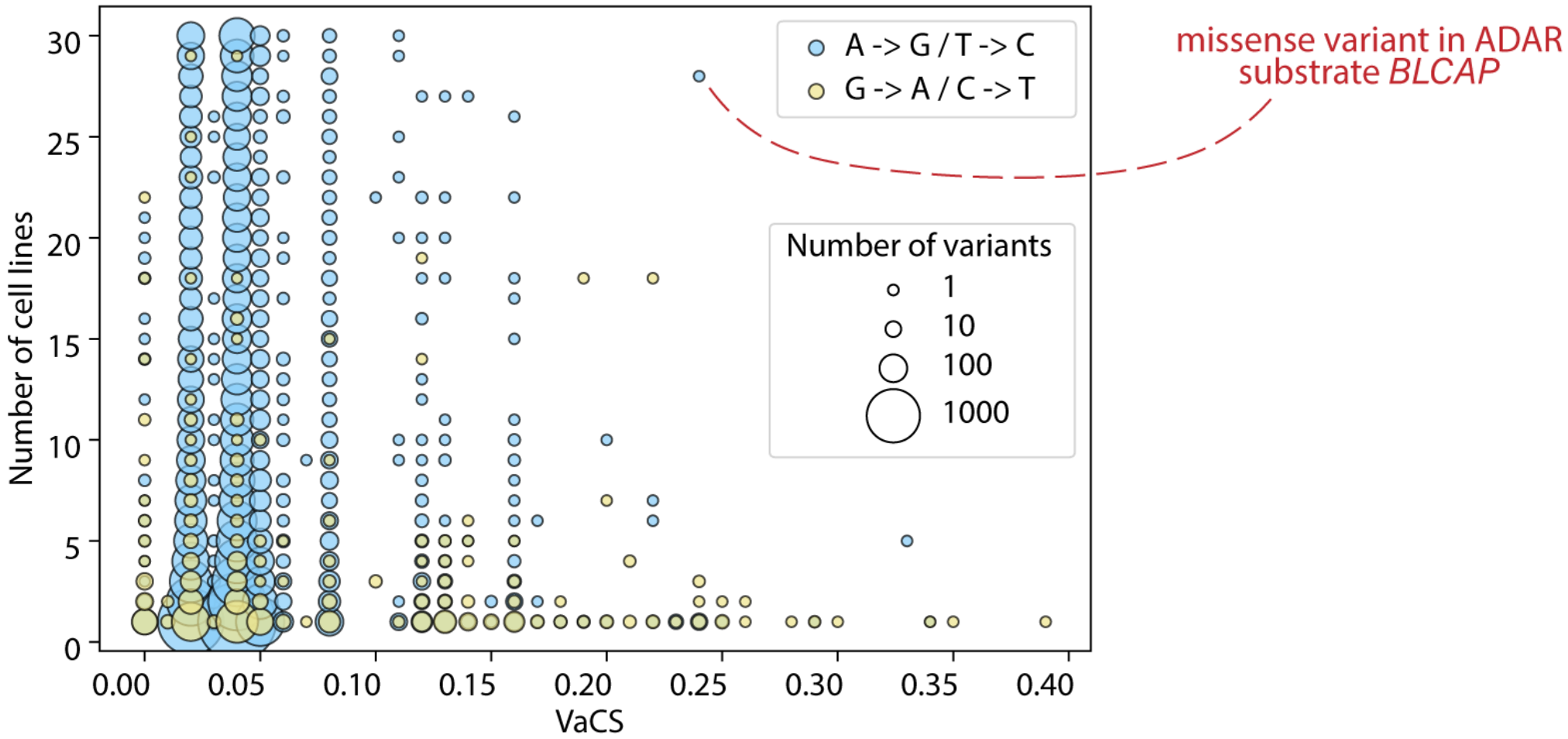
RNAseq-unique variants are potential RNA editing events



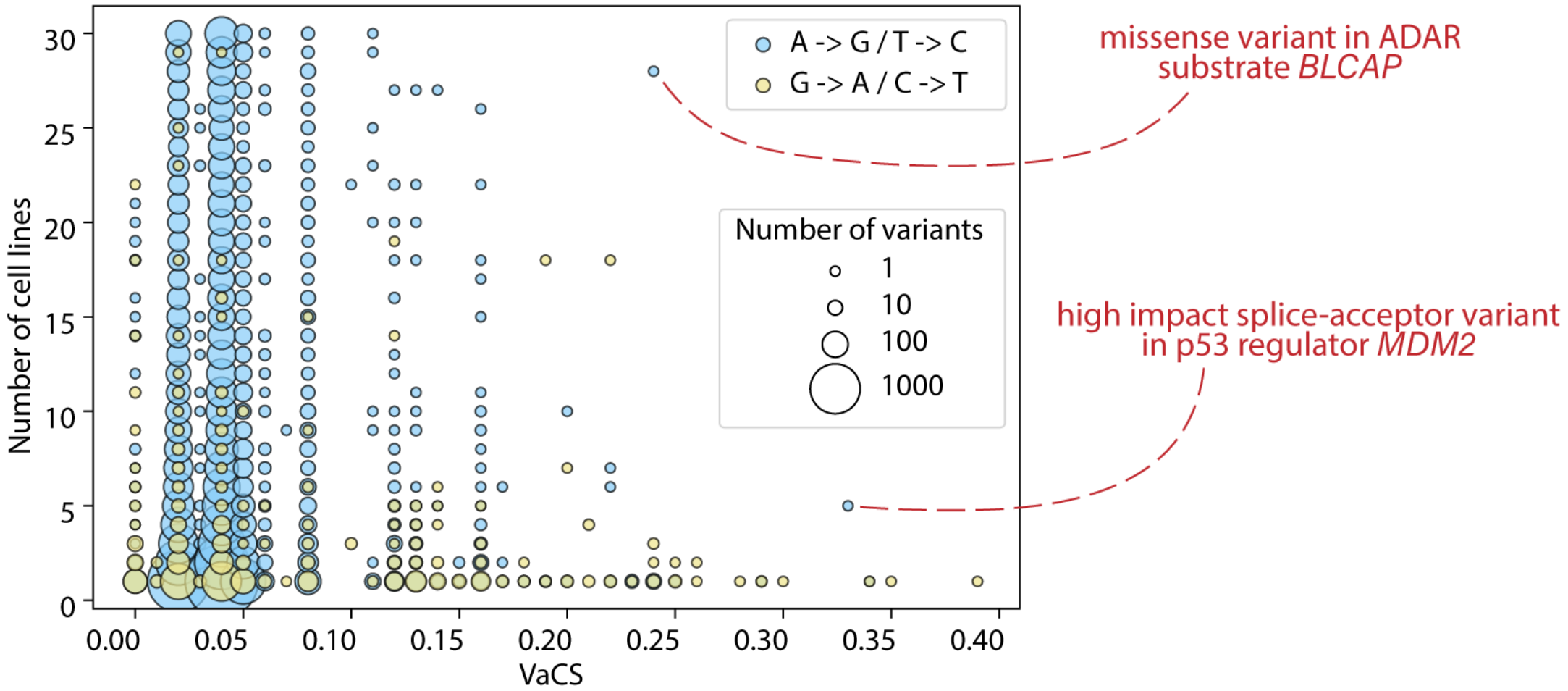
VaCS distinguishes potentially concerning RNA editing events



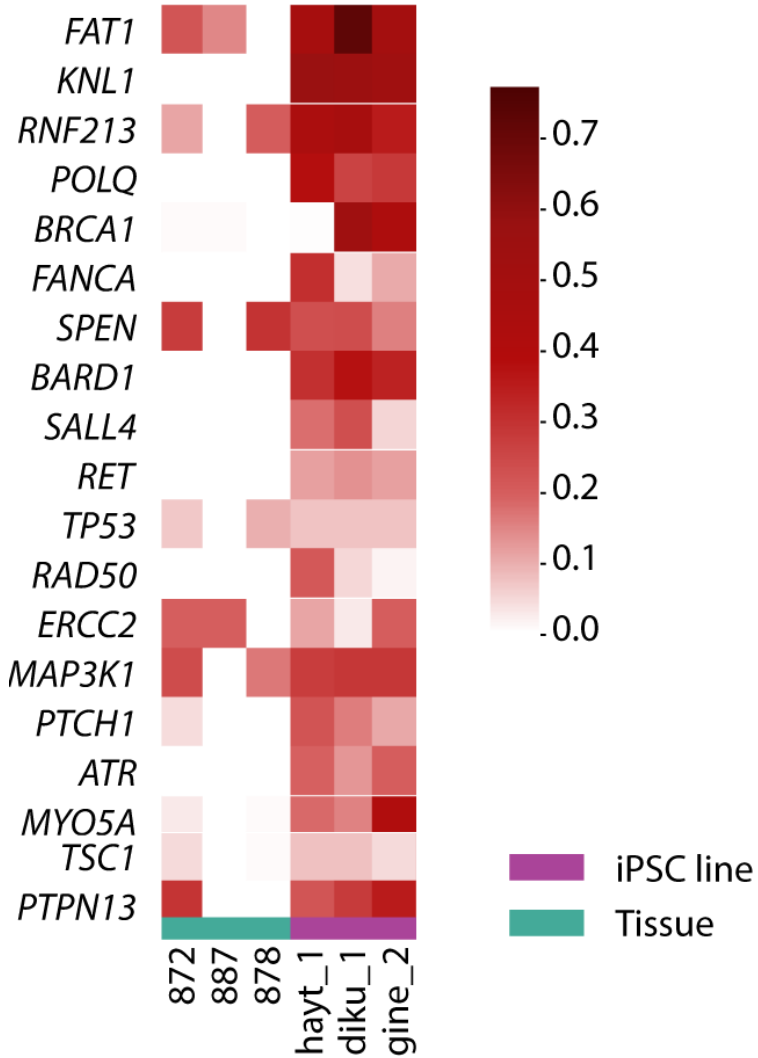
VaCS distinguishes potentially concerning RNA editing events



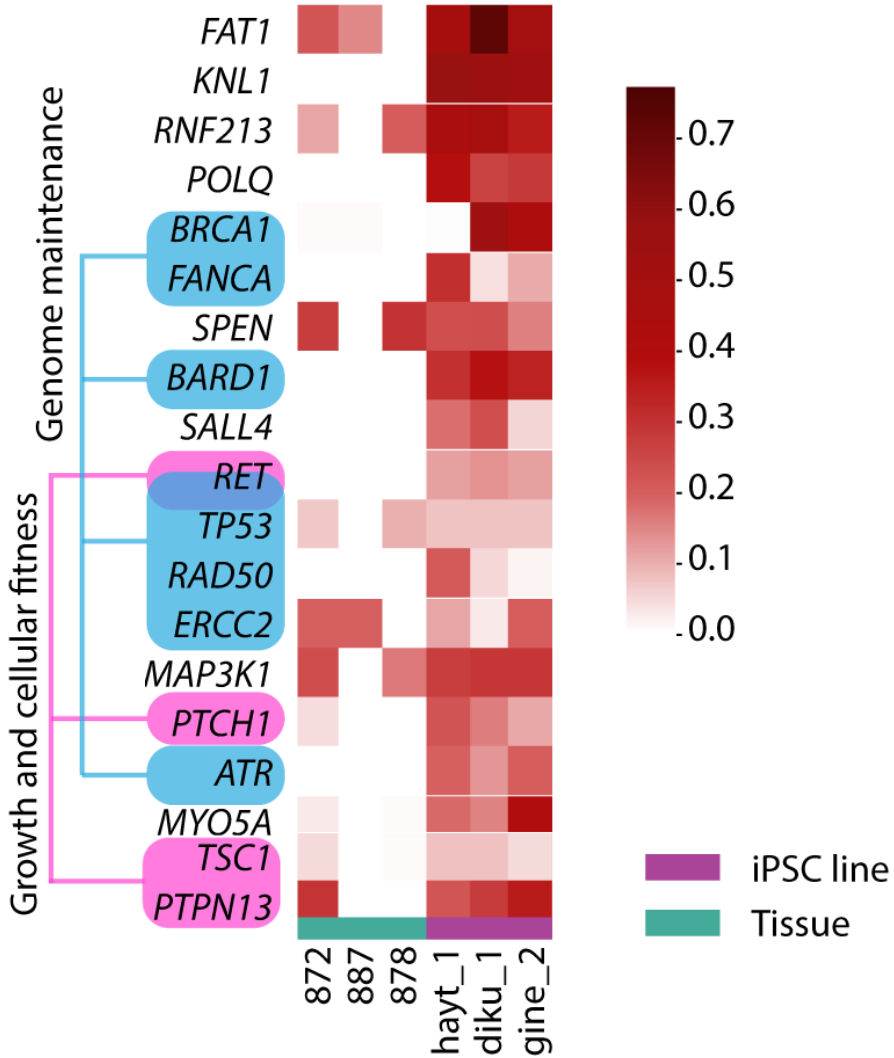
VaCS distinguishes potentially concerning RNA editing events



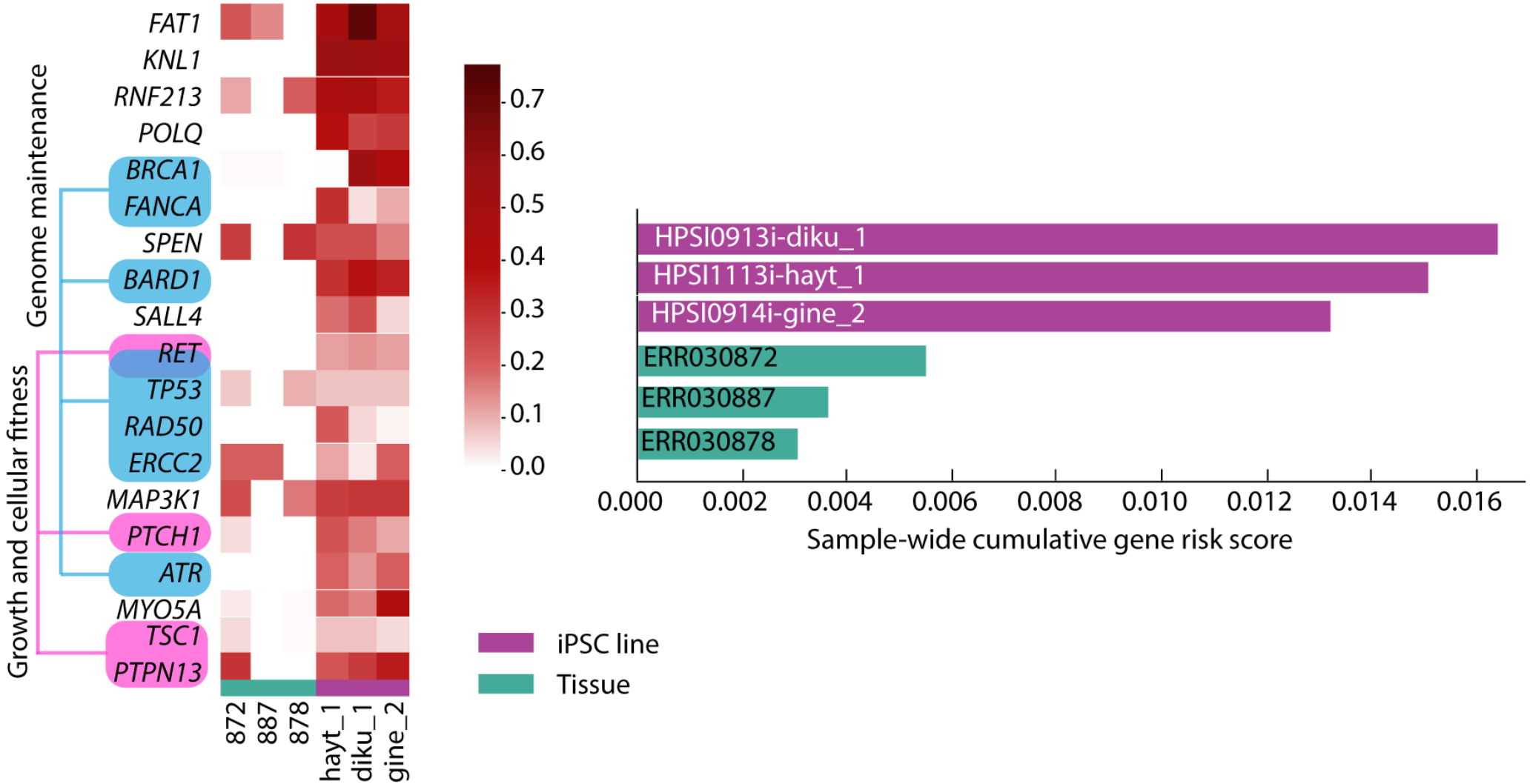
Summing aggregated gene-level VaCS enables sample-level comparison

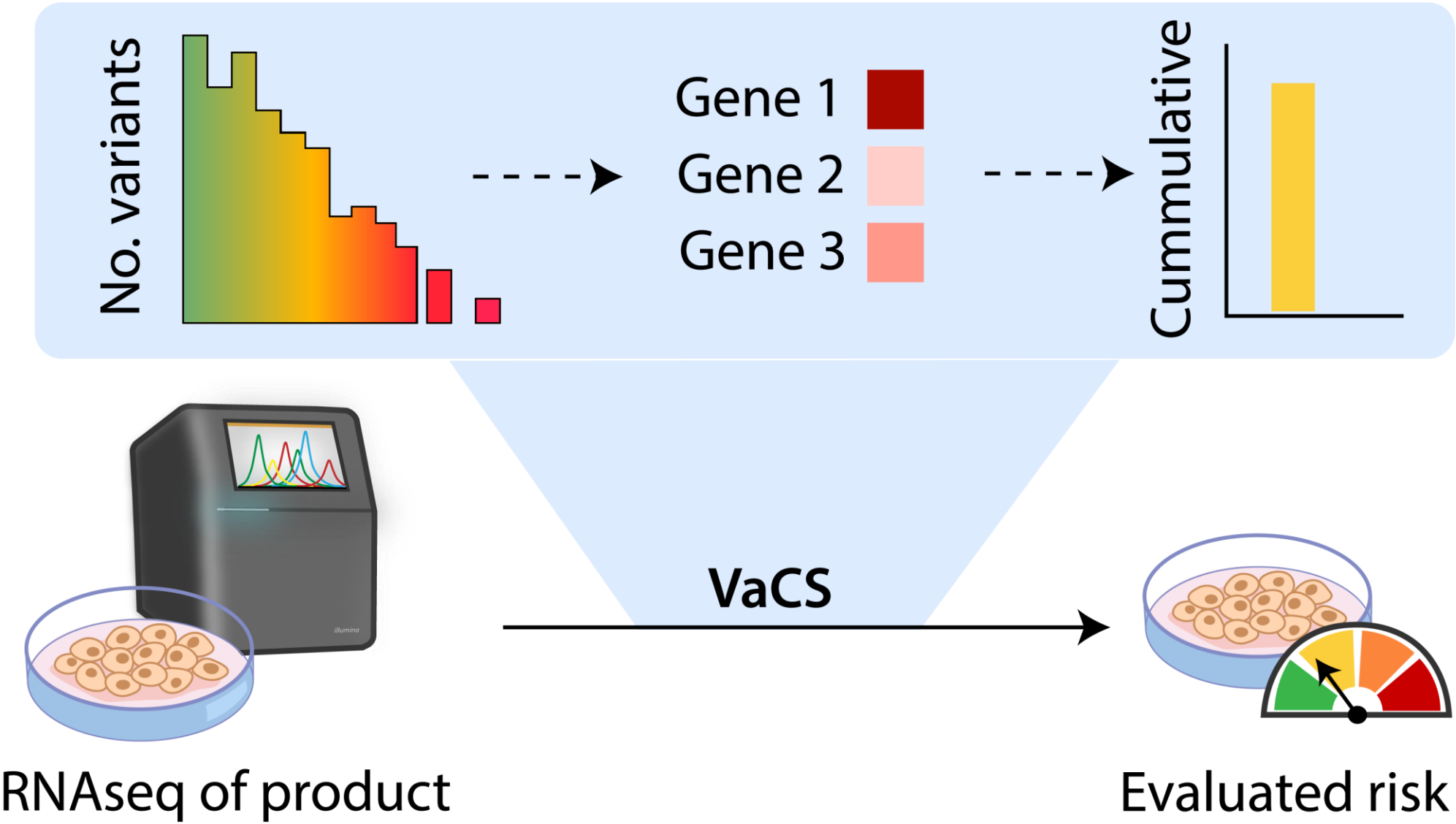


Summing aggregated gene-level VaCS enables sample-level comparison



Summing aggregated gene-level VaCS enables sample-level comparison




No. variants
Gene 1
Gene 2
Gene 3
Cummulative
VaCS
RNAseq of product
Evaluated risk

Universiteit Leiden

Prof. Micha Drukker, PhD
Christian Schröter, PhD

Andreea Iosif, MSc
Nina Schulten, MSc
Tiago Scarano, BSc

Max Fernkorn, PhD
Benjamin Tak, MSc
Ruben de Vries, MSc
Eliane Züger, MSc
Joey Zuijdervelt, MSc
Yubin Guo, MSc

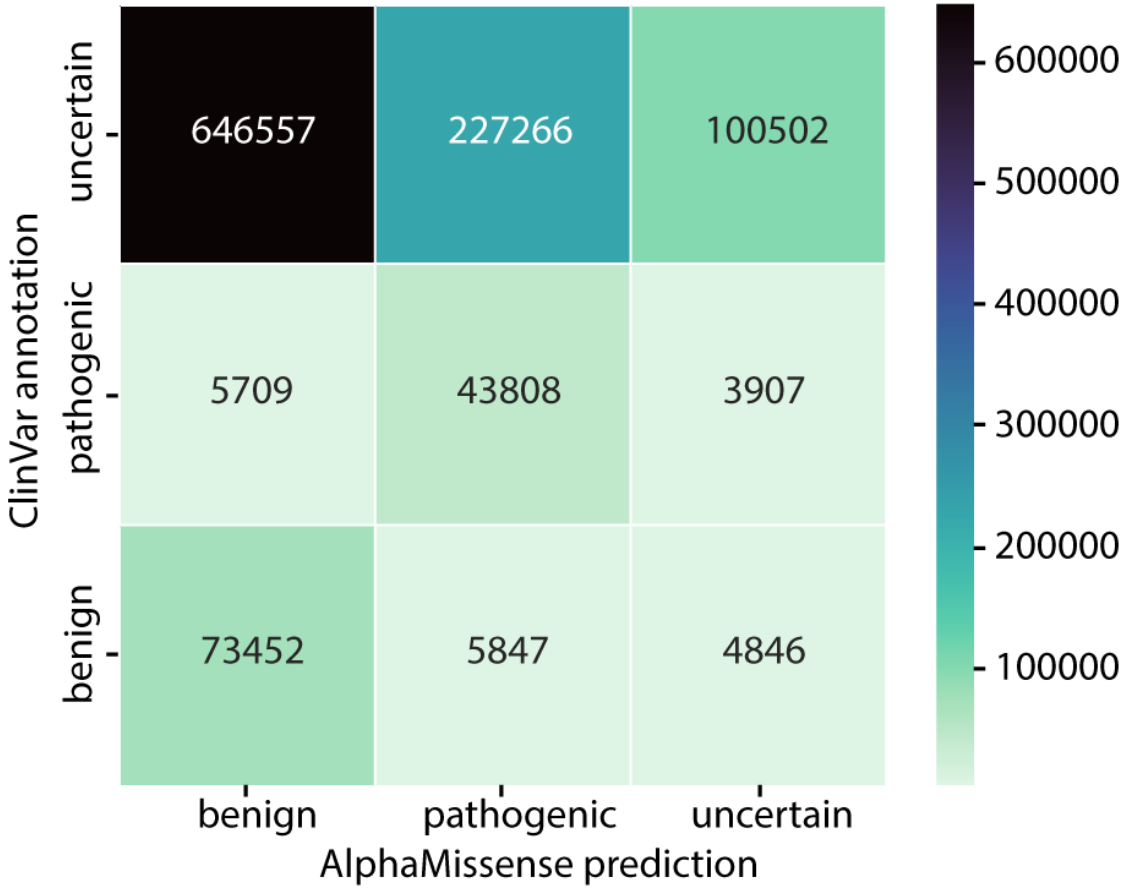
CBG MEB

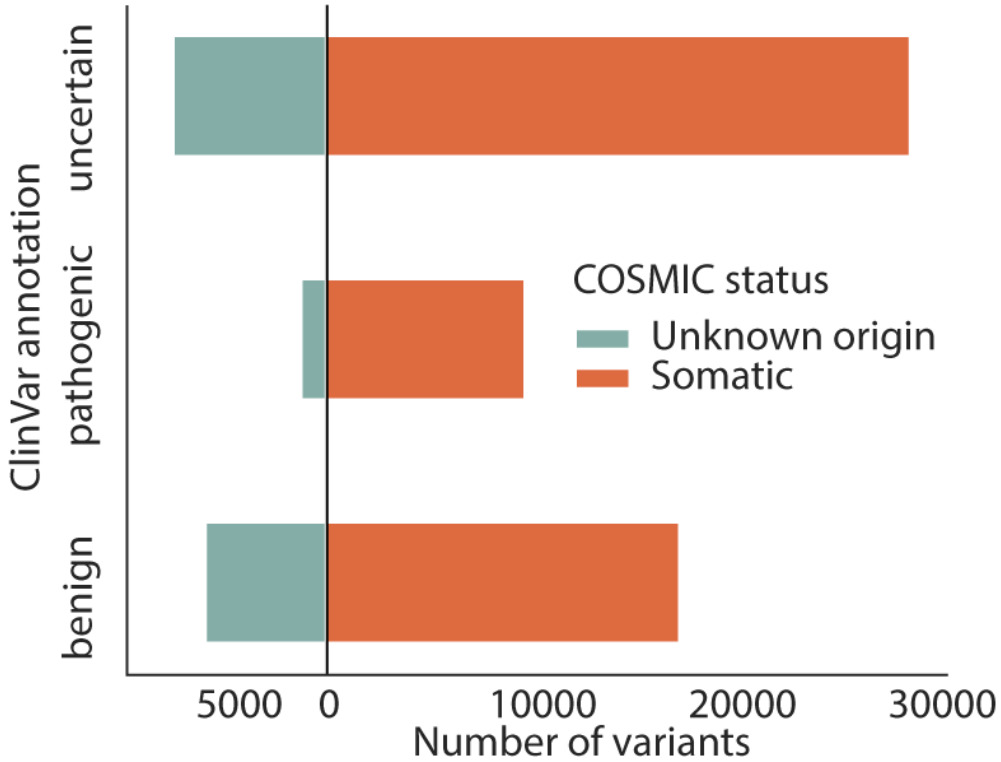
Tineke van den Hoorn, PhD
Marcel Hoefnagel, PhD



drukkerlab.nl

Existing resources provide conflicting pathogenicity annotations





VaCS resolves discordant annotations through evidence integration

