Integrated Long-Read Target Enrichment and Comprehensive Genomic Profiling for Hematologic Malignancies Using the SureSelect Cancer Pan Heme Assay

Brandyn Clark[1], Adam Janssen[1], Jeff Fox[1], Nedda Saremi[2], Kristi Stephenson[2], Kelle Hammock[2], Bahram Arezi[1]

[1]Agilent Technologies, La Jolla, CA [2]Agilent Technologies, Santa Clara, CA

AACR
POSTER
#2578

PR7004-1368

Introduction

Next-generation sequencing (NGS) has transformed cancer research, yet its effectiveness is often constrained by sample quality, tumor fraction, and limitations inherent to short-read sequencing—particularly in detecting structural variants, complex rearrangements, and alterations in repetitive or polymorphic regions. Long-read sequencing offers a solution, but cost and throughput challenges persist. To address these limitations, we developed a flexible, automation-compatible library prep and target enrichment workflow that supports DNA inputs as low as 200 ng, accommodates both enzymatic and mechanical shearing, and enables fast hybridization (90 minutes). Coupled with Oxford Nanopore's long-read sequencing and the SureSelect Cancer Pan Heme assay, this platform enables detection of diverse genomic alterations—including SNVs, indels, CNVs, and gene fusions—within a single assay.

The SureSelect Cancer Pan Heme panel, codeveloped with Roswell Park Comprehensive Cancer Center, interrogates 359 DNA and 124 RNA genes, delivering integrated DNA/RNA analysis that may help address certain limitations of conventional single-analyte methods such as karyotyping, FISH, and PCR. We demonstrate high enrichment efficiency using a novel fast hybridization buffer, achieving on-target rates of ~80% for libraries with insert sizes up to 4–5 kb . Comparative analysis reveals superior coverage in challenging genomic regions using enriched long-read sequencing versus short-read approaches. This solution offers a scalable, automation-compatible, and streamlined workflow for molecular laboratories, supporting comprehensive genomic profiling applications relevant to hematologic malignancies.

Experimental

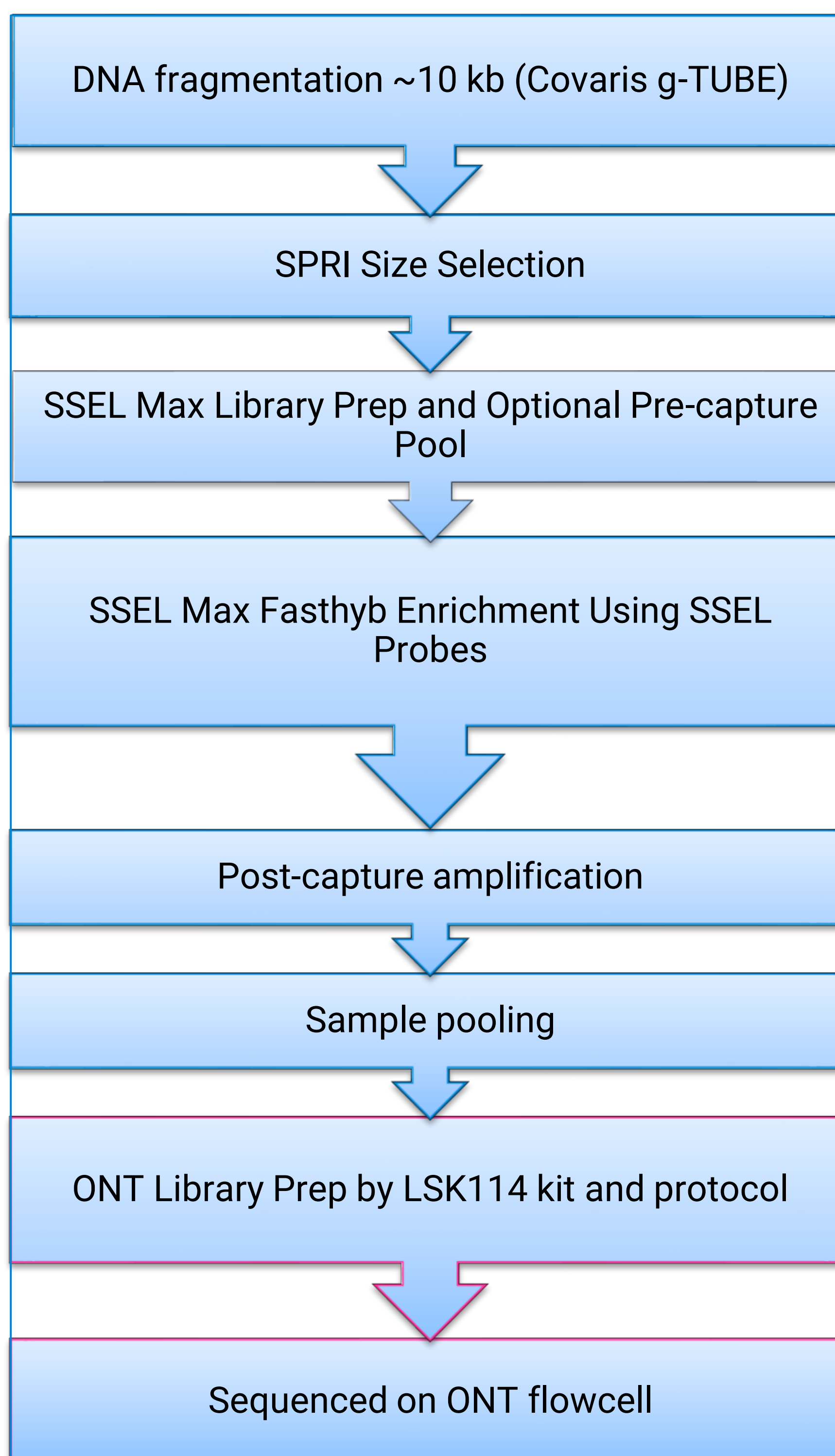


Figure 1. Workflow diagram for library prep and capture with SureSelect Max chemistry, followed by ONT Sequencing. This SureSelect Max for Long Read workflow has been further optimized since these libraries were prepared.

Experimental

Sample Details

Coriell samples NA12878, NA24385, NA12149, and NA12143 were used. 24 samples were prepped in total; 8 samples were captured individually; 16 samples were pooled together into a single capture pool using Agilent SureSelect Max library prep and target enrichment chemistry. All captures were performed using the Agilent SureSelect Cancer Pan Heme panel. After capture, samples were prepared for ONT sequencing using Ligation Sequencing kit library prep (LSK114) and sequenced on an ONT PromethION.

Analysis Methods

Basecalling was performed with Guppy[1] v5.2.0 using the SUP model. Libraries were demultiplexed using ONT's Dorado® software using Agilent SureSelect Max dual indices. Cutadapt[2] was used for adapter and UMI removal, and reads were mapped with minimap2[3] using the lr:hq model to GRCh38. The tools listed below were used for further analysis of the sequenced libraries.

Tool	Parameters
Picard[4] CollectHSMetrics	•NEAR_DISTANCE=5000 •MINIMUM_MAPPING_QUALITY=0 •MINIMUM_BASE_QUALITY=0 •CLIP_OVERLAPPING_READS=false
Picard[4] CollectGcBiasMetrics	•SCAN_WINDOW_SIZE 100
samtools[5] stats	•RL (Read lengths)
Sawfish[6]	–disable-cnv for non-WGS input

Results and Discussion

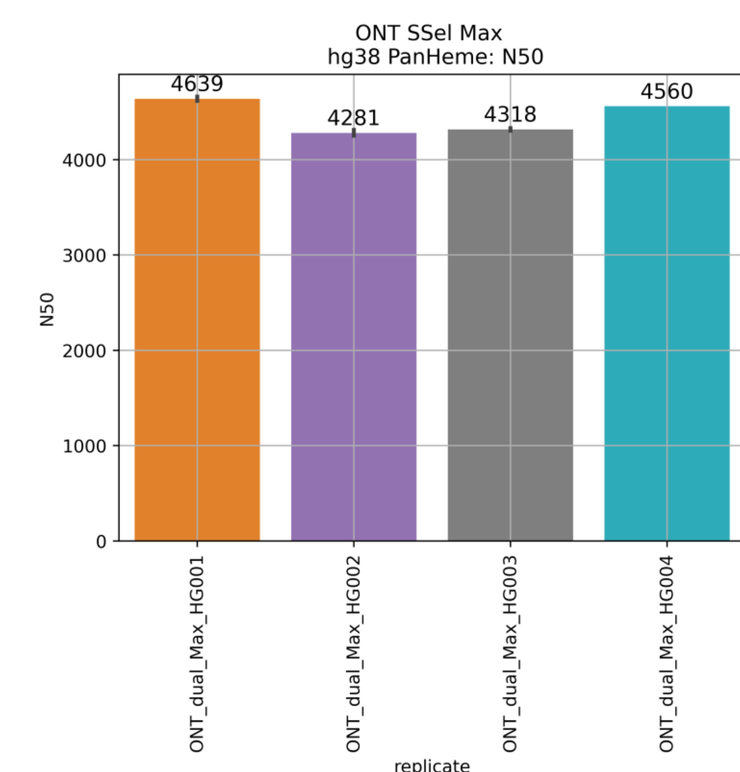


Figure 2. N50 for DNA inputs from NA12878, NA24385, NA12149, and NA12143 show long inserts exceeding 4 kb. For each sample, the plotted N50 represents the average calculated from two independent replicates.

Sample	ONT dual end HG001	ONT dual end HG002	ONT dual end HG003	ONT dual end HG004
Mean Fragment	3,859	3,390	3,542	3,896
Median Fragment	3,185	3,252	3,526	3,892
Total Reads	4.21E6	4.81E6	5.11E6	4.02E6

Table 1. Mean, median average base pair value and total reads for NA12878, NA24385, NA12149, and NA12143.

For Research Use Only (RUO). Not For Use in Diagnostic Procedures.

Results and Discussion

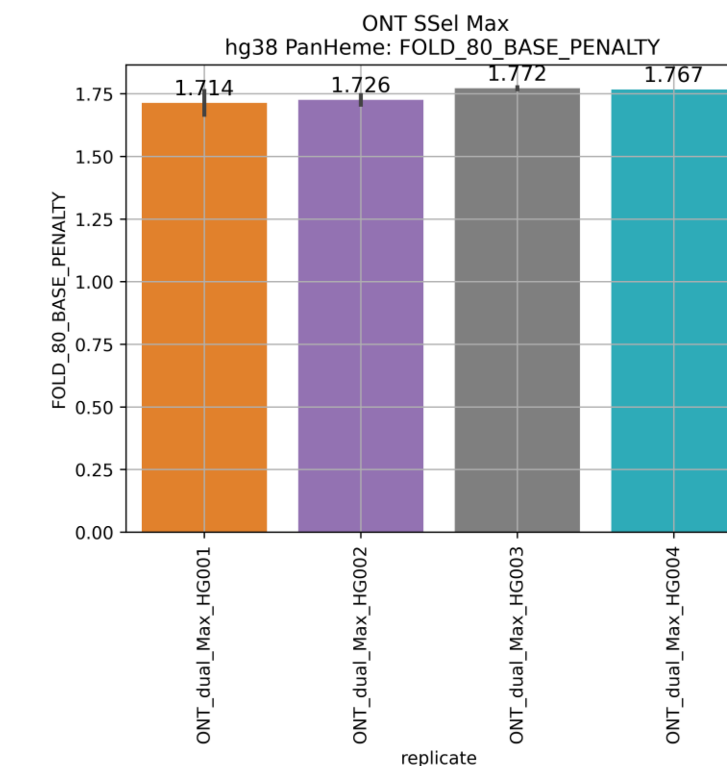
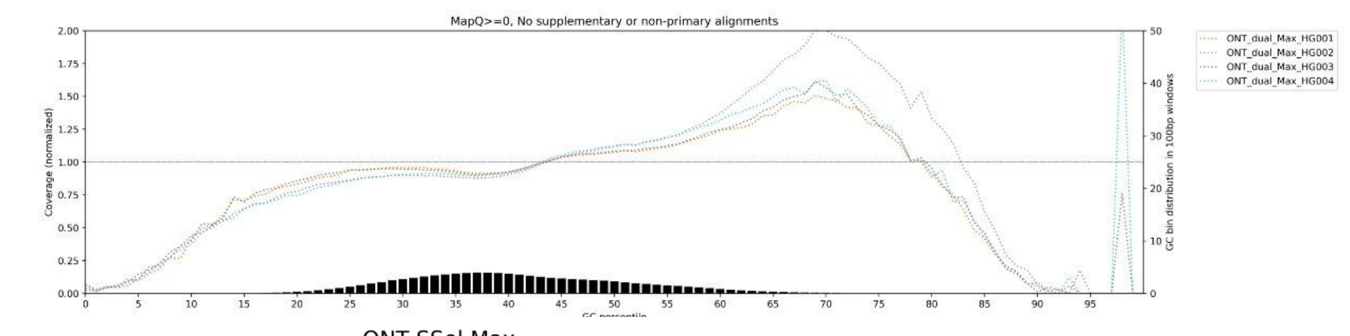


Figure 3. Fold 80 base penalty metrics are shown for each DNA input. All samples demonstrate Fold 80 values below 1.8, indicating efficient and uniform coverage across inputs when looking at gene-only content. The values are tightly clustered, reflecting highly comparable performance and minimal variability. All total reads were used in analysis.

Figure 4. GC-bias profiles are consistent across all samples, showing nearly identical coverage trends from low to high GC content.



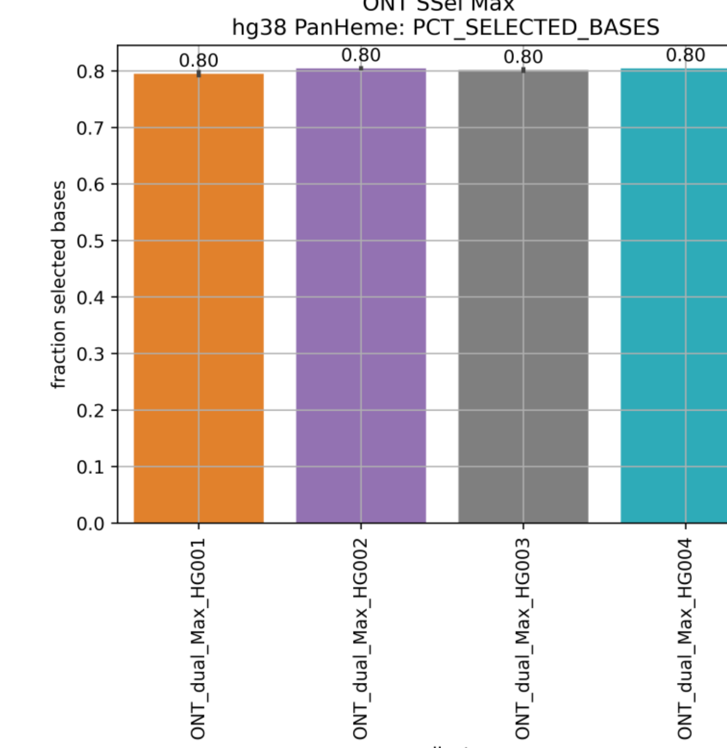
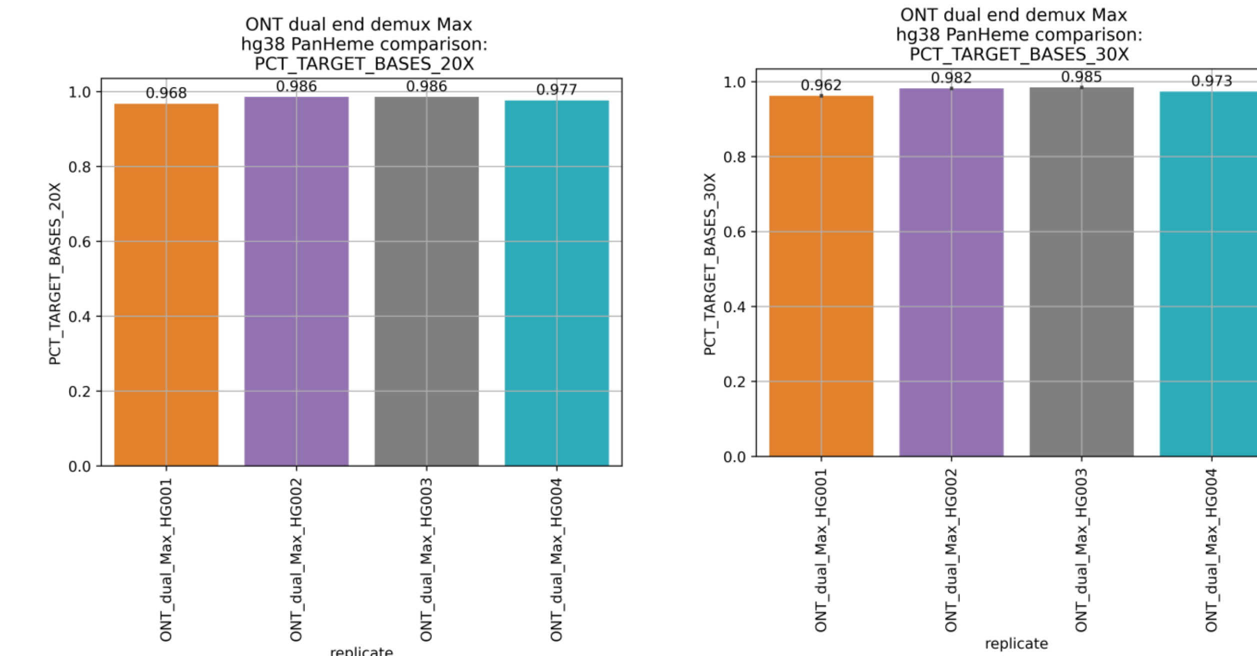


Figure 5. Percent selected bases reflects the proportion of reads mapping to intended target regions. Approximately 80% of total bases are on-target across all replicates, indicating consistent and efficient panel capture.

Figure 6. Percent of targeted bases at 20X and 30X coverage are above 96% when looking at gene-only content. This is consistent across all sequenced libraries within this study. All total reads were used in this analysis.



Results and Discussion

Figure 7. IGV screenshot showing a 3,988 bp deletion on chr13 in HG002 with clear support from long read data. The long reads (bottom 2 tracks) span the regions flanking the deletion, providing contiguous coverage across the breakpoints, whereas the short reads (top 4 tracks) show less definitive support for the deletion.

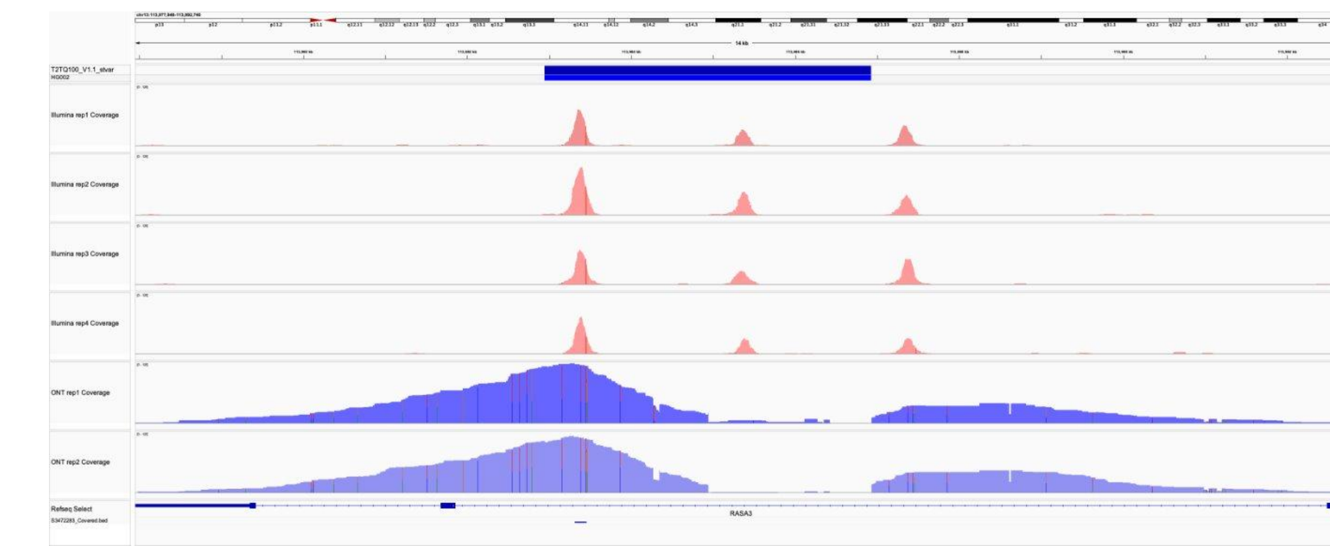
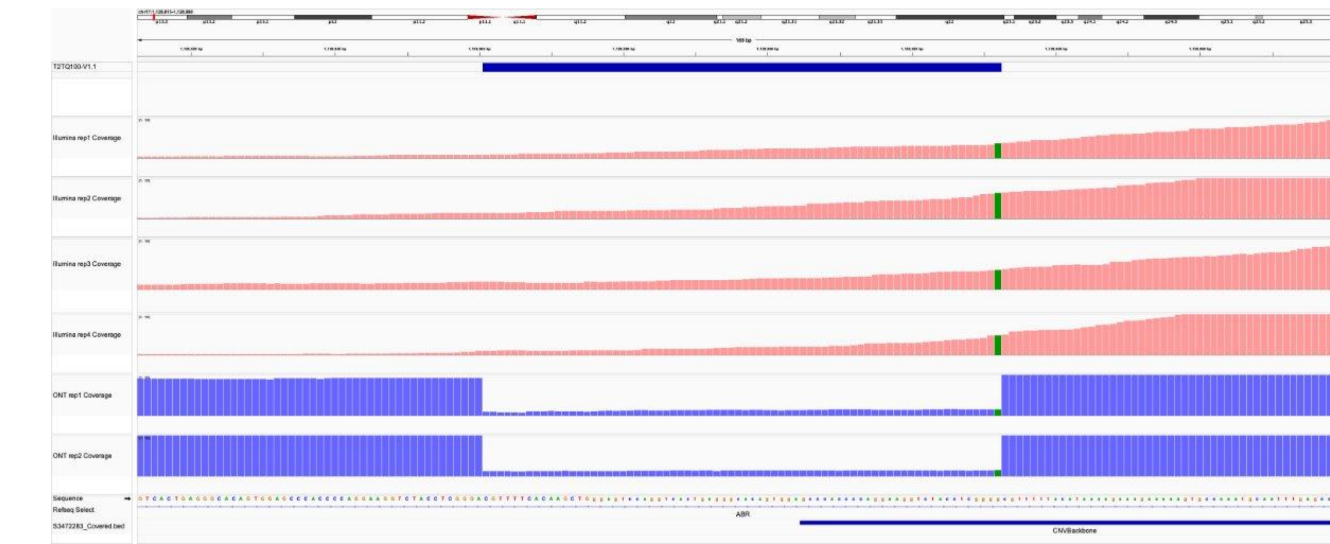


Figure 8. IGV screenshot of a 72 bp deletion on chr17 in HG002 detected by long read sequencing (bottom 2 tracks) and poorly resolved with short read sequencing (top 4 tracks).



Conclusions

- Off-the-shelf SureSelect panel optimized for short-read sequencing generated robust performance. Modifications in bait panel design (or using a spike-in) may help improve performance further.
- Off-the-shelf SureSelect Max kit can be used with minimal changes in the workflow and no need for additional enzymes, beads, etc.
- We have also successfully enriched libraries prepared using SeqWell's enzymatic Long Plex Multiplexing kit plus SSEL Max hybridization and enrichment.
- An automated Magnis long read alpha protocol is available. Contact your Agilent representative to discuss research opportunities.

References

1. Oxford Nanopore Technologies, https://nanoporetech.com/software/other/guppy
2. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200
3. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100.
4. Broad Institute. Picard Toolkit. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/.
5. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021). https://doi.org/10.1093/gigascience/giab008
6. Saunders, C.T. et al. Sawfish: improving long-read structural variant discovery and genotyping with local haplotype modeling. *Bioinformatics* **41**, btaf136(2025). https://doi.org/10.1093/bioinformatics/btaf136