PureTarget: An amplification-free workflow for genetic and epigenetic profiling of short tandem repeat expansions

Poster #: P14.004.B
Guilherme De Sena Brandine, Valeriya Gaysinskaya, Janet Aiyedun, Julian Rocha, Duncan Kilburn, Sarah Kingan, Egor Dolzhenko, Zoi Kontogeorgiou, Anita Szabo, Christina Zarouchlioti, Robert Thaenert, Fabio Fuligni, Aidan Hennigan, Chelsea Roselund, Alesia Piselli, Pilar Alvarez Jerez, Kimberley Billingsley, Sonia Lameiras, Sylvain Baulande, Petra Liskova, Alice Davidson, Georgios Koutsis, Georgia Karadima, Stéphanie Tomé, Michael Eberle

Goal

The PureTarget repeat expansion panel allows base-resolution genotyping of 20 short tandem repeat loci. It is an amplification-free protocol which enables simultaneous genotyping and methylation profiling for a precise and scalable sequencing of common repeat expansion-associated disease targets (Fig. 1).

The PureTarget protocol

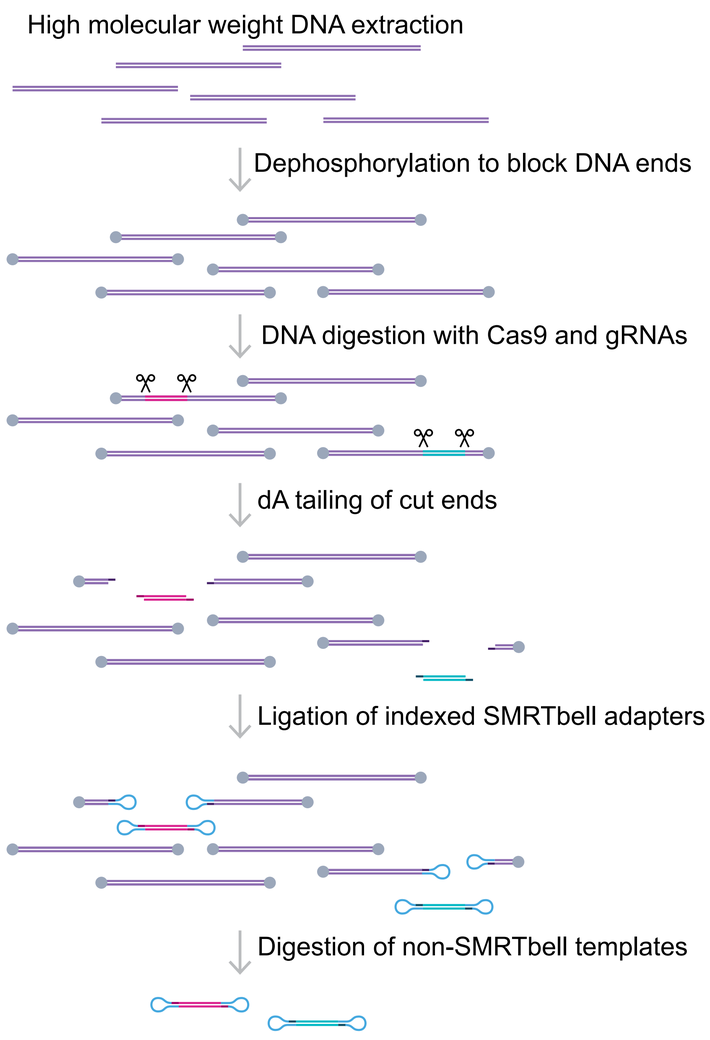
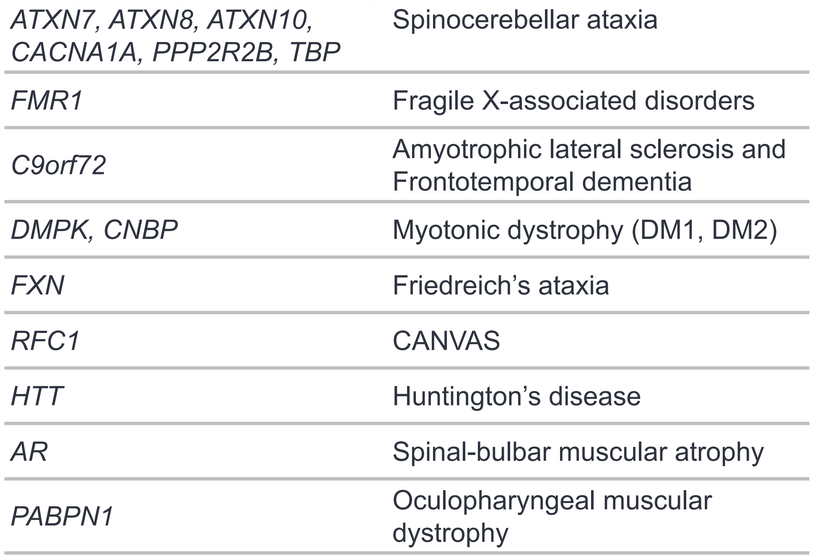


Figure 1. The targeting and library prep workflow uses CRISPR/Cas9 system and single guide RNAs to cut high molecular weight DNA ~2-3kb upstream and downstream of 20 short-tandem repeat targets. The final step is a nuclease treatment to remove non-SMRTbell library templates resulting in a library enriched for the panel targets [1] .

Repeat expansion panel of 20 targets

Gene(s)	Associated disease
ATN1, ATXN1, ATXN2, ATXN3, ATXN7, ATXN8, ATXN10, CACNA1A, PPP2R2B, TBP	Spinocerebellar ataxia
FMR1	Fragile X-associated disorders
C9orf72	Amyotrophic lateral sclerosis and Frontotemporal dementia
DMPK, CNBP	Myotonic dystrophy (DM1, DM2)
FXN	Friedreich's ataxia
RFC1	CANVAS
HTT	Huntington's disease
AR	Spinal-bulbar muscular atrophy
PABPN1	Oculopharyngeal muscular dystrophy
TCF4	Fuchs endothelial corneal dystrophy

Computational workflow



Workflow is available as push-button analysis using the SMRT Link v13.1 software bundle.

Accuracy

We tested the protocol in 129 samples with verified expansions, totaling 2,580 loci. All datasets contain one confirmed expansion verified orthogonally. We verified consistency across 18 technical replicates (Fig. 2), scalability with multiplexing (Fig. 3) and identification of all expected expansions in dominant (Fig. 4) & recessive (Fig. 5) alleles.

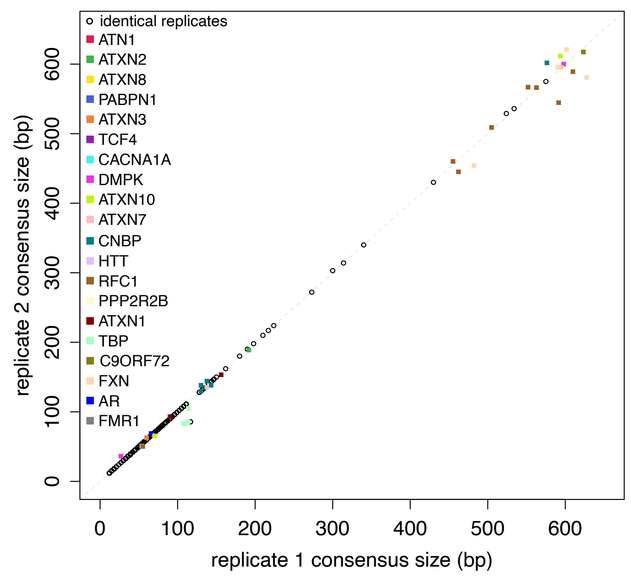


Figure 2. Consensus comparison in 18 pairs of technical replicates. 614/657 have identical consensus sequences, 627 are at most off by 1, and all (657/657) have concordant ranges, meaning the range of allele sizes overlap between replicates.

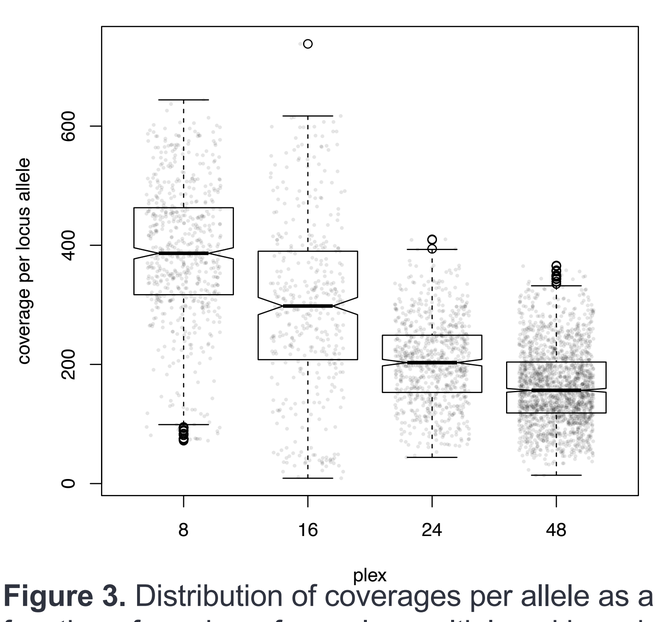


Figure 3. Distribution of coverages per allele as a function of number of samples multiplexed in a single run.

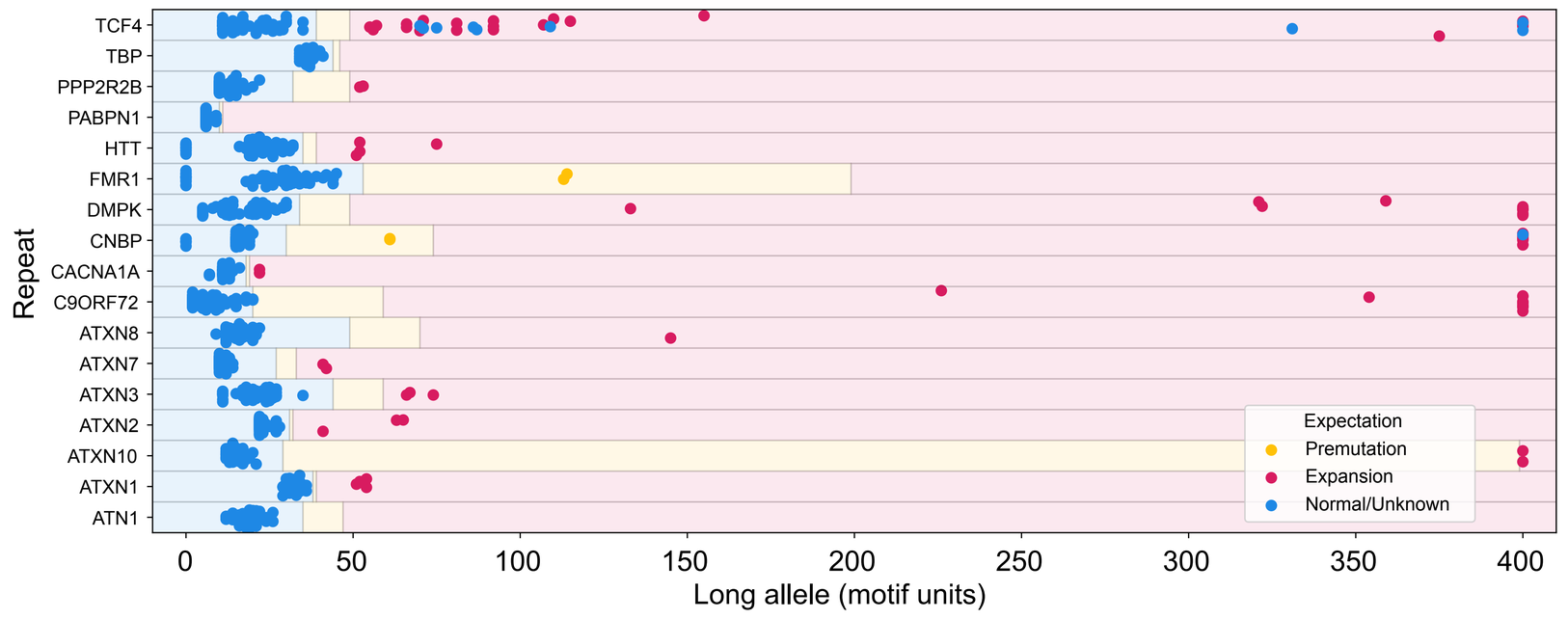


Figure 4. "Swim lane plot" showing the long allele length of 129 samples for 17 autosomal and X-linked dominant loci. Dots are colored by expected genotype. One *CNBP* and nine *TCF4* expansions in samples expected normal were identified.

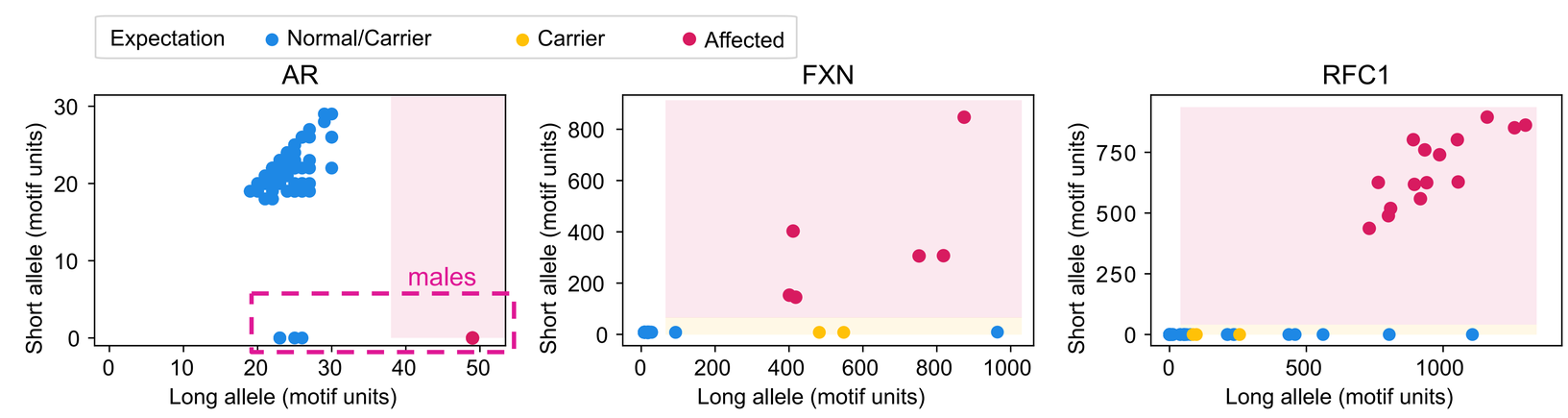


Figure 5. Scatterplot showing short and long allele lengths of 126 samples for 3 autosomal (FXN, RFC1) and X-linked (AR) recessive loci.

Interpretation

The TRGT software tool [2] allows flexible visualization of genotype and variation for methylation (Figure 6), mosaicism (Figure 7) and complex motif composition (Figure 8).

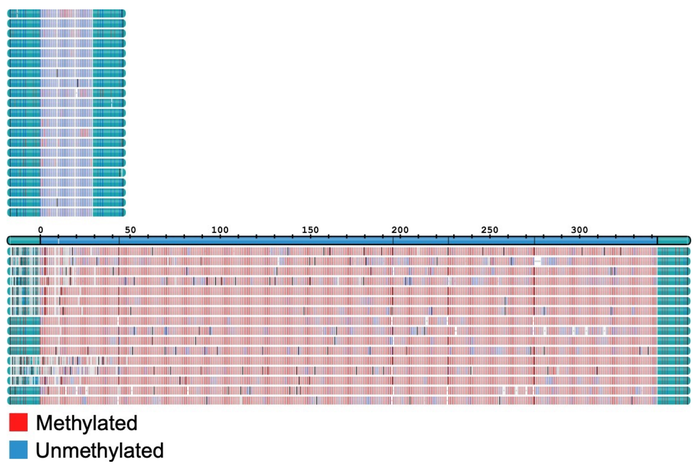


Figure 6. Allele methylation of the *FMR1* CGG motif is resolved at single-base resolution in sample NA07537.

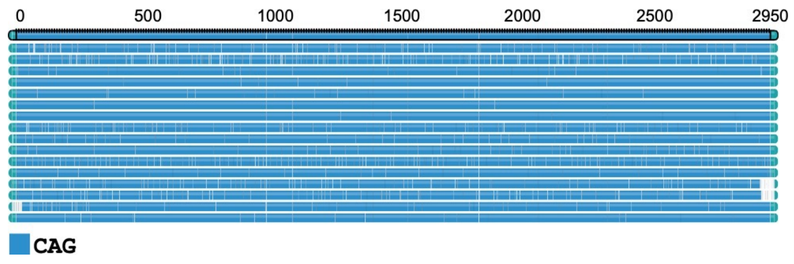


Figure 7. Long *DMPK* expansions with low mosaicism can be precisely genotyped.

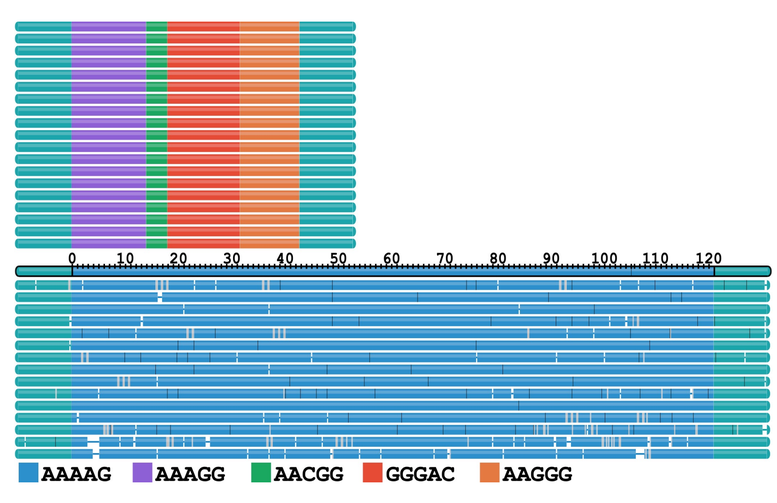


Figure 8. The protocol allows diverse motif composition in the *RFC1* locus.

Acknowledgements

The authors would like to thank Yu-Chih Tsai, John Harting, Jenny Ekholm, Ian McLaughlin, Cheryl Heiner, and Janet Ziegle for their contributions to the original No-amp protocol, on which this method was based.

References

[1] Tsai, Y. C., de Pontual, L., Heiner, C., Stojkovic, T., Furling, D., Bassez, G., ... & Tomé, S. (2022). Identification of a CCG-Enriched expanded allele in patients with myotonic dystrophy type 1 using amplification-free long-read sequencing. *The Journal of Molecular Diagnostics*, *24*(11), 1143-1154.

[2] Dolzhenko, E., English, A., Dashnow, H., De Sena Brandine, G., Mokveld, T., Rowell, W.J., Karniski, C., Kronenberg, Z., Danzi, M.C., Cheung, W.A. and Bi, C., 2024. Characterization and visualization of tandem repeats at genome scale. *Nature Biotechnology*, pp.1-9.v