Long-read sequencing panels to consolidate challenging germline targets – applications for carrier screening and repeat expansion disease in diagnostic research

G007

Abstract # 2169890
Sarah B. Kingan, Jeff Zhou, Anupam K Chakravarty, Guilherme De Sena Brandine, Tom Mokveld, Egor Dolzhenko, Xiao Chen, Duncan Kilburn, Heather Ferrao, Michael A. Eberle
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Native long read panels for complex genes

Targeted long-read sequencing enables efficient characterization of tandem repeats, structural variants, and copy number variants that currently require non-NGS assays like rpPCR, MLPA, long range PCR, and Sanger sequencing. Here we demonstrate that our PCR-free Cas-9 target enrichment panels can resolve these variants in clinically-relevant regions of the genome.

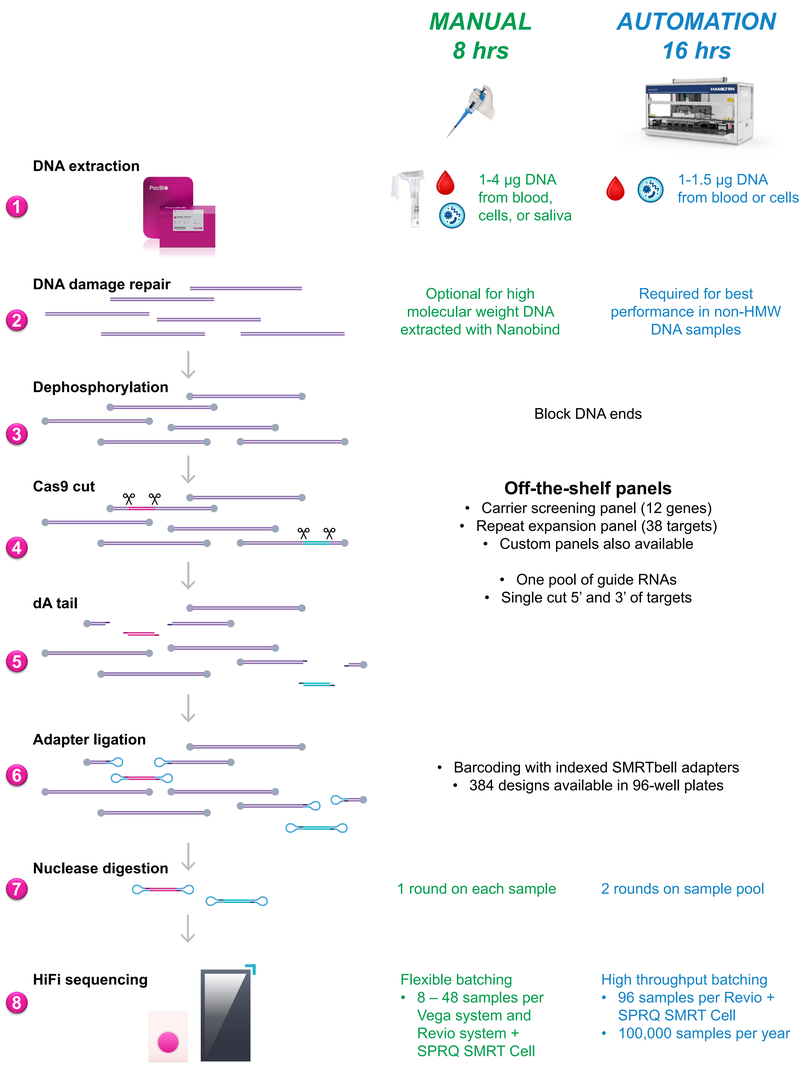
PureTarget carrier screening panel (12 genes)

- *AFF2, ARX, CYP21A2/TNXB, F8, FMR1, FXN, GBA, HBA1/2, HBB, RPGR, SMN1/2,*

PureTarget repeat expansion panel 2.0 (38 targets)

- *ATN1, ATXN1, ATXN2, ATXN3, ATXN7, ATXN8, ATXN10, CACNA1A, PPP2R2B, TBP, BEAN1, DAB1, FGF14, NOP56, ZFHX3 , FMR1 , AFF2 , AFF3 , C9ORF72 , FXN , RFC1 , NOTCH2NLC , DMPK, CNBP , HTT , JPH3 , TCF4 , AR , PABPN1 , ABCD3, GIPC1, LRP12, PILPL1 , HOXD13 , PHOX2B , PRNP , CSTB , SAMD12*

PCR-free library prep workflows



Benchmarking repeat expansion panel

Sample	Expected genotype	PureTarget genotype	Concordance
RE_1	ABCD3 7 / 320	GCC:8 / GCC:315	
RE_2	ATN1 15+-1 / 49+-1	CAG:12 CAA:2 / CAG: 48 CAA:2	
RE_3	ATXN1 32+-1 / 49+-2	CTG:29 ATG:2 / CTG:47	
RE_4	ATXN10 normal <32 / 120-125 ATTCT 1400-1500 ATTCC	ATTCT:12 / ATTCT:139,ATTCC:1353	
RE_5	ATXN2 normal <30 / 38+-1	CTG:20 TTG:2 / CTG:38	
RE_6	ATXN3 14+-1 / 78+-2	CTG:11 CTT:1 TTG:2 / CTG:75 CTT:1 TTG:2	
RE_7	ATXN8 normal <50 / 240-260	CTA:9 CTG:12 / CTA:12 CTG:253 CCG:4	
RE_9	CACNA1A normal <18 / 23+-1; ATXN3 intermediate (45-59)	CACNA1A CTG:13 / CTG:22; ATXN3 CTG:20 CTT:1 TTG:2 / CTG:42 CTT:1 TTG:2	
RE_10	AR 42+-1	GCA:50	
RE_11	CNBP normal < 26 / pathogenic > 75	CAGG:13 CAGA:11 CA:17 / CAGG:2714 CAGA:43 CA:147	
RE_12	DAB1 normal < 30 / pathogenic >31	AAAAT:15 / AAAAT:145 GAAAT:96	
RE_13	FGF14 normal <179 / 250-290	GAA:68 / GAA:250	
RE_14	FGF14 normal <179 / alternative 375 GAAGGA	GAA:9 / GAA:383 GGA:358	
RE_15	FGF14 normal <179 / 400-590	GAA:16 / GAA:645	
RE_16	HTT 17 / 40	CAG:18 CCG:8 CAA:1 CCA:1 / CAG:41 CCG:8 CAA:1 CCA:1	
RE_17	NOP56 normal <14 / pathogenic >650	GGCCTG:6 CGCCTG:2 / GGCCTG:1159 CGCCTG:2	
RE_18	RFC1 360 AAAGG / 520 AAGGG	AAGGG:1 AGAGG:1 AAAAG:7 AAAGG:286 AAAGGG:8 / AAGGG:548	
RE_19	RFC1 normal / **AGGGG expanded**	AAAAG:87 / AAAGG:5 AAGGG:29 AAAGGG:1 **AGGGG:96**	
RE_20	TBP normal <40 / 43+-1	GCA:32 ACA:5 / GCA:37 ACA:5	
RE_21	TCF4 normal 7 / 11	CAG:11 / CAG:15	
RE_22	ZFHX3 normal <26 / pathogenic > 46	GCC:18 ACT:1 ACC:2 / GCC:58	

Table 1. Functional concordance for 21 / 21 samples from MGZ lab.
PureTarget repeat expansion panel 2.0 was prepared with 24 DNA samples and sequenced on Revio + SPRQ. Positive samples extracted with NucleoMag and FlexiGene were previously genotyped with orthogonal methods and 2 Nanobind-extracted blood samples from a random donor were included as controls; 1 sample had no index added and was omitted from analysis. Samples were analyzed with PureTarget repeat expansion panel analysis workflow in SMRT Link 25.4 using TRGT[1] using custom BED file containing interruption sequences observed in the samples. Functional concordance was assessed using normal/intermediate/pathogenic repeat size cuts offs from strchive.org.

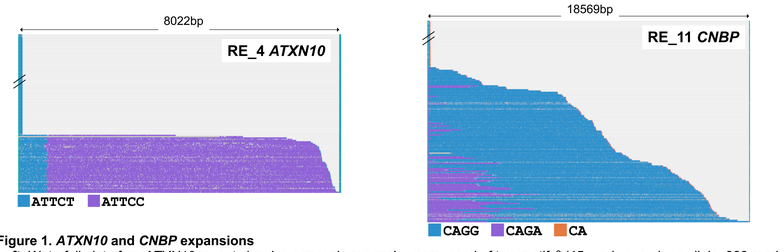
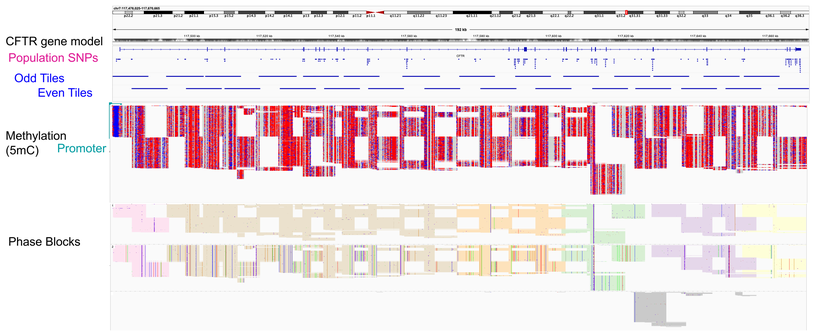


Figure 1. *ATXN10* and *CNBP* expansions
Left. Waterfall plot of an *ATXN10* repeat showing a mosaic expansion composed of two motifs[2] (45 reads span long allele, 389 reads span short allele). **Right**. Waterfall plot of a *CNBP* repeat showing a mosaic expansion spanning up to 18kb (101 reads span long allele, 619 reads span short allele). The short allele was cropped in both figures to better visualize the expansion.

Custom tiled designs for large genes



CFTR design implemented as proof of concept for gene phasing

- Full gene coverage (~190 kb) of overlapping tiles to enable phasing
- Guide RNA (gRNA) design informed by population SNPs to give robust coverage across samples from different human populations
- Final design: 28 gRNA pairs capturing ~10 kb fragments
- Library prep requires separate Cas9 cutting step for odd and even gRNA pools

Benefits of approach:

- 6 phase blocks resolved in random donor control DNA
- Methylation information retained

Benchmarking carrier genes (F8, SMA, CAH)

Expected genotype (F8)	PureTarget genotype	No. unique samples	Concordance
Intron 1 inversion 0 ; Intron 22 inversion 0	0 ; 0	2	
Intron 1 inversion 0 ; Intron 22 inversion 1	0 ; 1	2	
Intron 1 inversion 0/0 ; Intron 22 inversion 0/0	0/0 ; 0/0	1	
Intron 1 inversion 0/0 ; Intron 22 inversion 0/1	0/0 ; 0/1	8	
Intron 1 inversion 0/1 ; Intron 22 inversion 0/0	0/1 ; 0/0	2	

Table 2. Benchmarking of intron 1 and intron 22 inversion in *F8* (Hemophilia A). PureTarget guides were designed to span inversion breakpoints. Reads were re-aligned with Paraphase[3] to the reference and read alignments are analyzed to identify haplotypes and inversion breakpoints. Analysis of 15 samples containing 5 different genotypes with PTCP v3.0.0[4] was concordant with expectations.

Exp copy number (*SMN1*/ *SMN2*)	PureTarget copy number	No. Unique Samples	Concordance
<2 SMN1	1/2	1	
0/2	0/2	2	
0/3	0/3	2	
0/4	0/4	2	
1/1	1/1	3	
1/2	1/2	1	
1/2	2/2	1	HG01612
1/3	1/3	2	
1/4	1/4	1	

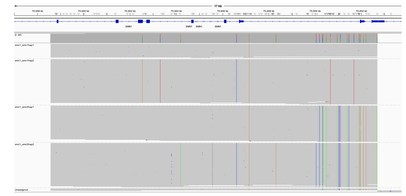
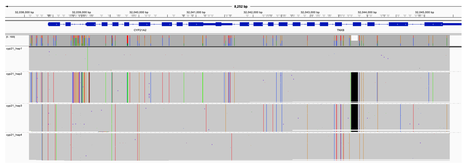


Table 3. Benchmarking *SMN1/2* copy number (Spinal muscular atrophy). Paraphase was used for mapping and genotyping, with additional read-depth correction for identical haplotypes. Samples (N=15) with 9 unique genotypes were analyzed with PTCP v3.0.0. Concordant copy number calls were found for 14/15 samples. Sample HG01612 (inset) clearly shows 4 haplotypes, so the source of discordance with expectations remains unclear.

Exp copy number (*CYP21A2 / CYP21A2P*)	Expected variants	PureTarget copy number	PureTarget variants	Concordance
Conversion	N/A	1/3	N/A	
Deletion	c.844G>T	1/3	c.844G>T	
Pseudogene deletion	N/A	2/1	N/A	
N/A	c.844G>T	2/2	c.844G>T	
N/A	c.955C>T	2/2	c.955C>T	
N/A	c.844G>T	2/2	c.844G>T	
Conversion	c.1360C>T	2/2	c.1360C>T	
Hybrid allele	c.597A>T	2/2	c.597A>T, c.518T>A (hap2)	
N/A	c.518T>A (het), c.1226G>A (het), c.1439G>T (het)	2/2	c.518T>A (hap2), c.1226G>A (hap3), c.1439G>T (hap3)	
Duplication	N/A	2/3	N/A	
N/A	c.518T>A, c.844G>T	2/3	c.518T>A (hap2), c.844G>T (hap4)	
Hybrid alleles	c.597A>T	3/1	c.597A>T (hap2), c.597A>T (hap4)	
Duplication	c.955C>T	3/2	c.955C>T	
Duplication	c.955C>T	3/2	c.955C>T	
Duplication	N/A	3/2	N/A	

Table 4. Benchmarking of *CYP21A2* copy number and variants (congenital adrenal hyperplasia). Paraphase was used for mapping and genotyping. Samples (N=15) were analyzed with PTCP v3.0.0, and complete concordance was observed for both copy number and variant calls. Example of sample with 1 gene and 3 pseudogene copies (inset).



Conclusion

- PureTarget long-read assay combined with our software pipeline enables analysis of clinically-important variation in regions that are difficult to profile with short-read technologies
- Can consolidate multiple legacy genotyping assays like MLPA and PCR
- PCR-free method requires 1 µg of DNA
- Automation workflow capable of processing 100,000 samples on PacBio Revio + SPRQ
- Curated panels for carrier screening and repeat expansions available in addition to custom panels

Acknowledgements

The authors acknowledge members of the MGZ Medical Genetics Center for beta testing PureTarget and generously sharing data for this poster.

References

1. Dolzhenko, E., et al. (2024). Characterization and visualization of tandem repeats at genome scale. Nat Biotechnol. 2024 doi: 10.1038/s41587-023-02057-3.
2. Sedov et al. 2025. Novel ATXN10 repeat motif patterns in Peruvian families modify disease age at onset. Neurology: Genetics accepted. https://doi.org/10.21203/rs.3.rs-5989910/v1
3. Chen X, Baker D, Dolzhenko E, et al. Genome-wide profiling of highly similar paralogous genes using HiFi sequencing. Nature Communications. 2025. doi:10.1038/s41467-025-57505-2
4. https://github.com/PacificBiosciences/ptcp