Enrichment of unamplified DNA and long-read SMRT Sequencing in unlocking the underlying biological disease mechanisms of repeat expansion disorders





CBIO®

For many of the repeat expansion disorders, the disease gene has been discovered, however the underlying biological mechanisms have not yet been fully understood. This is mainly due to technological limitations that do not allow for the needed base-pair resolution of the long, repetitive genomic regions.

We have developed a novel, amplification-free enrichment technique that uses the CRISPR/Cas9 system to target large repeat expansions. This method, in conjunction with PacBio's long reads and uniform coverage, enables sequencing of these complex genomic regions. By using a PCR-free amplification method, we are able to access not only the repetitive elements and interruption sequences accurately, but also the epigenetic information.

CRISPR/Cas9 system for targeted amplification-free enrichment

Complexity Reduction		Total Reads	нтт	ALS	FMR1	SCA10	% of Reads	
	DNA Input	1 SMRT Cell	Chr4 CAG	Chr9 GGGGCC	ChrX CGG	Chr22 ATTCT (various)	"On Target"	
None	1-5 μg	44,031	162	41	224	518	2.1%	
2 RE	5-10 μg	51,806	574	174	607	1254	5.0%	
4 RE	10-20 μg	45,676	1145	203	1016	1971	9.5%	
							J	

Number of individual molecules sequenced

Improved on-target rate with complexity reduction:

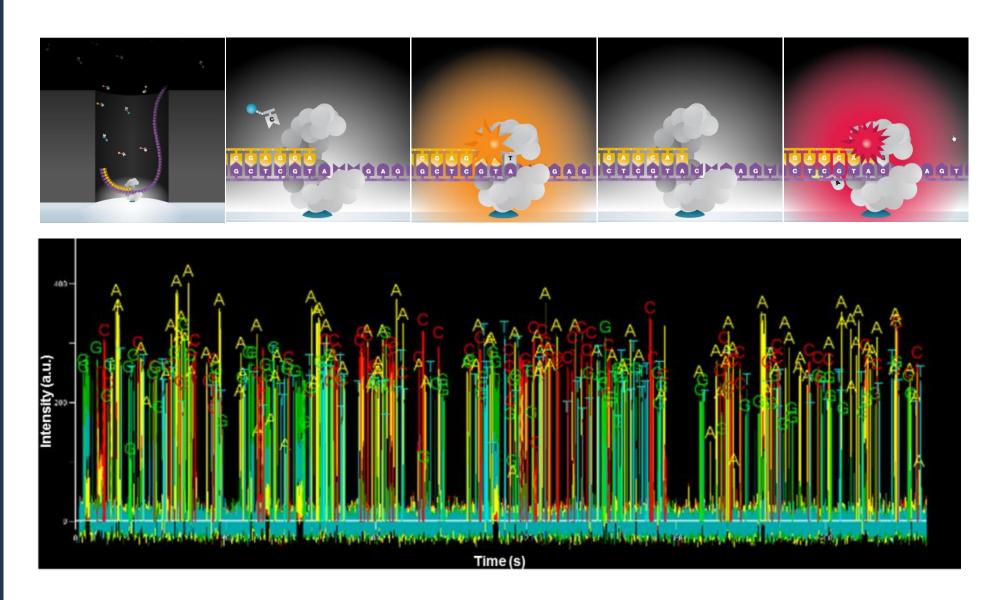
Characterization of *FMR1* pre-mutation sample

Repeat expansion disorders are challenging to interrogate due to the long repetitive regions. Using CRISRP/Cas9 we were able to access the repeat counts, interruption sequences, as well as epigenetic information, without introducing PCR bias. In addition, long-read sequencing allows for assembly-free, full-length isoform characterization.

Sequencing of *FMR1* gene, >700x CGG

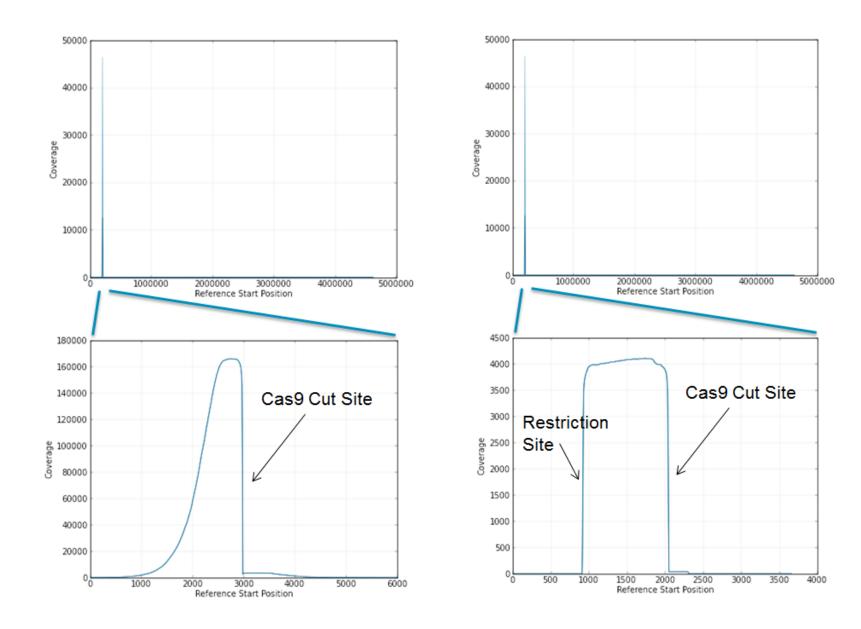
10	20) 30) 4	0 50	60	70	80	90	100	110	120	130	140	150
ATAGGATCAT 1	TTGATTTCCC	ACGCACTGAG	TGCACCTCTG	CAGAAATGGC	GTTCTGGCCC 1	CGGAGGCAG	TGCGACTGTC	ACCCTTCAGC	CTTCCCGCCT (CACCAAGCC G	CGCACGCCG GCC	GCGCGTC TGT	CTTTCGA CCGC	ACCCCG
160	170) 180) 19	0 200	210	220	230	240	250	260	270	280	290	300
GCGTTCCCAG	CAGCGCGCAT	GCGCGCGCTC	CAGGCCACTT	GAAAGAGAGG	GCGGGCCGAG (GGGCTGAAGC	CCGCGGGGGAG	GGAACAGCGT	TGATCACGTG A	CGTGTTTCA G	IGTTACACC CGC	AGCGGGC CGG	GGTTCGG CCTC	AGTCAG
310	320) 330	34	0 350	360	370	380	390	400	410	420	430	440	450
													1.	1 I I

SMRT Sequencing overview



In addition to calling the bases, SMRT Sequencing uses the kinetic information from each nucleotide to distinguish between modified and native bases.

- Restriction enzymes are used to degrade unwanted SMRTbell templates
- Additional starting DNA is required to maintain input into Cas9 digestion step



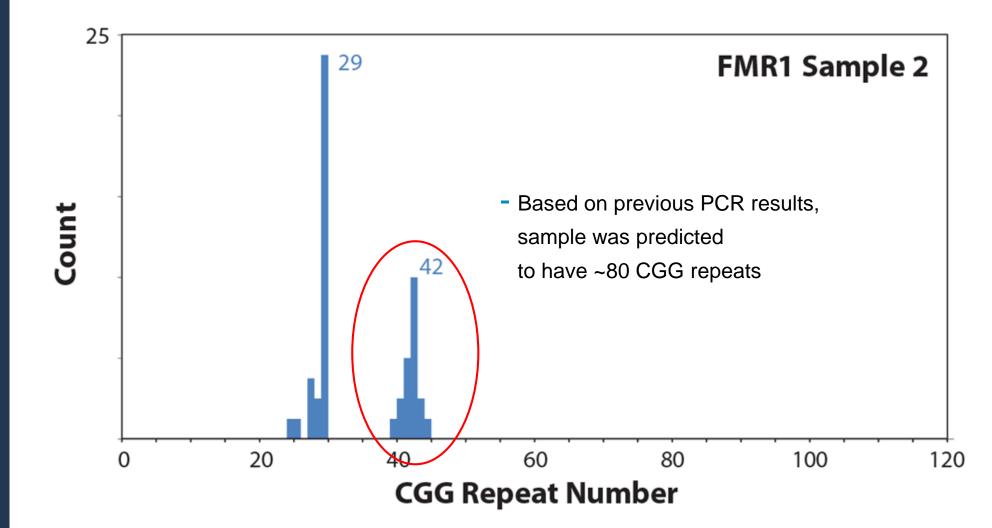
Multiplexing:

- Multiple regions can be targeted in the same reaction
- Patient samples could be barcoded during initial SMRTbell library preparation

46	0	470	480) 49	0 500	510	520	53	0 54	0 55	0 56	0 57	0 58	0 59	90	600
					CGGCGGCGGC											
61		620	630					68							40	750
CGGCGGCGGC 76		3GC GCG 770	780		OCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC										g geggeged 90	900
					CGGCGCGGCG											
91	-	920	930		0 950										040	1050
					CGGCGGCGGC											
10		1070	108		90 110										190	1200
					CGGCGGCGGC											
12		1220	123													1350
					CGCGGCGGCG											
13		1370	138												490	1500
CGGCGCGGGGG 15		1520	153		GCGGCGCGGC 40 155										GGGCGGCC	3GC 1650
					GGCGGCCGCG											
		1670	168		90 170										790 	1800
			GCGCGGC	GGCGGCGGCG	CGGCGGCGGC	GGCGCGCGCG	GGCGCGCGGC	GGCGCGGCGG	CGGGCGCGGC	GGCGGCGCGG	CGGCGGCGGG					
18		1820	183												940	1950
					CGCGGCGGCG											
19		1970	198												090	2100
					GCGGCGGCGG											
21		2120	213												240	2250
	60	2270	228		GCGGCGGCGG 90 230										390	2400
GCGGCGGGGCG 24		2420	CGGCGGC 243		40 245										G GCGCGCGC 540	2550
					40 24:											
GCGGCGGGCG	CGGCGGC	GC GCC	GCGCGCG	GCGGCGGCGG	CGGCGGCGGC	GCGGCGGCGC	GGCGCGGCGG	CGGCGGCGGC	GGCGGCGGCG	CGCGGCGCGG	CGGCCGGCGG	CTGGGCTCGA	GCGCCCGCAG	CCCACCTCTC	GGGGCGGG	CTC
25		2570	258		90 260										690	2700
					GTGAAGTGCG											
	10	2720	273		40 275											
					CGGCGGGATG											
					COCOUNTO		53.5150/1010									

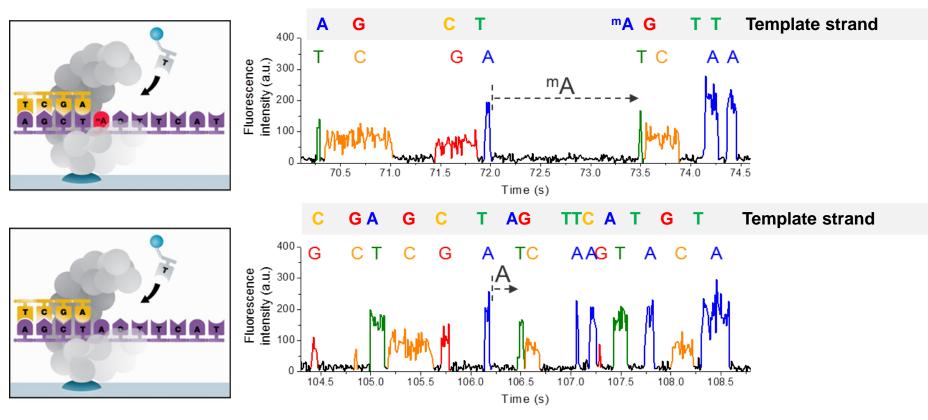
... CGG CGG CGG CGG <mark>AGG</mark> CGG CGG CGG CGG CGG CGG CGG CGG <mark>AGG</mark> CGG ...

Repeat counting of *FMR1* gene



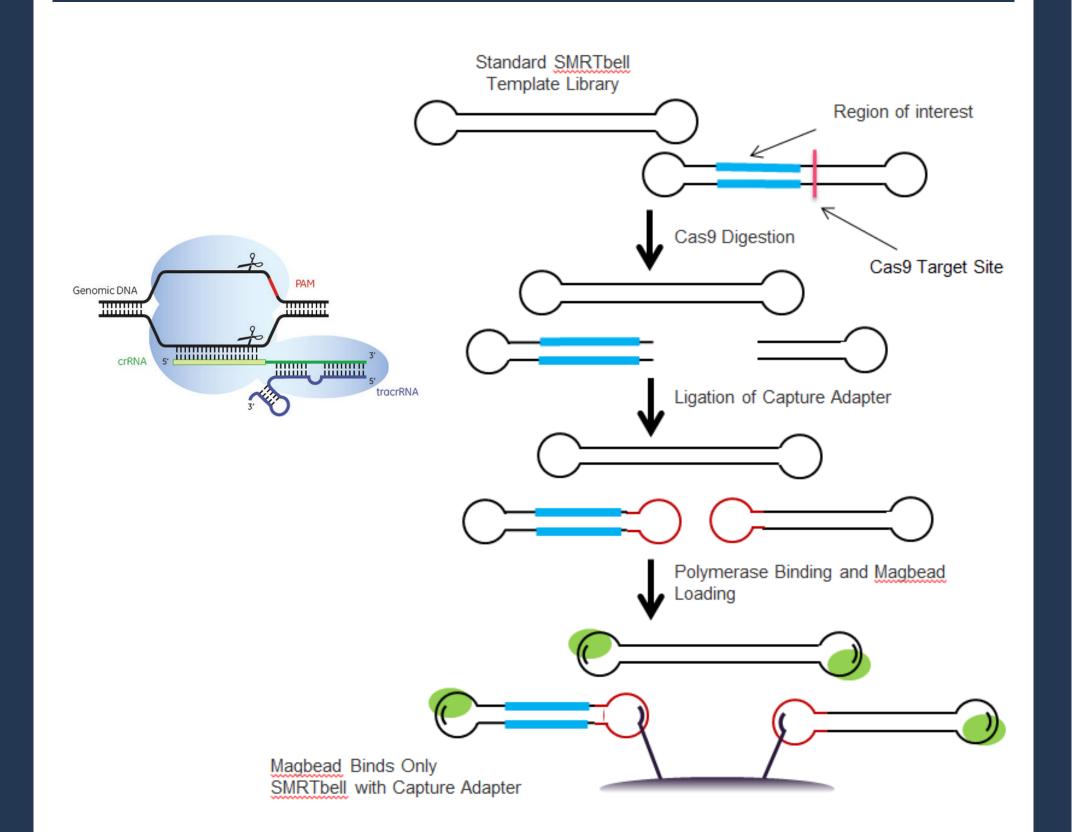
The mutant allele in this sample has a ~110 bp duplicated region preceding the CGG repeats.

Normal allele (29 CGG repeats) Mutated allele (~42 CGG repeats)



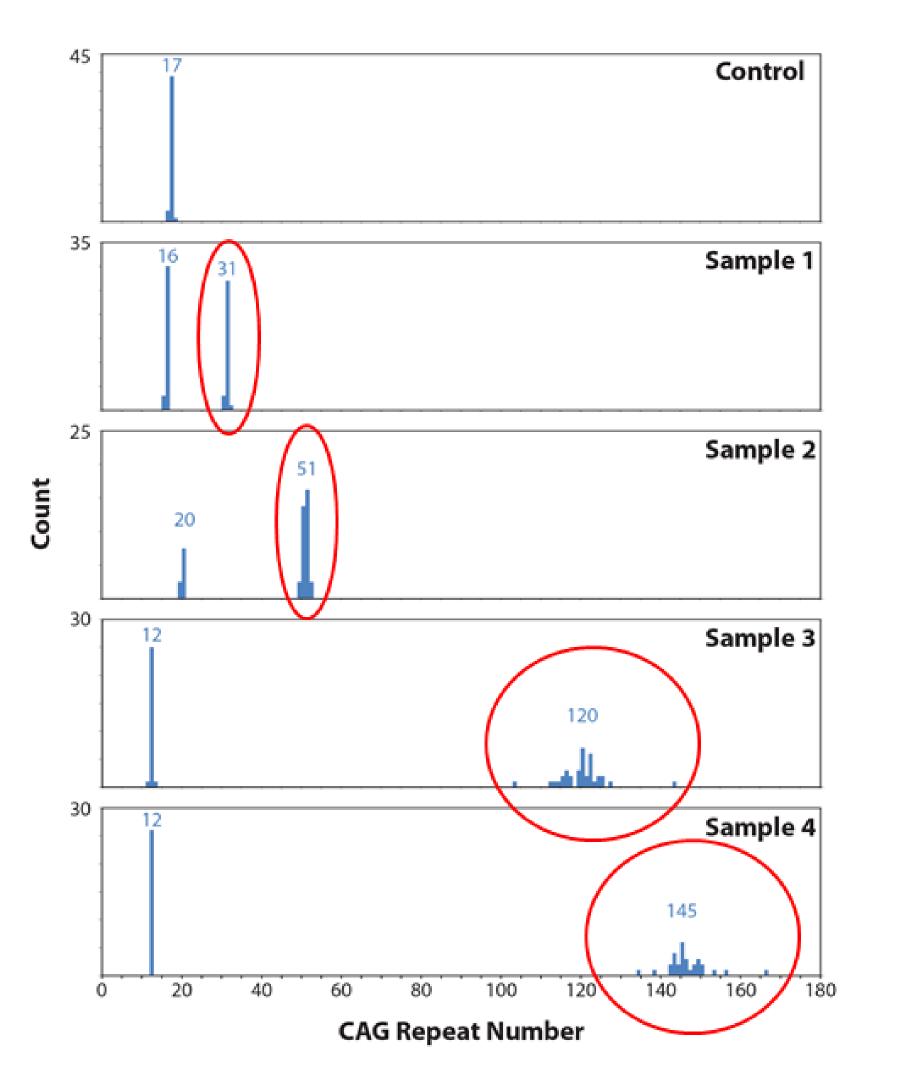
Example: N⁶-methyladenine

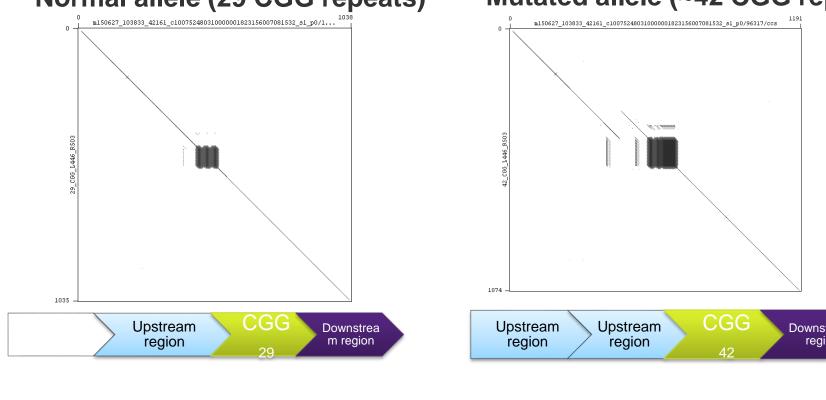
CRISPR/Cas9 system for targeted amplification-free enrichment



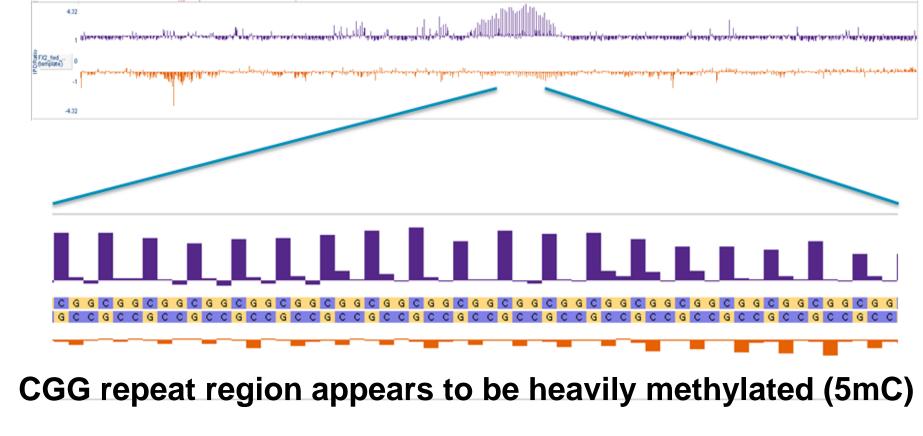
Repeat counting

The *Huntingtin* gene CAG repeat counts in HD patients





Direct Methylation Detection of FMR1 gene



Full Landscape of *FMR1* mRNA Isoforms

Pretto et al.¹⁾ provided the first comprehensive map of FMR1 PBMCs Brain Fibroblast

1. Prepare a standard SMRTbell library

- **2.** Use guide RNA to target the region of interest
- **3.** Digest the SMRTbell template using Cas9
- **4.** Ligate a capture adapter to the
- SMRTbell template that can be used to select for the SMRTbell
- templates that contain the region of interest
- **5**. Sequence the SMRTbell templates

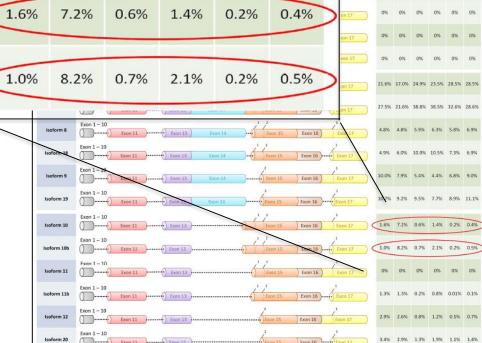
- CRISPR/Cas9 captures both the normal and the mutated allele successfully
- Widening repeat number distribution at the mutated allele is a known biological mechanism
- Obtained roughly equal number of sequenced molecules for normal and mutated alleles

RNA toxicity may arise from a relative increase of all FMR1 mRNA isoforms. However 2 isoforms were identified with differential expression in premutation carriers suggesting a functional relevance of these in the pathology of FMR1associated disorders

Identified novel alternative

gene isoforms

splicing event



References and Acknowledgements

¹⁾ Pretto DI et al. (2015) Differential increases of specific FMR1 mRNA isoforms in premutation carriers. J Med Genet.52(1):42-52.

Samples provided by Vanessa Wheeler (Harvard Medical School), Paul Hagerman (UC Davis)

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2016 by Pacific Biosciences, the Pacific Biosci BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx. All other trademarks are the sole property of their respective owners.