Supplemental Information

Supplemental figures

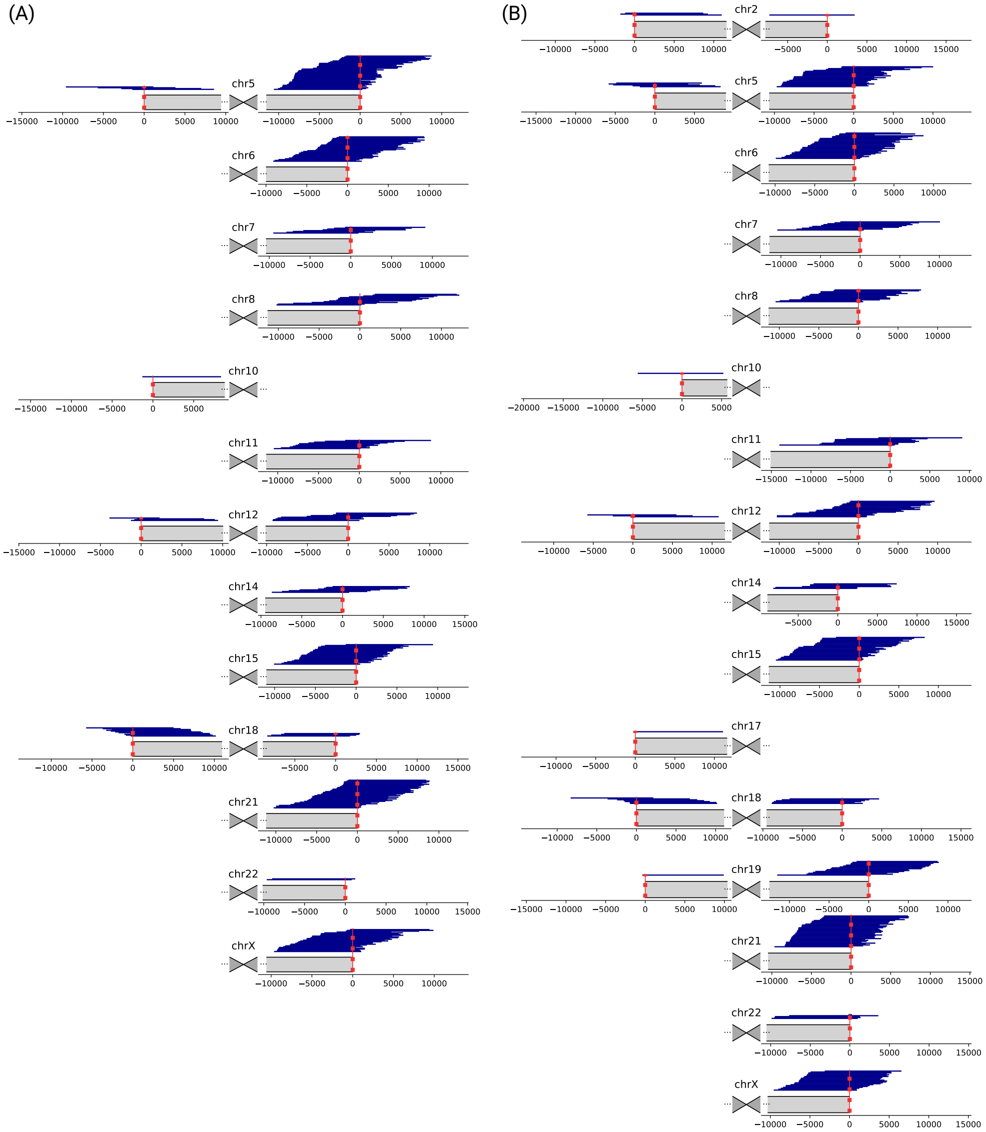

Figure S1: Mapping of candidate telomeric PacBio CCS reads from datasets (A) HG001 and (B) HG005. Chromosomes are displayed schematically, centered around the centromere, with only the arms shown to which candidate reads aligned. Vertical red dashed lines denote the position of the boundary of the annotated telomeric tract. Coordinates are given in bp, relative to the positions of the telomeric tract boundaries. Relates to: **Figure 1**.

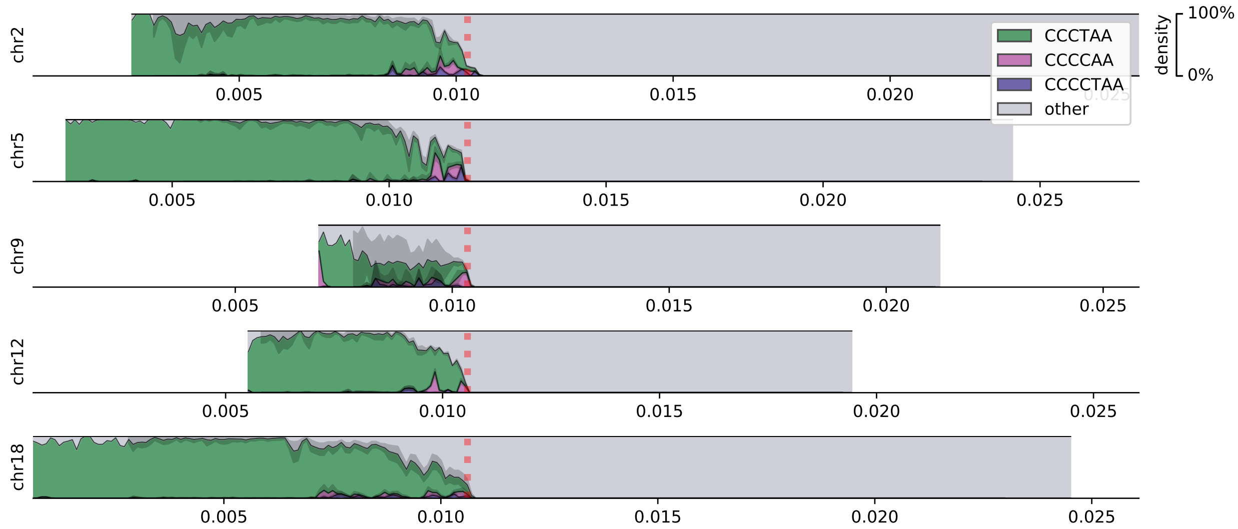

Figure S2: Densities of top three enriched motifs (contributing to at least 0.5% of the repeat content) at ends of chromosomal *p* arms of the HG002 dataset. Only the arms covered by at least 20 reads are displayed. Genomic coordinates are given in Mbp. Vertical red dashed lines denote the position of the boundary of the annotated telomeric tract. Relates to: **Figure 2**, **Table 1**.

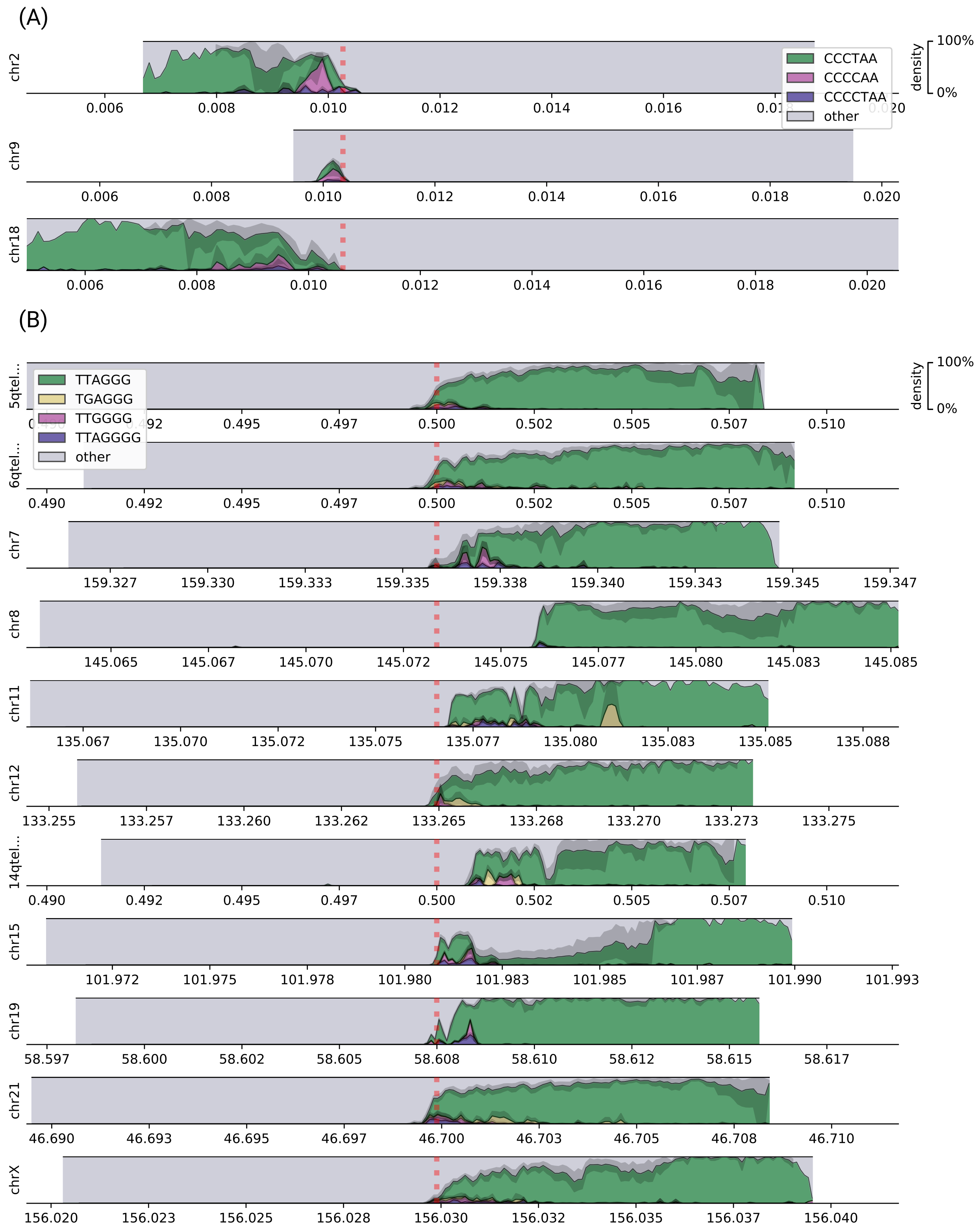

Figure S3: Motif densities at ends of chromosomal (A) *p* and (B) *q* arms of the HG001 dataset. Only the arms covered by at least 20 reads are displayed. Genomic coordinates are given in Mbp. Relates to: **Figure 2**, **Table 1**.

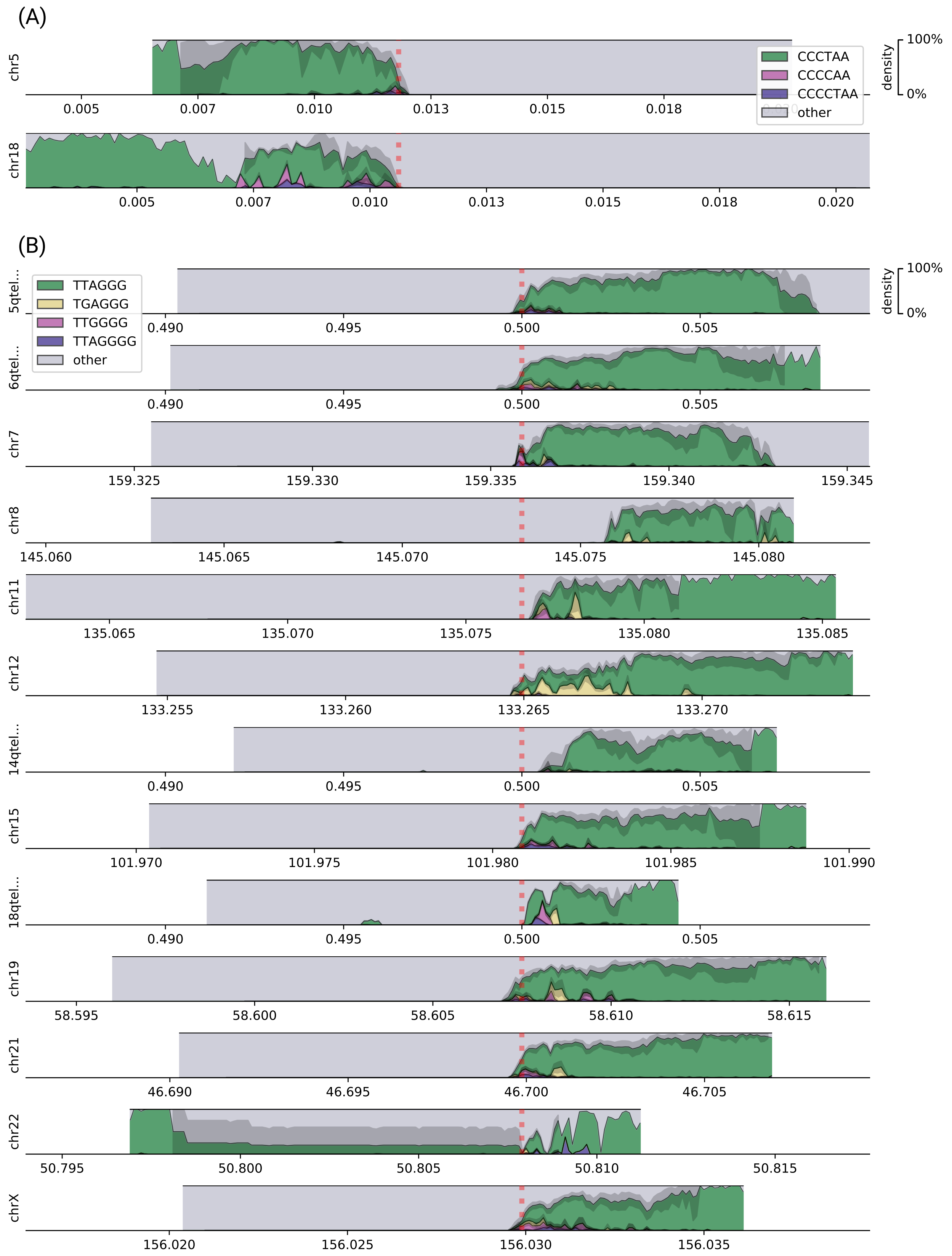

Figure S4: Motif densities at ends of chromosomal (A) *p* and (B) *q* arms of the HG005 dataset. Only the arms covered by at least 20 reads are displayed. Genomic coordinates are given in Mbp. Relates to: **Figure 2**, **Table 1**.

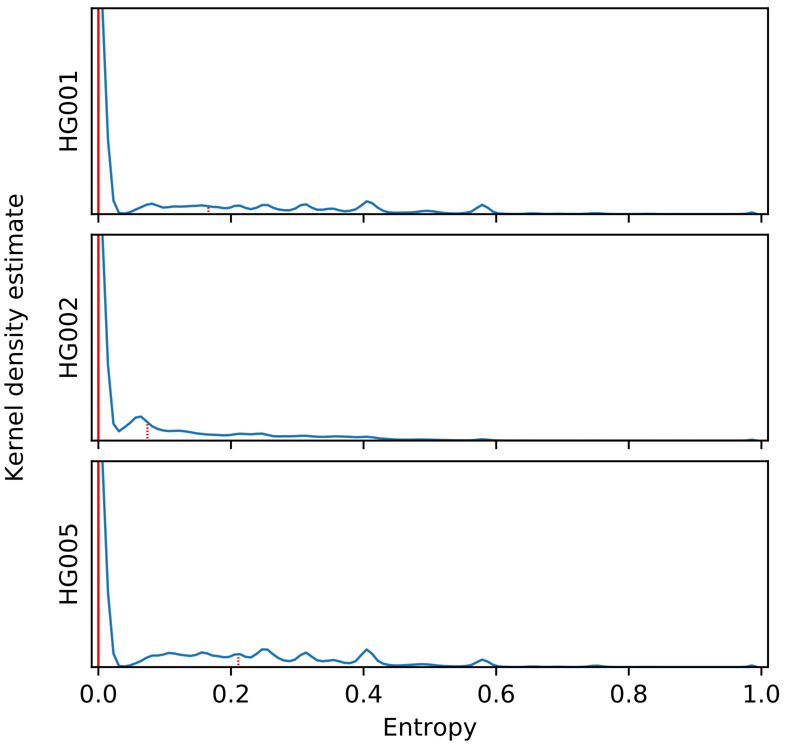

Figure S5: Distribution of motif entropies in 10 bp windows of candidate PacBio CCS reads aligning to the same chromosomal arms in GIAB datasets HG001, HG002, and HG005. Red solid lines denote the position of the median (0.000 in all three datasets), and red dashed lines denote the 3rd quartile (0.166, 0.074, and 0.211, respectively). Relates to: **STAR Methods, Evaluation of sequence concordance in telomeric long reads**.

Supplemental tables

Chromosome	Reference contig	Arm	HG001	HG002	HG005
chr2	2qtel_1-500K_1_12_12_rc	q	0	0	1
chr2	chr2	p	5	16	3
chr5	5qtel_1-500K_1_12_12_rc	q	42	53	23
chr5	chr5	p	4	15	5
chr6	6qtel_1-500K_1_12_12_rc	q	31	49	29
chr7	chr7	q	8	32	10
chr8	chr8	q	14	35	14
chr9	chr9	p	6	6	0
chr10	10qtel_1-500K_1_12_12_rc	q	0	1	0
chr10	chr10	p	1	2	1
chr11	chr11	q	11	31	9
chr12	chr12	q	10	27	18
chr12	chr12	p	4	5	3
chr14	14qtel_1-500K_1_12_12_rc	q	8	26	6
chr15	chr15	q	25	21	26
chr16	16qtel_1-500K_1_12_12_rc	q	0	2	0
chr16	chr16	p	1	0	0
chr17	17qtel_1-500K_1_12_12v2_rc	q	0	4	0
chr17	17ptel_1_500K_1_12_12	p	0	1	1
chr18	18qtel_1-500K_1_12_12_rc	q	4	26	6
chr18	chr18	p	11	35	7
chr19	19ptel_1-500K_1_12_12	p	0	1	1
chr19	chr19	q	6	0	16
chr21	chr21	q	35	77	35
chr22	chr22	q	2	51	5
chrX	chrX	q	28	54	22

Table S1: The number of telomeric reads on each arm identified in GIAB PacBio CCS datasets HG001, HG002, and HG005. Relates to: **Figure 1**, **Figure S1**.

Motif	Illumina datasets		10X datasets	
	Median abundance	Adjusted p-value	Median abundance	Adjusted p-value
TTAGGG	0.299068	0.00e+0	0.461711	0.00e+0
TGAGGG	0.007484	0.00e+0	0.018524	0.00e+0
TTGGGG	0.002495	0.00e+0	0.007190	0.00e+0
GGGG	0.020347	0.00e+0	0.006080	0.00e+0
TTAGGGG	0.003007	0.00e+0	0.005024	0.00e+0
TTTT	0.001294	0.00e+0	0.001490	0.00e+0
TTAAGGG	0.000664	1.39e-55	0.001124	1.58e-59
TTAGGGGTTAGGG	0.000533	1.04e-51	0.001020	0.00e+0
TAGGG	0.000619	0.00e+0	0.001020	0.00e+0
TTGGG	0.000500	0.00e+0	0.000989	0.00e+0
TTTAGGG	0.000622	6.40e-55	0.000884	1.02e-57
TAGGGTTAGGG	0.000312	4.24e-40	0.000503	0.00e+0
TTAGGGTTTAGGG	0.000176	4.41e-38	0.000284	6.22e-59
TTAGGGTTAAGGG	0.000145	6.63e-36	0.000264	4.15e-57
TTAGG	0.000241	8.13e-35	0.000213	1.10e-55
TTGGGTTAGGG	0.000127	4.47e-28	0.000178	3.34e-56
TTAGGGTTAGG	0.000066	1.99e-18	0.000092	7.82e-48
TTAGGGGG	0.000039	1.02e-14	0.000062	4.31e-40
TTAGGGTTGTTAGGG	0.000035	4.64e-09	0.000061	4.65e-57
TTAGAGGG	0.000036	5.44e-13	0.000053	2.66e-36
TTGGGGTTGGGGG	0.000002	4.51e-13	0.000014	5.84e-21
TTAGGGTGGTTAGGG	0.000007	5.39e-06	0.000013	5.42e-38

Table S2: Significantly enriched repeating motifs in telomeric candidate reads in short-read sequencing experiments, subset to motifs also observed in PacBio telomeric reads, with respect to reverse-complement equivalence. Relates to: **STAR Methods, Identification of repeat content**.

Chromosome	Haplotype	PacBio_CCS_10kb	PacBio_CCS_15kb
5qtel_1-500K_1_12_12_rc	1	11	24
5qtel_1-500K_1_12_12_rc	2	7	10
6qtel_1-500K_1_12_12_rc	1	8	12
6qtel_1-500K_1_12_12_rc	2	18	10
chr7	1	9	8
chr7	2	7	8
chr8	1	8	6
chr8	2	9	8
chr11	1	5	11
chr11	2	8	7
chr12	1	9	9
chr12	2	6	3
14qtel_1-500K_1_12_12_rc	1	3	8
14qtel_1-500K_1_12_12_rc	2	5	10
chr15	1	6	0
chr15	2	4	11
18qtel_1-500K_1_12_12_rc	1	4	9
18qtel_1-500K_1_12_12_rc	2	4	9
chr21	1	16	20
chr21	2	12	29
chr22	1	2	27
chr22	2	11	10
chrX	1	12	13
chrX	2	10	19

Table S3: Amounts of reads from the two HG002 PacBio CCS sequencing experiments contributing to each telomeric haplotype on the *q* arms. Relates to: **Figure 3**.