Supplementary Information

Mapping G-Quadruplex structures in the genome

Enid Yi Ni Lam[1], Dario Beraldi[1], David Tannahill[1] and Shankar Balasubramanian[1,2,3]*

Author Affiliations:

[1] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, UK

[2] The University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge, UK

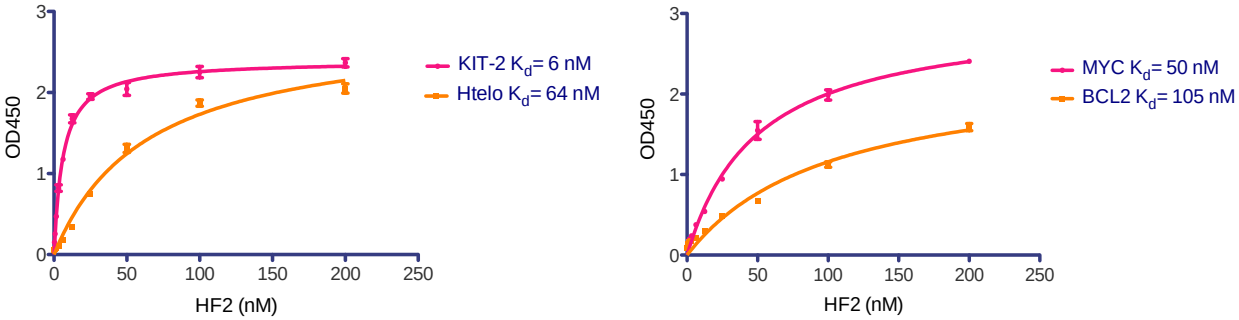
[3] School of Clinical Medicine, University of Cambridge, Cambridge, UK

* To whom correspondence should be addressed.

Email: sb10031@cam.ac.uk

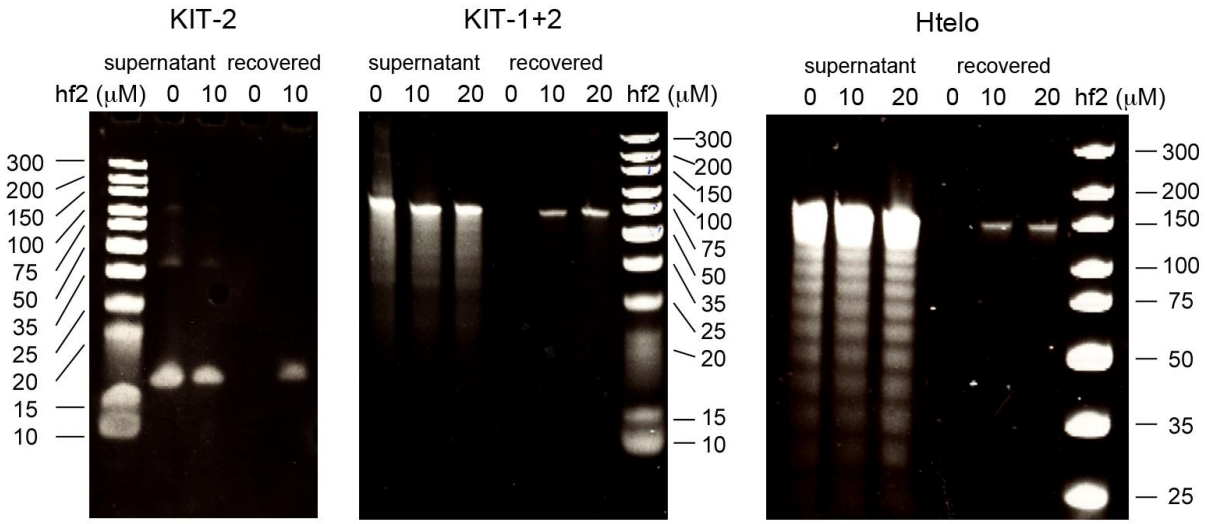
Supplementary Results

Supplementary Figure S1 The hf2 single chain antibody binds G-quadruplex-structured oligonucleotides.



Binding curves as determined by ELISA for hf2 and different G-quadruplex oligonucleotides folded into a G-quadruplex structure. The dissociation constants (K_d) are indicated.

Supplementary Figure S2 Pull-down of a variety of G-quadruplex-structured oligonucleotides by the hf2 single chain antibody.

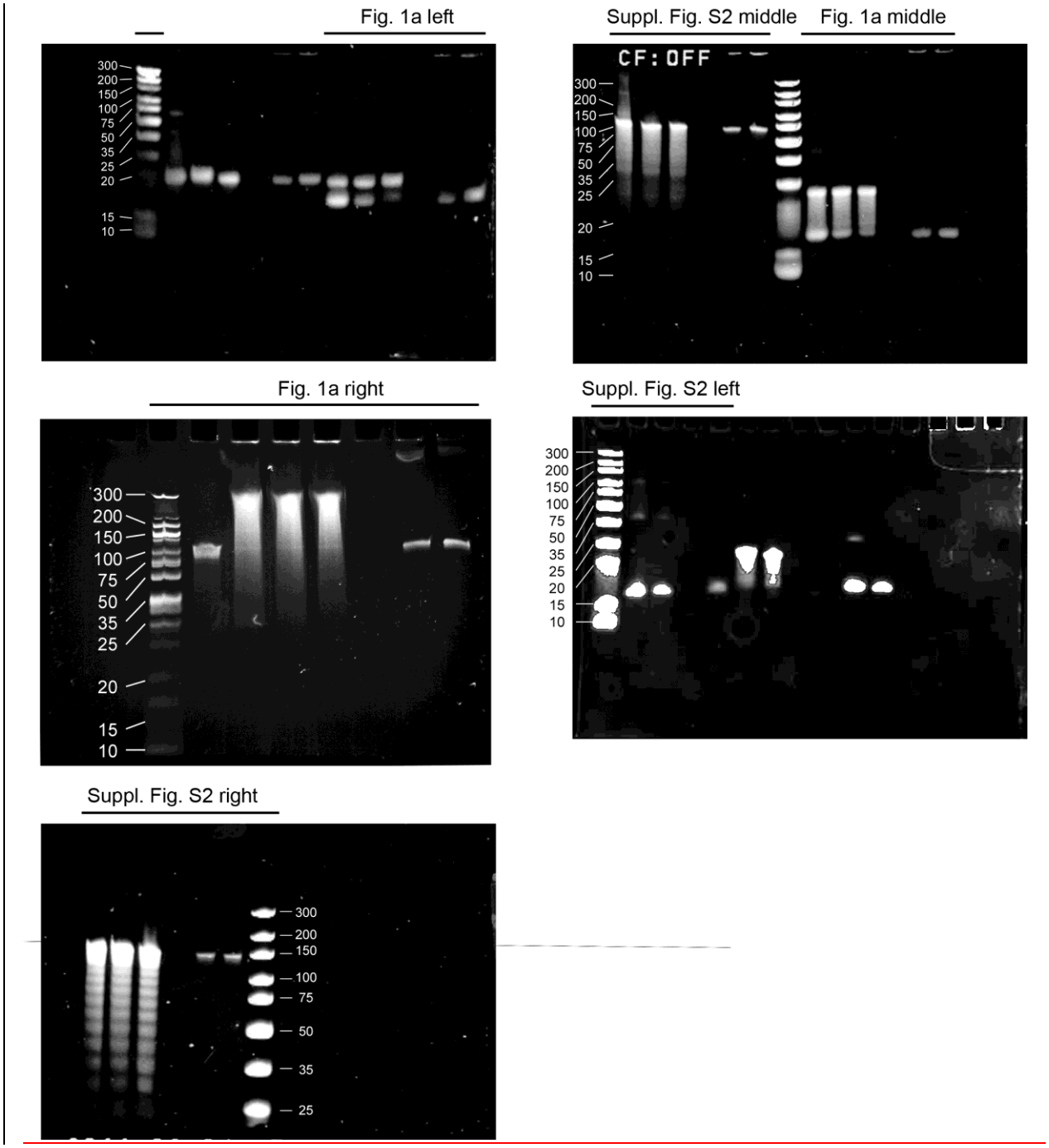


DNA recovered after binding to hf2 and unbound DNA (supernatant) analyzed on TBE-urea gels. The oligonucleotides KIT-2 (left), KIT-1+2 (middle) and Htelo (right) are pulled down by hf2. Molecular weight size markers are indicated.

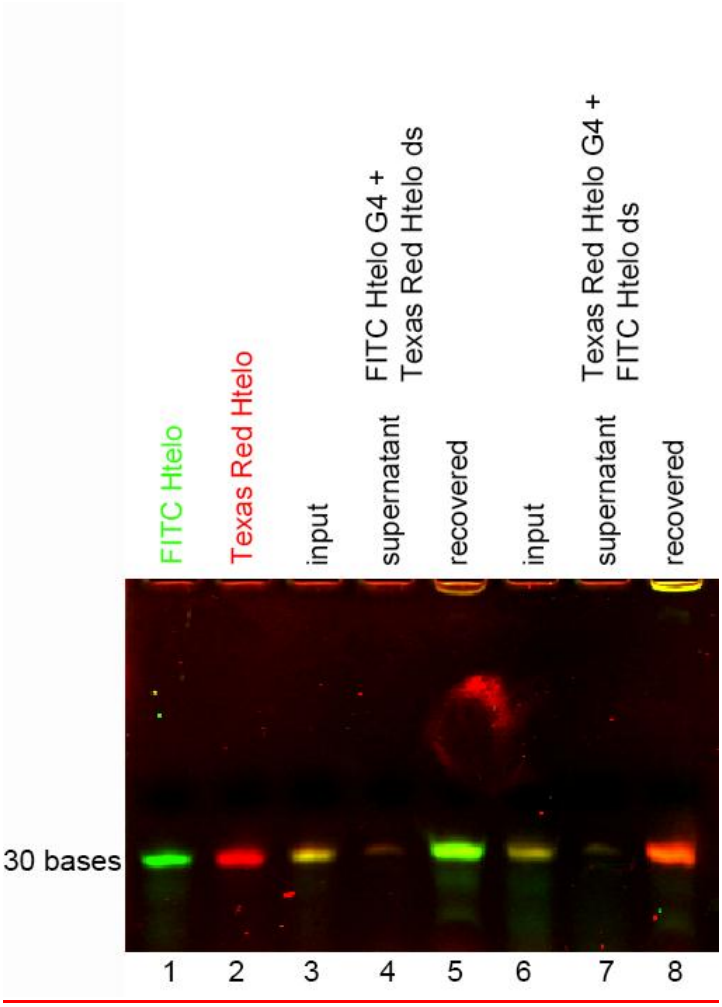
Supplementary Figure S3 Full images of gels shown in Figure 1 and Supplementary Figure S2.

The portion of gels displayed in Figure 1 and Supplementary Figure S2 are indicated by the black line.

Molecular weight size markers are indicated.



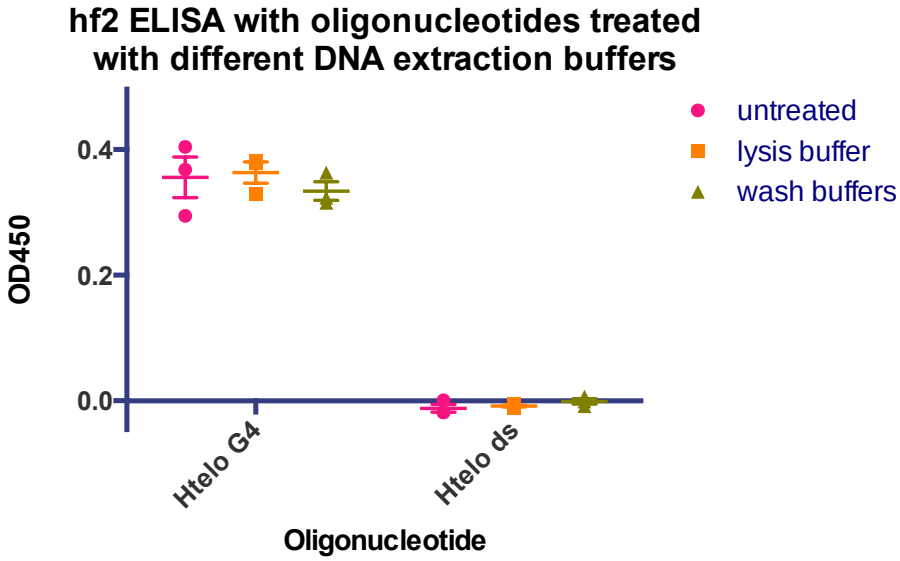
Supplementary Figure S4 Sonication does not induce or destroy G-quadruplex structures.



Mixtures of pre-folded G-quadruplex-structured (G4), and double-stranded (ds) fluorescent oligonucleotides were sonicated, prior to pull-down with hf2. Analysis by urea-PAGE confirms that the sonication procedure does not affect the pre-folded G-quadruplex oligonucleotide structure nor induce G-quadruplex formation from the double-stranded oligonucleotide.

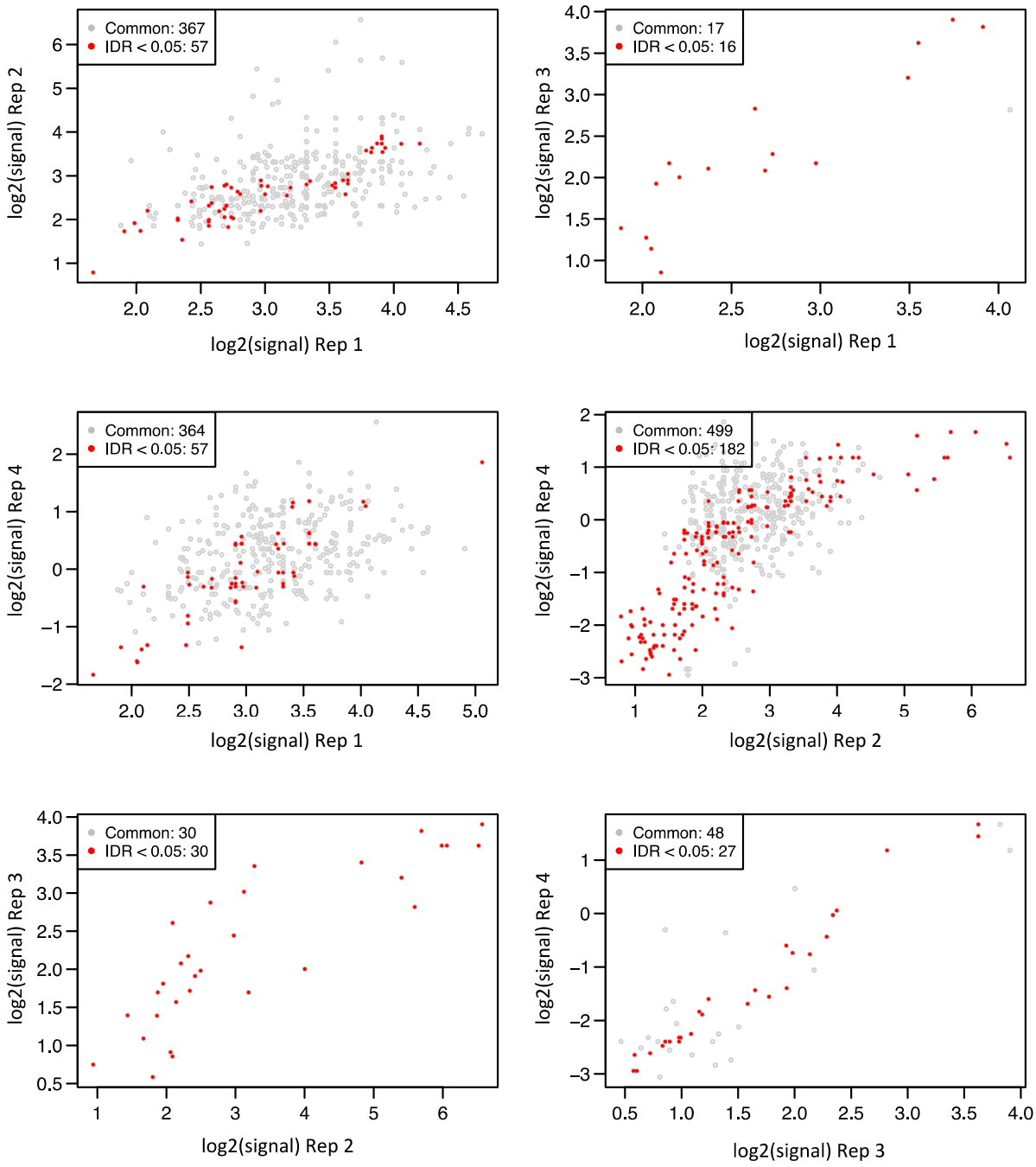
Lane 1, FITC Htelo oligonucleotide folded into a G4 structure alone, size control.
Lane 2, Texas Red Htelo oligonucleotide folded into a G4 structure alone, size control.
Lanes 3-5, Hf2 pull-down of a mixture of pre-folded FITC Htelo G4 oligonuecleotide and pre-folded Texas Red Htelo ds oligonucleotide after sonication- only the FITC Htelo oligo pre-folded into a G-quadruplex is recovered, while the Texas Red oligonucleotide pre-folded into ds is not recovered.
Lanes 6-8, Pull-down of a mixture of Texas Red Htelo G4 oligonucleotide and FITC Htelo ds oligonucleotide after sonication- only the Texas Red Htelo oligonucleotide pre-folded into a G-quadruplex is recovered, but not the FITC Htelo pre-folded ds oligonucleotide.

Supplementary Figure S5 Qiagen DNA extraction buffers do not destroy or induce G-quadruplex structures.



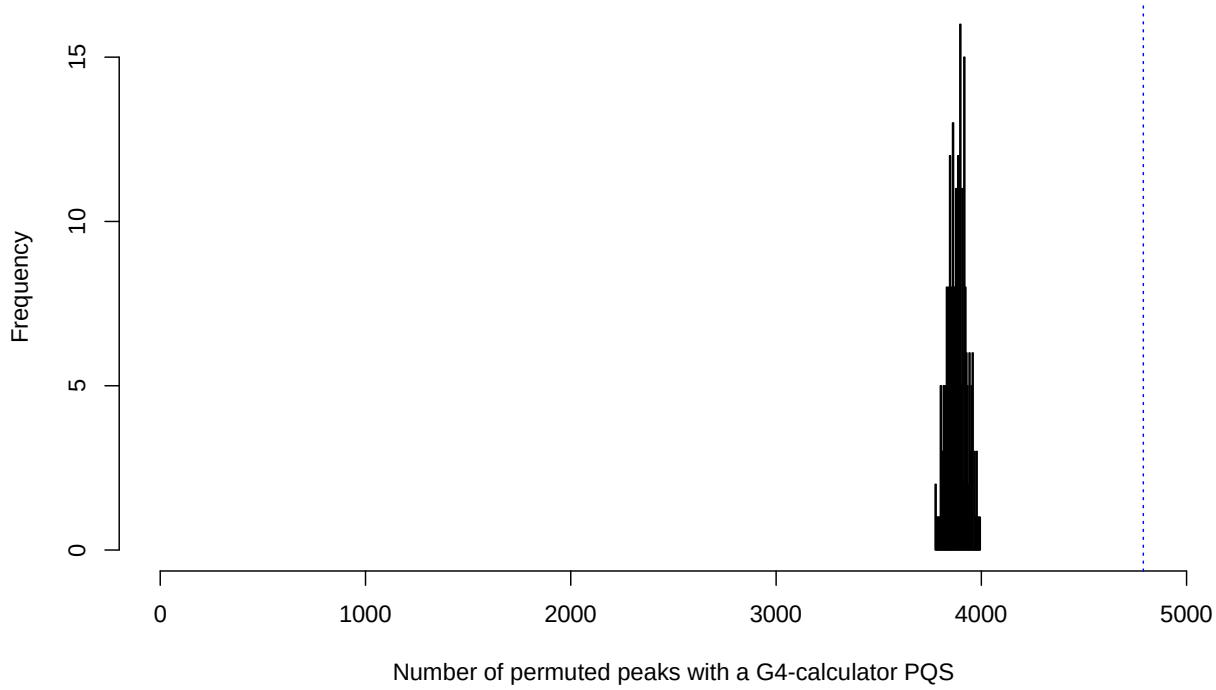
ELISA with biotinylated Htelo oligonucleotides either annealed in K+ buffer into G-quadruplex structure or annealed with the complementary strand in buffer without K+ into duplex structure. After binding of oligonucleotides to streptavidin-coated plates, the wells were treated in triplicate with either the lysis buffer (Qiagen buffer AL) for 10 minutes, or the two wash buffers (AW1 and AW2) sequentially for 5 minutes each. Three wells were left untreated as a control. After treatment with the buffers, an ELISA was performed with hf2 at its determined K_d concentration (64nM). The mean and standard error are plotted. Treatment of the Htelo G-quadruplex oligonucleotides with buffers used during genomic DNA isolation does not disrupt the G-quadruplex structure, and does not induce its formation from duplex DNA.

Supplementary Figure S6 Pair-wise comparison of the peaks called between the four libraries to assess their consistency between replicates.



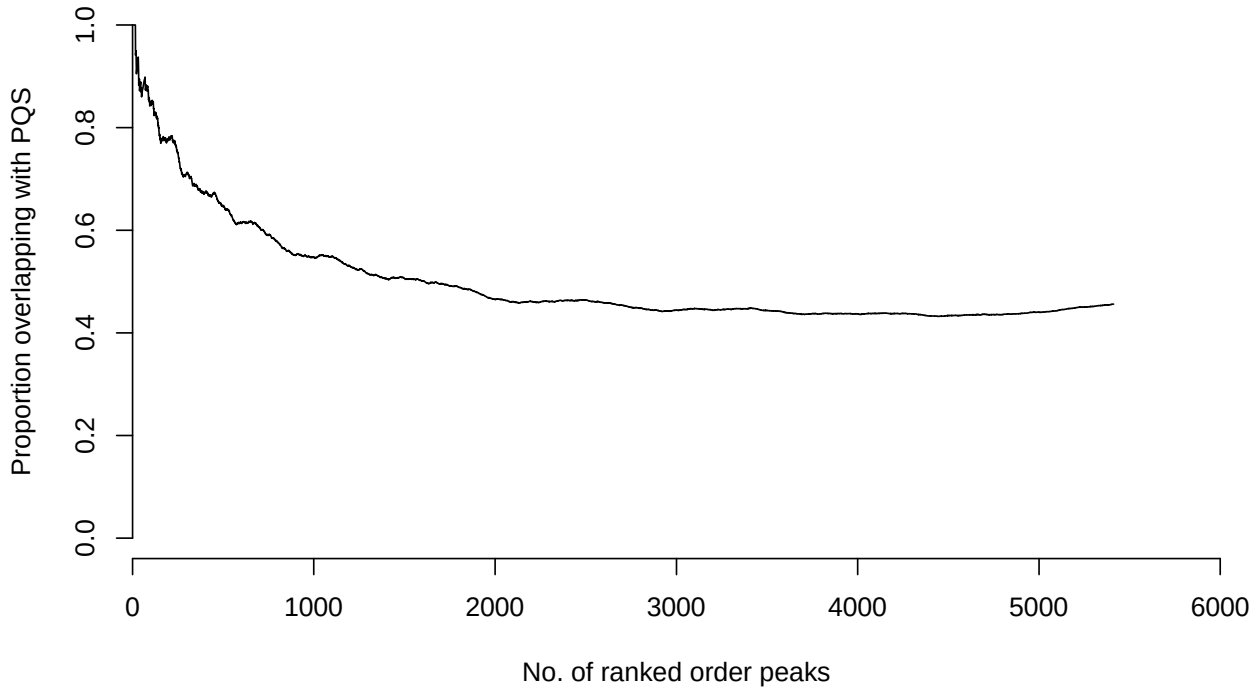
The enrichment of the peaks in common between each pair of libraries were plotted and the peaks with an irreproducible discovery rate (IDR) of 0.05 (IDR [56]) are highlighted in red, while peaks with an IDR of greater than 0.05 are plotted in grey.

Supplementary Figure S7 Statistical significant enrichment of G4-calculator PQS in the hf2 pull-down peaks.



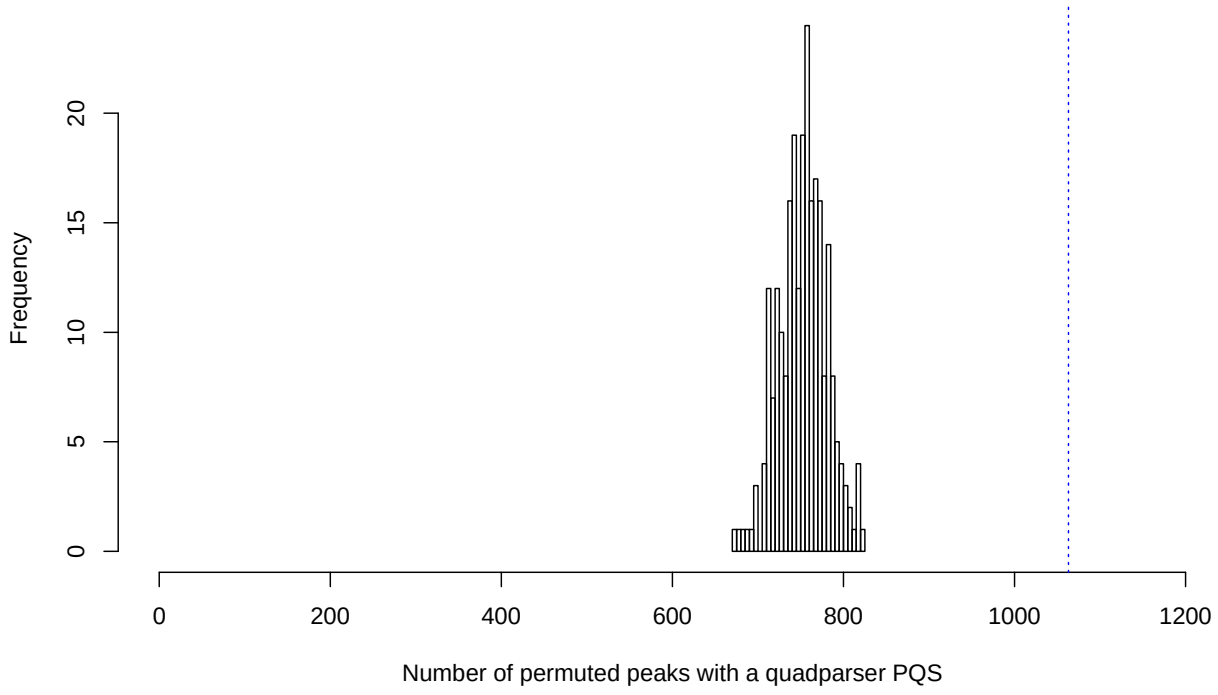
The histogram shows the number of peaks containing a G-quadruplex motif in random sequences compared with the 4789 observed number of pull-down peaks containing a G-quadruplex (blue line). Simulation was used to estimate the number of peaks containing a PQS by chance, that is, if peaks were unrelated to G-quadruplexes. For 250 times, the peaks on each chromosome were shuffled (thus maintaining the number of peaks and their size on each chromosome unchanged) and the number of peaks containing a G-quadruplexes computed. On average, 3885 peaks had a PQS (sd= 44.6) therefore the observed number of G-quadruplex-containing peaks is significantly larger (by 20.3 standard deviations) than expected by chance.

Supplementary Figure S8 Prevalence of G4 calculator predicted G-quadruplex motifs compared to peak rank order.



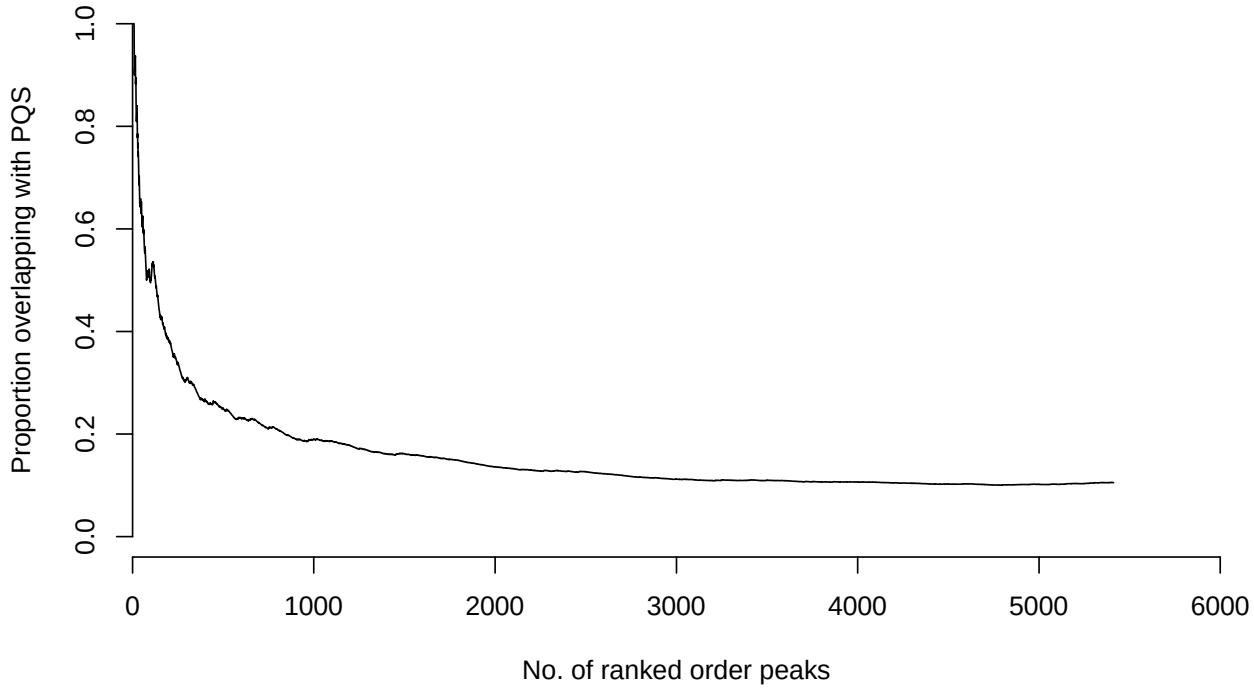
After normalizing for sequencing depth, reads for all libraries were combined and the resulting peaks ranked by enrichment over input were analyzed with G4 calculator. The proportion of the peaks overlapping with a G4 calculator predicted PQS is plotted for increasing numbers of ranked order peaks. This shows that PQS are most prevalent in approximately the top 500 enriched peaks.

Supplementary Figure S9 Statistical significant enrichment of G-quadruplex motifs in the hf2 pull-down peaks.



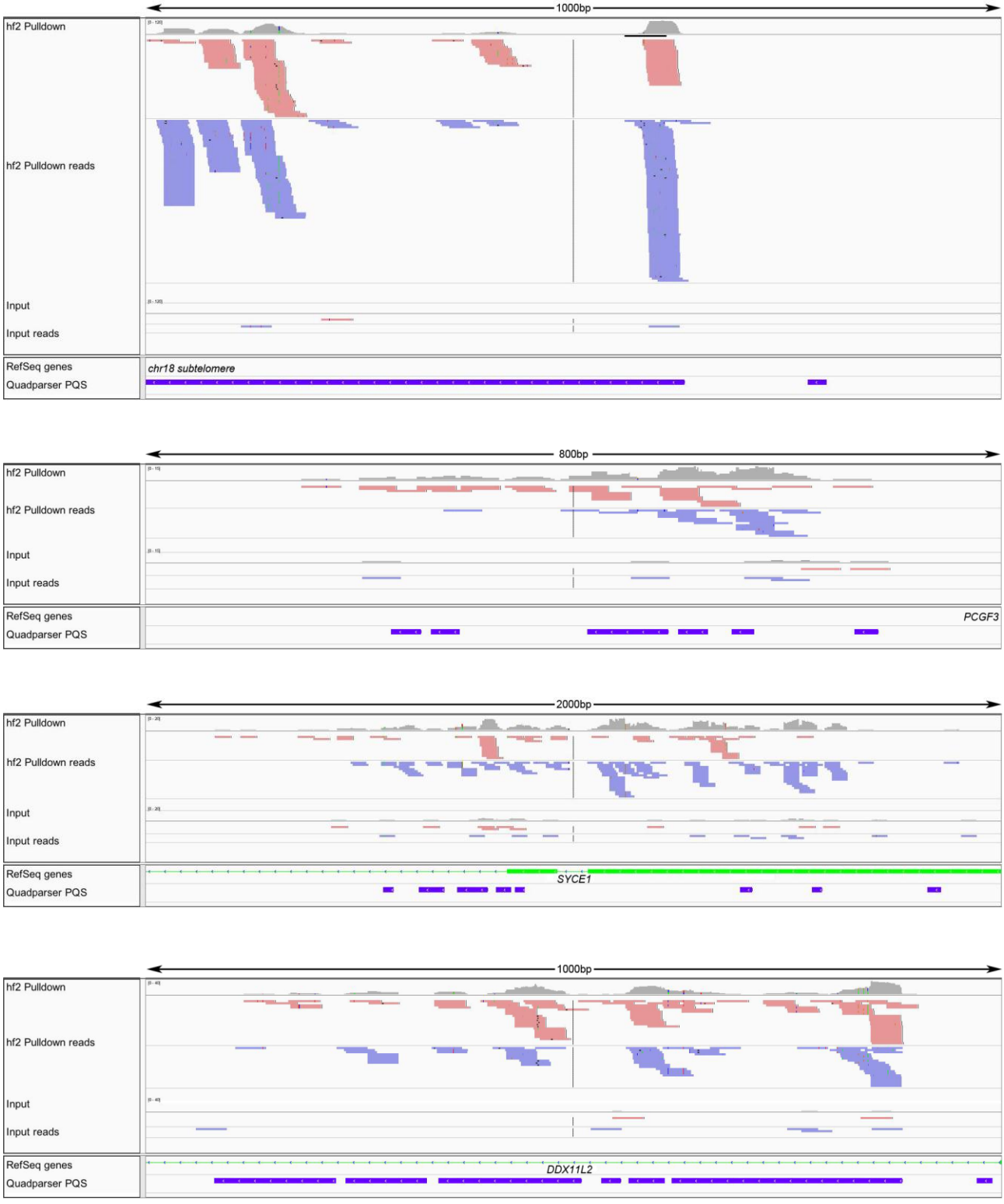
The histogram shows the number of peaks overlapping with a G-quadruplex motif in random sequences compared with the 1063 observed number of pull-down peaks containing a G-quadruplex (blue line). Simulation was used to estimate the number of peaks containing a PQS by chance, that is, if peaks were unrelated to G-quadruplexes. For 500 times, the peaks on each chromosome were shuffled (thus maintaining the number of peaks and their size on each chromosome unchanged) and the number of peaks overlapping a G-quadruplexes computed. On average, 755 peaks had a PQS (sd=27.6) therefore the observed number of G-quadruplex-containing peaks is significantly larger (by 11.1 standard deviations) than expected by chance.

Supplementary Figure S10 Prevalence of G-quadruplex motifs compared to peak rank order.



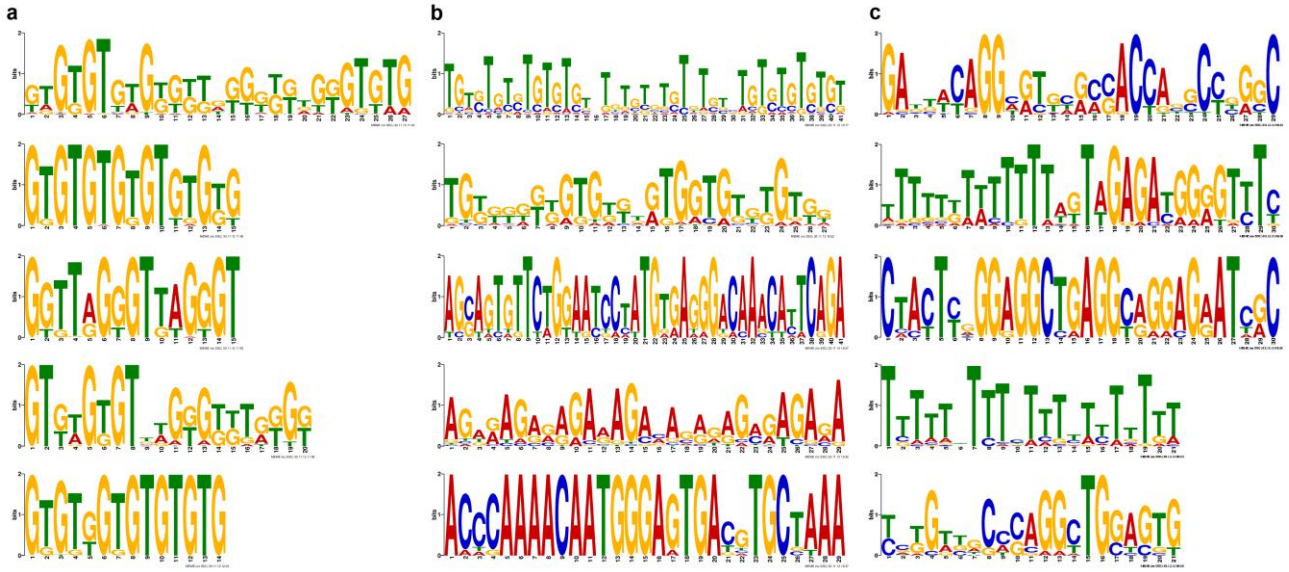
After normalizing for sequencing depth, reads for all libraries were combined and the resulting peaks ranked by enrichment over input were analyzed by *quadparser* for overlap with PQS. The proportion of the peaks overlapping with a PQS is plotted for increasing numbers of ranked order peaks. This shows that PQS are most prevalent in the top 200-300 enriched peaks.

Supplementary Figure S11 High resolution view of the Peaks identified by hf2 pull-down in **Figure 2**.



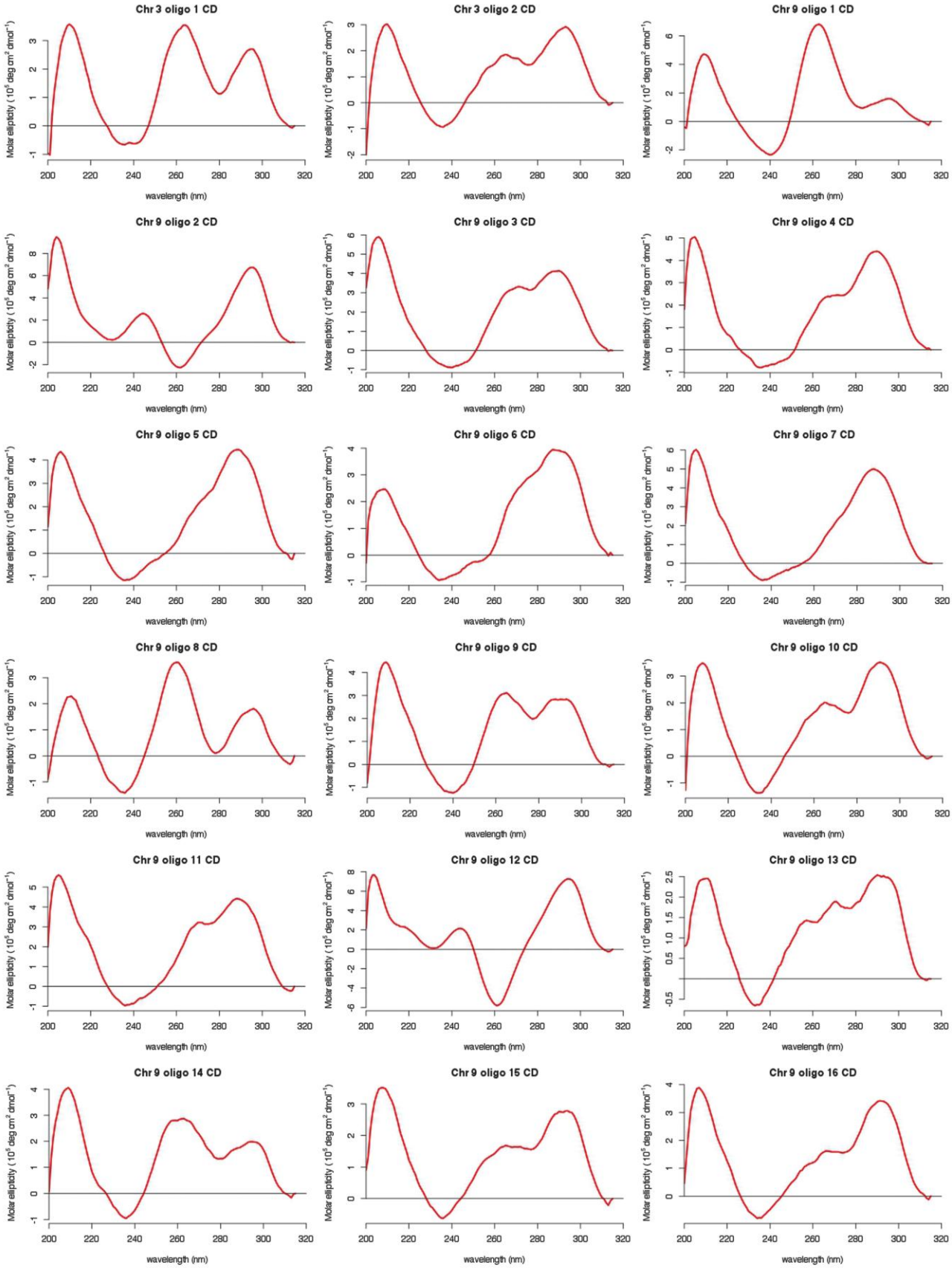
Genome browser view of the four peaks (grey) showing the reads mapping to the two strands separately in red and blue, respectively. RefSeq genes and quadparser PQS are shown in green and purple, respectively.

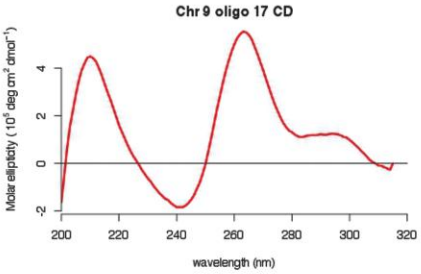
Supplementary Figure S12 Motif analysis of the enriched peaks by MEME.

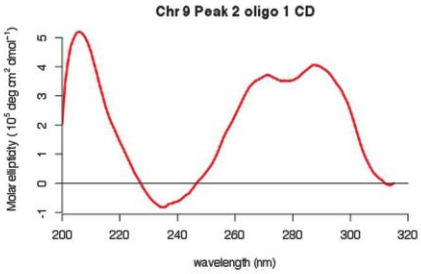


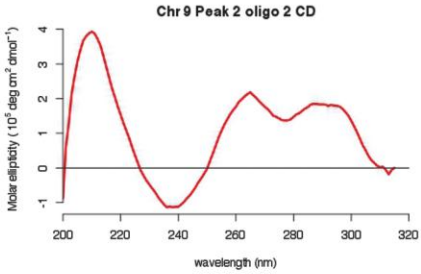
(a) The top five most enriched motifs by MEME of the 200 peaks with the highest enrichment over input in the combined hf2 pull-down library. (b) The top five most enriched motifs by MEME of the 200 peaks with the highest enrichment called in the input library. (c) The top five most enriched motifs by MEME of 200 sequences randomly selected from the genome. The G-rich strand is shown for all motifs to allow comparison.

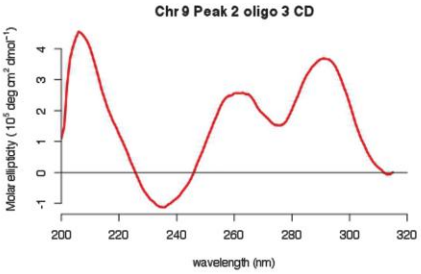
Supplementary Figure S13 Circular dichroism spectra of 44 oligonucleotides containing sequences from four peaks indicated above and folded into a G-quadruplex in the presence of K^+.

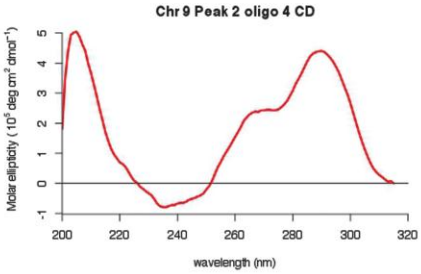


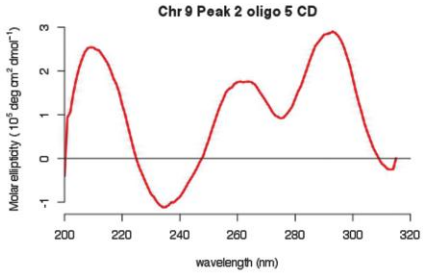

Chr 9 oligo 17 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

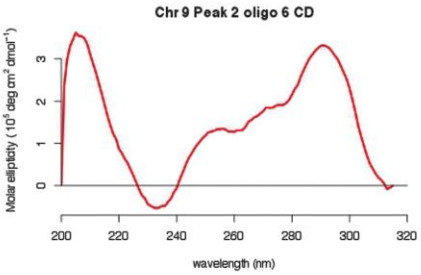

Chr 9 Peak 2 oligo 1 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

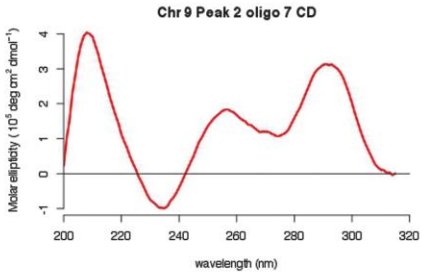

Chr 9 Peak 2 oligo 2 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

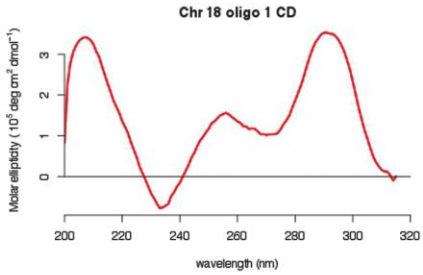

Chr 9 Peak 2 oligo 3 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

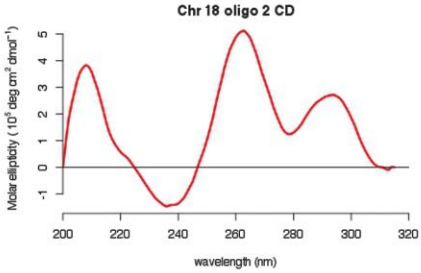

Chr 9 Peak 2 oligo 4 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

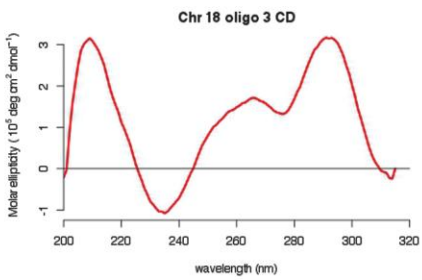

Chr 9 Peak 2 oligo 5 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

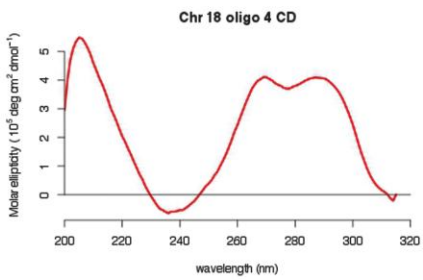

Chr 9 Peak 2 oligo 6 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

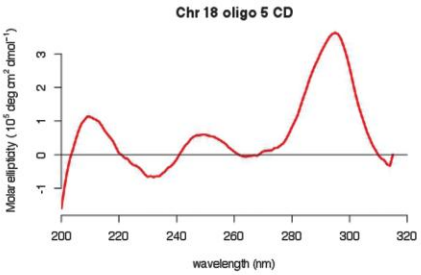

Chr 9 Peak 2 oligo 7 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

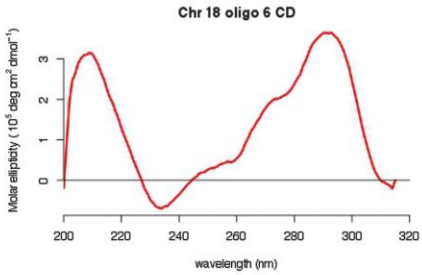

Chr 18 oligo 1 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

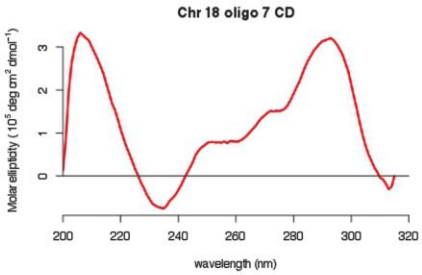

Chr 18 oligo 2 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

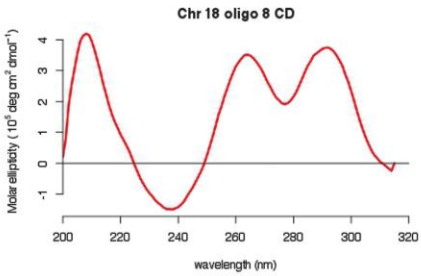

Chr 18 oligo 3 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

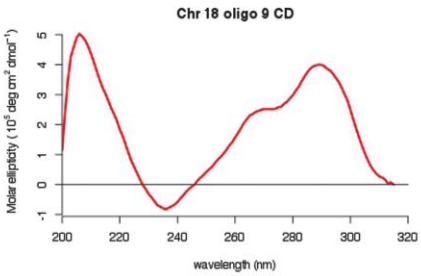

Chr 18 oligo 4 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

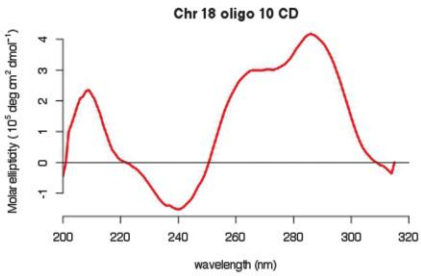

Chr 18 oligo 5 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)

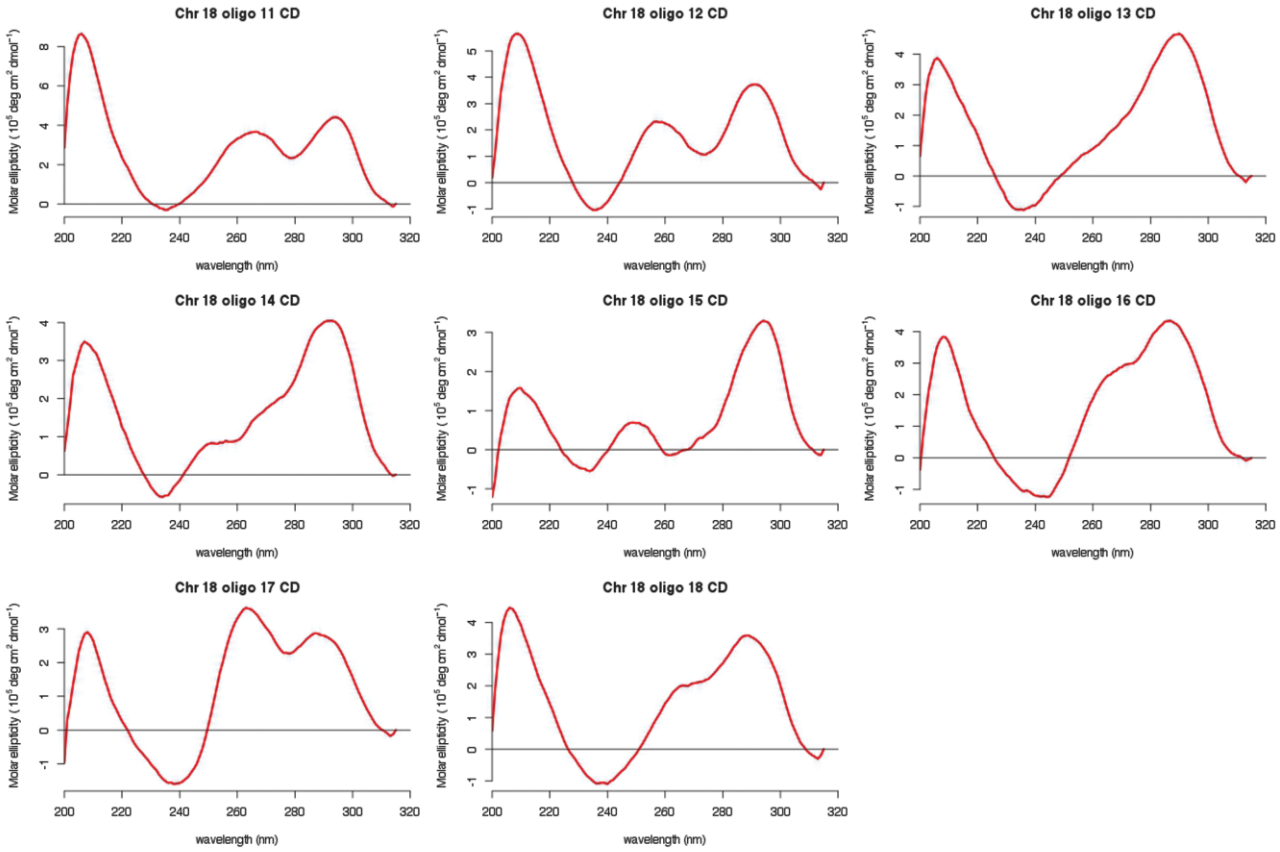

Chr 18 oligo 6 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)


Chr 18 oligo 7 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)


Chr 18 oligo 8 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)


Chr 18 oligo 9 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)


Chr 18 oligo 10 CD
Molar ellipticity (10^5 deg cm^2 dmol^{-1})
wavelength (nm)



The majority of the CD spectra show characteristics of both parallel and anti-parallel G-quadruplexes, suggesting they are either hybrid G-quadruplexes or a mixture of parallel and anti-parallel G-quadruplexes.

Supplementary Table S1 Summary of ENCODE quality metrics for four hf2 pull-down libraries.

Library	No. reads	No. mapq ≥15	Self cons IDR 0.02	SPOT	NSC	RSC	PBC
Rep 1	43363413	21694123	772 (2.5)	0.26	1.06	3.53	0.73
Rep 2	12312124	8213628	485 (1.6)	0.32	1.53	19.79	0.60
Rep 3	48155845	36225937	308 (1)	0.55	1.35	20.77	0.17
Rep 4	2869469	1183304	-	0.14	1.08	1.48	0.95
ENCODE guidelines	-	≥10000000	N in similar range (<2)	ENCODE avg: 0.46	>1.05	>0.8	Moderate: 0.5-0.8

The QC metrics for the libraries fall broadly within the acceptable thresholds recommended by the ENCODE consortium, and are comparable to transcription factor ChIP-Seq data [53](see http://encodeproject.org/ENCODE/qualityMetrics.html[54]) for the values of these metrics on the ENCODE transcription factor ChIP-Seq data). In downstream analysis, to reduce bias due to differences in library size, we down-sampled each library to 8 million reads. Note that the starting library size for replicates 2 and 4 is small giving a lower number of mapped reads. Also, the number of self-consistent peaks in replicate 1 is slightly higher then the one in other libraries while replicate 3 shows a PCR bottleneck.

No. reads: Total number of reads sequenced.

No. mapq ≥ 15: Number of reads uniquely mapped to the reference with MAPQ ≥ 15.

Self cons IDR 0.02: Self-consistent peaks with an irreproducibilty discovery rate (IDR) < 0.02 (number of peaks in common between the two halves of each library). In parenthesis the ratio between the number of peaks in the library and the smallest number of peaks in any library (308). The number of self-consistent peaks should be comparable across replicates (within a factor of 2). This metric is not available for replicate 4 due to the small library size.

SPOT: Fraction of reads in enriched regions as calculated by the hotspot program (referred to as ptih by hotpsot)[55].

NSC: Normalized Strand Cross-correlation coefficient, a measure of enrichment derived without dependence on prior determination of enriched regions.

RSC: Relative Strand Cross-correlation coefficient, a measure of enrichment derived without dependence on prior determination of enriched regions.

PBC PCR Bottleneck Coefficient, a measure of library complexity and calculated as number of non-redundant reads (reads mapped uniquely and at different positions) / Number of uniquely mapped reads. The range 0.5-0.8 is considered a moderate bottleneck, 0-0.5 a severe bottleneck.

For more information see also http://www.encodeproject.org/ENCODE/qualityMetrics.html#definitions[54].

Supplementary Table S2 Number of peaks called in hf2 pull-down libraries and intersecting with *G4-calculator* PQS.

	Number of peaks	**Number in *G4-calculator* PQS**
In only 1 library	8368	4221 (50.4%)
In only 2 libraries	503	367 (72.8%)
In 3 or 4 libraries	265	201 (75.8%)
Total	9136	4789 (52.4%)

A list of peaks from the pull-down libraries was created and any peaks that overlap by at least 1 bp was combined to create a single peak, resulting in a set of peaks found in either one, two, three or all four libraries. The number of peaks in common between any two, or three or more libraries, and those present in only one library, were calculated. This resulted in 9136 peaks, of which 4789 (52.4%) contained a predicted G-quadruplex sequence.

Supplementary Table S3 Number of peaks called in pull-down libraries and intersecting with *quadparser* PQS.

	Number of peaks	**Number in *quadparser* PQS**
In only 1 library	8368	888 (10.6%)
In only 2 libraries	503	122 (24.3%)
In 3 or 4 libraries	265	53 (20%)
Total	9136	1063 (11.6%)

A list of peaks from the pull-down libraries was created and any peaks that overlap by at least 1 bp was combined to create a single peak, resulting in a set of peaks found in either one, two, three or all four libraries. The number of peaks in common between any 2, or 3 or more libraries, and those present in only 1 library, were calculated. This resulted in 9136 peaks, of which 1063 (11.6%) overlapped with a predicted G-quadruplex sequence.

Supplementary Table S4 The chromosomal position of the 175 peaks identified in at least two of the four libraries that overlap with PQS by *quadparser*.

Chr	Start	End	Size	Possible no. of G-quadruplexes	Possible no. of simultaneous G-quadruplexes	Feature
chr1	8	519	511	69	18	1218bp upstream of DDX11L1
chr1	9042368	9043340	972	1	1	Intron 1 of SLC2A5
chr1	31676804	31677979	1175	46	12	Intron 8 of SERINC2
chr1	41662288	41663564	1276	2	1	53469bp downstream of EDN2
chr1	107106987	107110502	3515	1	1	290288bp upstream of PRMT6
chr1	113933419	113934116	697	1	1	Intron 4 of MAGI3
chr1	114397764	114399220	1456	1	1	34217bp downstream of SYT6
chr1	114562519	114563649	1130	1	1	64525bp upstream of SYT6
chr1	143769360	143770576	1216	4	1	Intron 7 of PDE4DIP
chr1	158370228	158370921	693	1	1	Intron 13 of ATP1A2
chr1	179662391	179663354	963	9	5	55955bp upstream of CACNA1E
chr1	200859764	200861200	1436	2	2	Intron 1 of SYT2
chr1	205163381	205165054	1673	1	1	1381bp upstream of FAIM3
chr1	226321620	226326612	4992	2	1	6026bp downstream of WNT3A
chr1	227180015	227180595	580	1	1	230977bp downstream of RHOU
chr10	99820	100673	853	19	7	14643bp upstream of TUBB8
chr10	3274083	3275206	1123	2	1	69051bp upstream of PITRM1
chr10	5529186	5530438	1252	1	1	220bp downstream of CALML5
chr10	8549891	8550961	1070	1	1	392722bp downstream of GATA3
chr10	11365225	11367034	1809	1	1	Intron 8 of CELF2
chr10	129485296	129486119	823	8	5	55857bp downstream of FOXI2
chr10	132478284	132480482	2198	4	3	170359bp downstream of MIR378C
chr10	134545250	134545875	625	1	1	Intron 27 of TTC40
chr10	135229754	135231090	1336	23	9	Exon 2 of SPRNP1
chr11	2104287	2105240	953	12	6	1678bp downstream of INS-IGF2
chr11	17741532	17742603	1071	1	1	Intron 1 of KCNC1

chr11	59293690	59294679	989	1	1	Intron 1 of STX3
chr11	68340473	68341539	1066	1	1	Intron 1 of CPT1A
chr11	69021482	69022975	1493	1	1	142079bp upstream of CCND1
chr11	70469951	70470667	716	4	4	Intron 7 of SHANK2
chr11	70828658	70829373	715	2	2	Intron 6 of DHCR7
chr12	120672689	120673517	828	17	9	Exon and intron 4 of TMEM120B
chr12	129489606	129490677	1071	10	5	Intron 8 of RIMBP2
chr13	19864378	19866450	2072	3	1	9356bp downstream of CRYL1
chr13	28963008	28964794	1786	1	1	Intron 9 of MTUS2
chr13	30314309	30317148	2839	4	1	31528bp downstream of LINC00398
chr13	113635573	113636613	1040	1	1	Intron 3 of FAM70B
chr14	99703356	99704959	1603	4	3	7592bp upstream of DEGS2
chr14	104223054	104223970	916	6	4	3018bp upstream of INF2
chr14	105100301	105101980	1679	15	9	32718bp downstream of TMEM121
chr14	105106275	105108635	2360	1	1	38692bp downstream of TMEM121
chr14	105202820	105204090	1270	2	2	2869bp downstream of ELK2AP
chr14	105220341	105221616	1275	12	8	10153bp upstream of ELK2AP
chr14	105226011	105228871	2860	1	1	15823bp upstream of ELK2AP
chr14	105275354	105277318	1964	2	2	65166bp upstream of ELK2AP
chr15	28252133	28252958	825	2	2	22573bp upstream of DKFZP434L187
chr15	38129839	38130686	847	1	1	Intron 2 of NR_040062
chr15	77171121	77172372	1251	1	1	852bp upstream of RASGRF1
chr15	77531162	77532163	1001	1	1	Intron 1 of KIAA1024
chr16	967660	968443	783	4	1	Intron 3 of LMF1
chr16	10041147	10042110	963	1	1	Intron 2 of GRIN2A
chr16	26292786	26293697	911	2	2	236277bp downstream of HS3ST4
chr16	28007581	28009051	1470	9	4	7766bp downstream of XPO6
chr16	29963847	29964796	949	1	1	7116bp upstream of ALDOA
chr16	71842895	71843546	651	2	1	134659bp upstream of LOC100506172
chr16	71877818	71879596	1778	3	2	98609bp upstream of LOC100506172
chr16	78416324	78417106	782	1	1	224202bp upstream of MAF

chr16	83521280	83522414	1134	32	12	20664bp downstream of CRISPLD2
chr16	84162831	84164170	1339	1	1	38360bp upstream of KIAA0182
chr16	85293975	85294735	760	8	4	121171bp downstream of FOXL1
chr16	85500290	85501114	824	1	1	327486bp downstream of FOXL1
chr16	85563243	85564448	1205	6	3	329457bp downstream of C16orf95
chr16	86115229	86116480	1251	2	1	32269bp upstream of ZCCHC14
chr16	87113048	87114877	1829	14	6	Intron 4 of ZFPM1
chr16	87605071	87606830	1759	3	1	34067bp upstream of CBFA2T3
chr16	88228335	88229442	1107	8	3	Intron 2 of DPEP1
chr17	7018376	7019958	1582	1	1	Intron 2 of ASGR1
chr17	9070925	9072004	1079	1	1	Intron 6 of NTN1
chr17	21079212	21080502	1290	4	2	2274bp downstream of C17orf103
chr17	21250806	21252490	1684	3	3	Intron 1 of KCNJ12
chr17	54350821	54353967	3146	1	1	Intron 1 of PPM1E
chr17	55561798	55563216	1418	1	1	18868bp upstream of CA4
chr17	55928915	55935478	6563	1	1	Exon 2 of APPBP2
chr17	56703580	56707890	4310	1	1	Intron 23 of BCAS3
chr17	56976149	56984472	8323	1	1	38104bp downstream of NACA2
chr17	58165206	58165845	639	1	1	Intron 6 of MARCH10
chr17	60926093	60927420	1327	1	1	27725bp downstream of AXIN2
chr17	62452279	62453637	1358	1	1	Intron 3 of CACNG4
chr17	77013084	77014744	1660	10	3	Intron 1 of BAHCC1
chr17	77535787	77536782	995	6	3	Intron 3 of ASPSCR1
chr18	53391	54431	1040	35	11	44634bp upstream of ROCK1P1
chr18	74466821	74468208	1387	1	1	373055bp upstream of SALL3
chr18*	2	789	787	102	26	98276bp upstream of ROCK1P1
chr19	196333	197139	806	45	14	34905bp downstream of PPAP2C
chr19	52684952	52685598	646	1	1	Intron 9 of NAPA
chr19	59328816	59329791	975	8	5	1855bp downstream of PRPF31
chr19	63789576	63790088	512	21	6	2003bp downstream of MGC2752
chr2	7	799	792	57	17	28015bp downstream of FAM110C

chr2	67294386	67295929	1543	1	1	Intron 1 of LOC644838
chr2	71698006	71699580	1574	1	1	Intron 40 of DYSF
chr2	91408643	91409274	631	10	4	74764bp downstream of GGT8P
chr2	114076583	114077750	1167	107	31	Exon and Intron 1 of DDX11L2
chr2	121607959	121608989	1030	1	1	81645bp downstream of TFCP2L1
chr2	240735321	240736372	1051	2	1	6576bp upstream of OTOS
chr20	28259785	28261460	1675	3	3	7987bp upstream of MLLT10P1
chr20	45239349	45242327	2978	12	5	Intron 12 of EYA2
chr20	45248764	45254467	5703	1	1	Exon 16 of EYA2
chr20	45467778	45470464	2686	1	1	48898bp upstream of ZMYND8
chr20	45492655	45496322	3667	1	1	67686bp upstream of NCOA3
chr20	45602175	45604933	2758	18	5	Intron 1 of NCOA3
chr20	46511302	46512503	1201	11	5	78515bp downstream of LINC00494
chr20	46738917	46740334	1417	2	2	Intron 9 of PREX1
chr20	46855288	46859216	3928	3	2	Intron 1 of PREX1
chr20	46887904	46889539	1635	1	1	10078bp upstream of PREX1
chr20	46923635	46925834	2199	3	1	45809bp upstream of PREX1
chr20	48784912	48792215	7303	1	1	Exon 2 of PARD6B
chr20	51746773	51752096	5323	1	1	113731bp upstream of ZNF217
chr20	52518710	52524246	5536	2	2	1427bp upstream of DOK5
chr20	52977654	52979623	1969	2	1	276538bp downstream of DOK5
chr20	55039856	55041017	1161	1	1	136199bp downstream of BMP7
chr20	55425486	55426232	746	2	2	7695bp downstream of RBM38
chr20	55668405	55669679	1274	2	1	Intron 1 of PMEPA1
chr20	55982590	55984285	1695	1	1	78685bp downstream of MIR4532
chr20	56528113	56529073	960	8	4	Intron 2 of LOC149773
chr20	56751874	56753431	1557	5	2	27568bp downstream of NPEPL1,STX16-NPEPL1
chr20	57646617	57648176	1559	3	3	Intron 1 of PHACTR3
chr20	58508632	58510305	1673	5	3	21999bp downstream of MIR4533
chr20	58772085	58773473	1388	1	1	285452bp downstream of MIR4533

chr20	62388336	62389608	1272	97	29	2574bp upstream of LINC00266-1
chr21	40279061	40280278	1217	1	1	25935bp downstream of DSCAM
chr21	43452868	43453992	1124	5	2	8218bp upstream of CRYAA
chr21	43466670	43467170	500	1	1	689bp downstream of CRYAA
chr21	43711963	43713103	1140	1	1	Intron 3 of LINC00313
chr21	45228050	45228831	781	14	5	5376bp downstream of LINC00163
chr22	43109232	43110876	1644	2	1	69169bp upstream of KIAA1644
chr3	14270948	14272195	1247	1	1	56076bp downstream of LSM3
chr3	134617828	134619365	1537	1	1	Intron 1 of BFSP2
chr3	197905523	197907044	1521	16	10	10501bp downstream of CEP19
chr3	199383862	199386032	2170	146	44	Intron 6 of FAM157A
chr3*	10889757	10890887	1130	4	2	Intron 5 of SLC6A11
chr4	9	366	357	24	7	42861bp upstream of ZNF595
chr4	688654	689387	733	18	8	186bp upstream of PCGF3
chr4	1623428	1624722	1294	1	1	Intron 6 of FAM53A
chr4	3549463	3550658	1195	1	1	Intron 1 of FLJ35424
chr4	7692045	7692700	655	1	1	Intron 4 of SORCS2
chr4	37921144	37922210	1066	1	1	103954bp downstream of TBC1D1
chr4	169836325	169837467	1142	2	1	Intron 3 of PALLD
chr5	398015	399095	1080	2	2	Intron 2 of AHRR
chr5	637777	638547	770	2	2	26858bp upstream of CEP72
chr5	1036358	1037195	837	1	1	4230bp downstream of LOC100506688
chr5	1094591	1102038	7447	1	1	1451bp downstream of SLC12A7
chr5	11364409	11365512	1103	11	4	Intron 9 of CTNND2
chr5	11495798	11496702	904	2	1	Intron 3 of CTNND2
chr5	170958025	170959090	1065	1	1	140791bp downstream of FGF18
chr5	176189693	176190793	1100	1	1	Intron 1 of UNC5A
chr5	177907130	177907912	782	20	8	Intron 2 of COL23A1
chr5	178304828	178305503	675	1	1	Intron 2 of ZNF454
chr5	179954198	179955353	1155	9	5	3106bp upstream of SCGB3A1
chr6	3721808	3722556	748	1	1	24564bp upstream of PXDC1

chr6	43813676	43814720	1044	2	2	31204bp upstream of VEGFA
chr7	284572	285283	711	1	1	2769bp upstream of FAM20C
chr7	1585592	1586639	1047	1	1	Intron 3 of KIAA1908
chr7	1762046	1763079	1033	4	1	7931bp downstream of ELFN1
chr7	76463952	76466431	2479	3	3	Intron 2 of DTX2P1-UPK3BP1-PMS2P11
chr7	98381793	98382517	724	2	1	Intron 32 of TRRAP
chr7	150237585	150238894	1309	1	1	34083bp downstream of KCNH2
chr7	150525496	150526443	947	3	2	9413bp downstream of ABCF2
chr7	154629961	154631827	1866	10	5	88592bp upstream of INSIG1
chr7	156784071	156785423	1352	2	1	29245bp downstream of UBE3C
chr8	82916991	82917900	909	6	2	Exon 1 of SNX16
chr8	102733468	102734329	861	1	1	Intron 14 of GRHL2
chr8	125818323	125819253	930	2	2	8413bp upstream of MTSS1
chr8	129637743	129638857	1114	1	1	406128bp downstream of MIR1208
chr8	134485671	134486724	1053	1	1	49549bp downstream of ST3GAL1
chr8	143574081	143575023	942	3	3	Intron 16 of BAI1
chr8	143821463	143823604	2141	5	3	633bp upstream of SLURP1
chr8	144545239	144546104	865	1	1	7707bp downstream of RHPN1
chr9	67906956	67910038	3082	1	1	7565bp downstream of LOC642236
chr9	89436616	89438253	1637	2	1	Intron 3 of DAPK1
chr9	123697261	123698412	1151	1	1	Intron 6 of TTLL11
chr9	136685978	136686856	878	7	5	Intron 1 of COL5A1
chr9	137805634	137806436	802	1	1	Intron 19 of KCNT1
chr9*	5	634	629	68	18	1353bp upstream of DDX11L5
chr9*	140173853	140174454	601	19	7	Intron 1 of TUBBP5
chrX	67789045	67790557	1512	1	1	Intron 1 of STARD8

The number of predicted G-quadruplexes and the nearest gene are shown. * denotes peaks analyzed by CD in **Supplementary Figure 12**.

Supplementary Table S5 Oligonucleotide sequences analyzed by circular dichroism

chr9 CD oligo1	GGGTGAGGGTGAGGGTGGGGG
chr9 CD oligo2	GGGGGTGGGGGTTGGGGTTGGGG
chr9 CD oligo3	GGGGTTAGGGTTCGGGTTCGGG
chr9 CD oligo4	GGGTTCGGGTTCGGGTTCGGG
chr9 CD oligo5	GGGTTGGGTTAGGGTTAGGG
chr9 CD oligo6	GGGTAGGGTTAGGGTTTAGGG
chr9 CD oligo7	GGGTTTAGGGTTTAGGGTTTAGGG
chr9 CD oligo8	GGGGTTAGGGGTTAGGGTTAGGG
chr9 CD oligo9	GGGTTAGGGTTAGGGGTTAGGGG
chr9 CD oligo10	GGGGTTAGGGTTAGGGTTAGGG
chr9 CD oligo11	GGGTTAGGGTTAGGGTTAAGGG
chr9 CD oligo12	GGGGTTAGGGGTTAGGGGTTAGGGG
chr9 CD oligo13	GGGTAGGGTTAGGGTGAGGG
chr9 CD oligo14	GGGTGAGGGTGAGGGTGAGGG
chr9 CD oligo15	GGGTGAGGGTTGGGTTAGGG
chr9 CD oligo16	GGGTTAGGGTTGGGTTAGGG
chr9 CD oligo17	GGGTTAGGGTTGGGTTGGGG
chr18 CD oligo 1	GGGTGAGGGTTAGGGTTAGGG
chr18 CD oligo 2	GGGGTAGGGGTAGGGGTAGGG
chr18 CD oligo 3	GGGGTAGGGTTAGGGTTAGGG
chr18 CD oligo 4	GGGTTAAGGGTTAGGGTTAGGG
chr18 CD oligo 5	GGGTTAGGGTAGGGTAGGG
chr18 CD oligo 6	GGGTAGGGTTAGGGTTAGGG
chr18 CD oligo 7	GGGTTAGGGTTAGGGTTAGGG
chr18 CD oligo 8	GGGTAGGGGTTAGGGGTTAGGG
chr18 CD oligo 9	GGGGTTTAGGGTTAGGGTTAGGG
chr18 CD oligo 10	GGGTAGGGTTGGGGTTGGGG
chr18 CD oligo 11	GGGGTTGGGGTTGGGGTTGGGG
chr18 CD oligo 12	GGGGTTGGGGTAGGGTTAGGG
chr18 CD oligo 13	GGGTTAGGGTTAAGGGTTAGGG
chr18 CD oligo 14	GGGTTAGGGTTAGGGTAGGG
chr18 CD oligo 15	GGGTTAGGGTAGGGTTAGGG
chr18 CD oligo 16	GGGTTAGGGTAGGGGTTAGGGG
chr18 CD oligo 17	GGGTTGGGGTTGGGGTTGGGG
chr18 CD oligo 18	GGGTTAGGGTTAGGGGTAAGGG
chr3 CD oligo 1	GGGTAGGGTAGGGTAGGG
chr3 CD oligo 2	GGGTAGGGTATGGTAGGGTAGGG
chr9 Peak 2 CD oligo 1	GGGTTAAGGGTTAGGGTGAGGG
chr9 Peak 2 CD oligo 2	GGGTGAGGGTTAGGGTTAGGGG
chr9 Peak 2 CD oligo 3	GGGTTAGGGCTAGGGTTGGGG
chr9 Peak 2 CD oligo 4	GGGTTAAGGGTTGGGGTTGGGGG
chr9 Peak 2 CD oligo 5	GGGTTAGGGGTTAGGGTAAGGG
chr9 Peak 2 CD oligo 6	GGGTTAGGGTTTGGGTTAGGG
chr9 Peak 2 CD oligo 7	GGGGTTAGGGCTAGGGCTAGGG

Supplementary References:

53 Landt, S. G. *et al. Genome Res* **22**, 1813-1831, (2012).
54 *Quality Metrics*, <http://encodeproject.org/ENCODE/qualityMetrics.html> (2012).
55 John, S. *et al. Nature genetics* **43**, 264-268, (2011).
56 Li, Q. H., Brown, J. B., Huang, H. Y. & Bickel, P. J. *Ann Appl Stat* **5**, 1752-1779, (2011).