

Research article

Open Access

Association of poly-purine/poly-pyrimidine sequences with meiotic recombination hot spots

Andrew TM Bagshaw*¹, Joel PW Pitt² and Neil J Gemmell¹

Address: ¹School of Biological Sciences, University of Canterbury, New Zealand and ²Bioprotection and Ecology Division, Lincoln University, New Zealand

Email: Andrew TM Bagshaw* - atb20@student.canterbury.ac.nz; Joel PW Pitt - joel.pitt@gmail.com; Neil J Gemmell - neil.gemmell@canterbury.ac.nz

* Corresponding author

Published: 18 July 2006

Received: 18 January 2006

BMC Genomics 2006, 7:179 doi:10.1186/1471-2164-7-179

Accepted: 18 July 2006

This article is available from: <http://www.biomedcentral.com/1471-2164/7/179>

© 2006 Bagshaw et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Meiotic recombination events have been found to concentrate in 1–2.5 kilo base regions, but these recombination hot spots do not share a consensus sequence and why they occur at specific sites is not fully understood. Some previous evidence suggests that poly-purine/poly-pyrimidine (poly-pu/py) tracts (PPTs), a class of sequence with distinctive biochemical properties, could be involved in recombination, but no general association of PPTs with meiotic recombination hot spots has previously been reported.

Results: We used computational methods to investigate in detail the relationship between PPTs and hot spots. We show statistical associations of PPT frequency with hot spots of meiotic recombination initiating lesions, double-strand breaks, in the genome of the yeast *S. cerevisiae* and with experimentally well characterized human meiotic recombination hot spots. Supporting a possible role of poly-pu/py-rich sequences in hot spot recombination, we also found that all three single nucleotide polymorphisms previously shown to be associated with human hot spot activity changes occur within sequence contexts of 14 bp or longer that are 85% or more poly-pu/py and at least 70% G/C. These polymorphisms are all close to the hot spot mid points. Comparing the sequences of experimentally characterized human hot spots with the orthologous regions of the chimpanzee genome previously shown not to contain hot spots, we found that in all five cases in which comparisons for the hot spot central regions are possible with publicly available sequence data, there are differences near the human hot spot mid points within sequences 14 bp or longer consisting of more than 80% poly-pu/py and at least 50% G/C.

Conclusion: Our results, along with previous evidence for the unique biochemical properties and recombination-stimulating potential of poly-pu/py-rich sequences, suggest that the possible functional involvement of this type of sequence in meiotic recombination hot spots deserves further experimental exploration.

Background

Crossovers between chromosomes occur during meiotic cell division resulting in heritable genetic recombination.

These crossovers have a complex, non-random distribution, and in the last decade recombination hot spots 1–2.5 kilo bases (kb) wide have been experimentally well

characterized in yeast, mice and humans (reviewed in [1,2]). Hot spots show a wide range of crossover frequencies, which are occasionally several hundred times greater than the level expected if crossovers were distributed randomly across chromosomes. Experimental studies that have located multiple hot spots in contiguous regions of the human genome have found that they often occur in clusters, which are separated by regions of about 50–100 kb showing very low recombination frequencies [3–5].

Hot spots do not share a consensus sequence, and the mechanisms responsible for regulating their distribution and activity levels are not well understood, but several molecular features of hot spot recombination have been described (reviewed in [6]). These include a locally open chromatin structure, presumably allowing access to recombination machinery [7,8], and a requirement for a chromosomal double-strand break (DSB) to initiate recombination [9,10]. A less well explained feature of hot spots is the influence on activity levels of sequence context, including flanking sequences several kb away [11–13]. This suggests that epigenetic, or distal sequence, factors may have a greater influence than local sequences on hot spot regulation, and consistent with this idea are recent studies showing that the locations of hot spots in humans and chimpanzees do not correspond despite more than 98% sequence similarity between the two species [14,15].

Local sequences clearly have some influence on hot spot recombination, however, as activity levels can be sensitive to local, single nucleotide changes [16–19]. The molecular basis of this sensitivity is largely unexplained, with the exception of the *Schizosaccharomyces pombe ade 6* hot spot, in which a single nucleotide change promotes transcription factor-binding, creating the hot spot [20,21], though its activity is also influenced by flanking sequences [22,23]. Other sequences have been found to be associated with hot spots including GC-rich DNA [24], tandem repeats [19,25], transcription factor binding sites [26–28], poly-A/T tracts [29] and some specific motifs less than 10 bp long [30]. Direct experimental evidence for a role of such sequence features in hot spot activity is lacking in most cases, the exceptions being transcription factor binding sites [20,21,26–28], and a 14 bp poly-A/T tract, deletion of which was found to reduce gene conversion activity at the yeast *ARG4* hot spot by 75% [29].

Human recombination rates estimated across windows hundreds of thousands to millions of bp wide have been reported to correlate negatively with poly-A/T tract frequencies [31], but positive correlations with broad scale recombination rates have been shown for other simple sequences, including the class of sequence with purines (A or G) on one strand of the DNA duplex, and pyrimidines

(T or C) on the complementary strand, of which poly-A/T is a subset [31,32]. Experimental evidence also suggests that poly-purine/poly-pyrimidine (poly-pu/py) tracts (PPTs) in general deserve further attention, for example they have been shown to affect recombination [33], replication [34], and gene expression [35] in model systems and these effects have often been linked to the ability of PPTs with some GC-content readily to form stable intramolecular secondary structures under physiological conditions [33,34,36,37]. These structures can be sensitive to single nucleotide changes [38,39], but cannot be formed by poly-A/T [40], though poly-A/T duplex DNA can form intermolecular three-stranded aggregates with synthetic oligonucleotides [41]. Recently, a study predicting hot spot locations throughout the human genome based on statistical analysis of haplotype data found several poly-pu/py-rich motifs of 5–9 bp to be associated with hot spots [30], providing further support for the possible importance of poly-pu/py-rich sequences in meiotic recombination.

We investigated in detail the relationship between PPTs and hot spots in humans and the yeast *S. cerevisiae*. We show significantly elevated PPT frequencies in hot spots in both species. We also found that the three single nucleotide changes shown to be associated with human hot spot recombination rates all occur in high GC-content poly-pu/py-rich sequences of at least 14 bp, and that sequence differences between humans and chimpanzees in regions where there is a hot spot in humans but none in chimpanzees occur in similar poly-pu/py-rich sequence contexts.

Results and discussion

Hot spots investigated

Recent experimental studies in humans [3–5] and yeast [24] have reported multiple meiotic recombination hot spots and cold regions across contiguous segments of genomic DNA, allowing comparison of sequence patterns between the two types of regions. In the yeast *S. cerevisiae*, recombination intensity has been assayed indirectly throughout the entire genome using microarray analysis of DSB frequency patterns [24]. This study identified 303 hot and 49 cold open reading frames (ORFs), and, combining adjacent ones, defined 177 hot spots, which encompassed all previously known hot spots in the species, and 40 cold spots. For the purposes of our investigation, we extended the hot and cold spots to include the intergenic regions (IGRs) adjacent to the ORFs concerned, since yeast hot spots are typically centred on IGRs, in which most DSBs occur [42]. The hot spots as we defined them had a mean length of 3466 bp. DSBs have also been mapped to 76 much narrower sites on *S. cerevisiae* chromosome 3 [42], but many of these had very low levels of DSBs and may therefore indicate background recombina-

tion events rather than hot spots as normally defined, so we analysed them separately.

In humans, multiple hot spots have been mapped by experimental methods in the MHC Class II region on chromosome 6, in which seven hot spots have been identified over 292 kb [4,5,43], and in a 206 kb segment of chromosome 1, in which eight hot spots have been mapped [3,19]. In each region, areas between hot spots showed very low levels of recombination. Two other human hot spots have been well characterized experimentally. In the Beta-Globin gene cluster, a hot spot was mapped to within 11 kb, with a 90 kb adjacent cold region also identified [44]. The location of the Beta-Globin hot spot was later narrowed to a region of under 2 kb [45]. Finally, in the pseudoautosomal region of the Y chromosome, a 9.9 kb section of the *SHOX* gene was assayed for recombination and found to contain a hot spot [46]. All of the human hot spots included in our analysis were 2.5 kb or less in size. We limited our investigation in humans to these 17 well-characterized hot spots known at the time of writing. Hot spots have also been predicted throughout the human genome using statistical analysis of haplotype data to infer past recombination events [30], but recent evidence indicates that these methods are not always reliable for predicting hot spots in the present generation [3,5,47,48].

Hot spots show elevated frequencies of PPTs

We used a pattern-matching computer algorithm to detect PPTs in the hot spots and cold regions reported in the above-mentioned studies. A 12 bp PPT has been shown to form a stable intramolecular quadruplex [49], but in our search of the literature we did not find reports of shorter PPTs forming stable secondary structures, so we initially searched for PPTs of at least 12 bp. In yeast, mean frequencies per kb were 1.92 in hot spots, and 0.97 in cold spots, which was a statistically significant difference ($p = 1.74 \times 10^{-10}$). Because most DSBs occur in IGRs [42], we repeated the analysis excluding ORFs and found that the difference between hot spots and cold spots increased, with the mean per kb frequencies 3.93 in hot spot IGRs and 1.62 in cold spot IGRs ($p = 1.65 \times 10^{-9}$). The mean length of PPTs of at least 12 bp was 15.49 bp in hot spots and 14.22 bp cold spots, also a statistically significant difference ($p = 0.0036$). In yeast, there were no significant differences between hot and cold spots for mean PPT GC-content.

Transcription factors can bind to PPTs as short as 5 bp [50], and we wished to know whether very long PPTs were associated with hot spots, so we varied the minimum PPT length between 5 and 40 bp. We found greater mean PPT frequencies in hot spots than cold spots, with the differences statistically significant, for length minima 7 through 33. Frequencies for the 233 remaining regions of the

genome not classified as hot or cold spots had similar frequencies to cold spots, showing that the differences were primarily due to elevated frequencies in hot spots rather than lowered frequencies in cold spots (Figure 1).

The 17 human hot spots did not show significantly elevated PPT frequencies for length minima 5–12 bp. Increasing the length minimum increased the ratio of frequencies between hot spots and cold regions, however, with the differences statistically significant for minima of 13 bp (means per kb 2.02 in hot spots and 1.33 in cold regions, $p = 0.045$), 14 bp (means per kb 1.50 in hot spots and 0.99 in cold regions, $p = 0.036$), and 15 bp (means per kb 1.21 in hot spots and 0.73 in cold regions, $p = 0.036$). In humans, no significant differences between hot spots and cold regions were found for mean PPT length, but human hot spot-associated PPTs had a higher mean GC-content (45.4%) than those in cold regions (37.5%), which was a statistically significant difference ($p = 0.001$). The difference was not due to a high GC-content ratio between hot spots and cold areas (1.00). Evidence indicates that high GC-content PPTs are more likely to form secondary structures [40,51-53], so we repeated the searches looking only for PPTs with GC-contents above the mean, as calculated across all PPTs of at least 12 bp

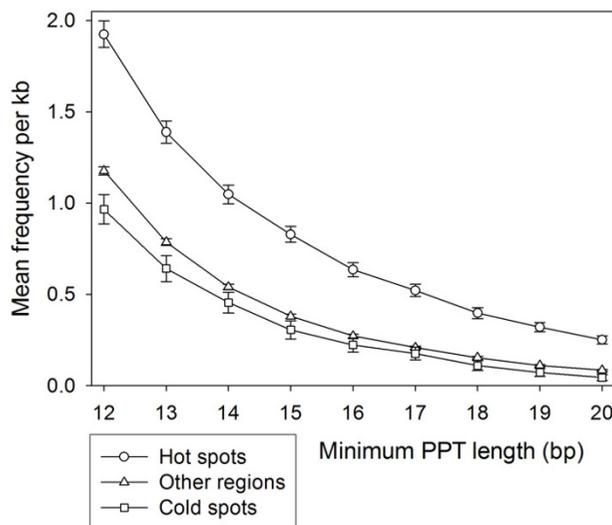


Figure 1
Association of high PPT frequencies with yeast hot spots. Mean frequencies of PPTs with length minima between 12 and 20 bp in 177 hot spots of meiotic double-strand breaks (DSBs) in the genome of the yeast *S. cerevisiae* compared with the mean frequencies in 40 DSB cold spots and 233 other regions comprising the remainder of the genome sequence. Error bars are plus and minus one standard error of the mean.

found in the study regions (38% G/C). We found that this limitation increased the levels of enrichment of PPTs in human hot spots. The associations were statistically significant for size minima 13 bp (mean frequencies per kb 1.48 in hot spots and 0.795 in cold regions, $p = 0.0079$) and 14 bp (mean frequencies per kb 1.04 in hot spots and 0.533 in cold regions, $p = 0.017$). In contrast, the association between PPTs and hot spots in yeast was weaker for high GC-content PPTs.

Because PPTs with mismatches to the homopurine/homopyrimidine motif are much more common than pure PPTs, we repeated all the searches allowing some mismatches. In humans, mean frequency differences for tracts of at least 20 bp (means per kb 1.51 in hot spots and 0.80 in cold regions, $p = 0.022$) and 24 bp (means per kb 0.92 in hot spots and 0.34 in cold regions $p = 0.036$) were statistically significant for PPTs with one mismatch per 10 bp allowed. No significant differences were found with one mismatch allowed every 5 bp. In yeast, allowing mismatches generally reduced the frequency ratios between hot spots and cold spots, but did increase the greatest size minimum for which the difference was statistically significant from 33 bp (for pure PPTs) to 39 bp when one mismatch was allowed per 10 bp, and 41 bp when one mismatch was allowed per 5 bp.

In both yeast and humans hot spot intensities vary greatly, so we asked if hot spot intensity was correlated with PPT frequency. We found no significant correlations in humans, but in yeast we did find significant positive correlations between previously reported hot spot intensity [24] and mean PPT frequencies with the length minima 18, 23, 25 and 26 bp, and for minima 26, 29 and 32 through 34 with one mismatch allowed per 10 bp. Even longer PPTs were correlated when one mismatch was allowed per 5 bp, with significance for minima 33, 35 and 37 through 39. The correlations were weak in all cases, however, with coefficients ranging between 0.15 and 0.19 (Spearman's rho).

The lack of a strong correlation between PPT frequency and hot spot intensity suggests that a high density of PPTs is not in itself a primary determinant of hot spot activity levels. High PPT density is also not sufficient in itself to cause a hot spot, since hot spot locations in humans and chimpanzees do not correspond despite more than 98% sequence similarity between the two species [14,15]. High PPT frequency could be a factor in hot spots, however, and it is also possible that only certain types of PPT may be involved in recombination. Below, we discuss evidence that PPTs could stimulate recombination through secondary structure formation and/or protein binding. Exact sequence requirements are not fully understood, either for binding of recombination-related proteins, or secondary

structure formation by PPTs. It is possible, therefore, that the association between PPT frequency and hot spots may be due to a greater likelihood that regions with a high frequency of PPTs will contain a functional tract.

PPTs are associated with DSB sites mapped with a resolution of about 500 bp

Baudat and Nicolas (1997) mapped meiotic DSBs throughout chromosome 3 of the genome of the yeast *S. cerevisiae* and identified 6 ORFs and 70 IGRs subject to at least one DSB [42]. Overall, these DSB-containing ORFs and IGRs averaged 567 bp in length. We found that 53 of the 70 DSB-containing IGRs had at least one PPT of 12 bp or longer (76%), and 35 of them had a PPT of at least 15 bp (50%). Of the 92 IGRs on the chromosome without a mapped DSB, 56 had a PPT of at least 12 bp (61%), and 26 had one of at least 15 bp (28%). Frequencies of PPTs with length minima 5–15 bp were significantly elevated in DSB-containing IGRs compared with the remaining IGRs on the chromosome. The strongest enrichment was observed with a 15 bp minimum length ($p = 0.000791$); mean per kb frequencies 1.83 in DSB-containing IGRs and 0.925 in IGRs without a DSB. When mismatches were allowed, mean frequencies were greater in DSB-containing IGRs than non-DSB IGRs with the differences statistically significant for size minima 10 through 23 bp with one mismatch allowed per 10 bp, and 11 through 26 bp with one mismatch allowed per 5 bp. We found no significant differences between the 6 DSB-containing ORFs and other ORFs on the chromosome.

Most of the 70 DSB-containing IGRs showed very low levels of DSBs (see Figure 2 in ref [42]), and 48 of them occurred outside hot spots reported in the genome-wide survey by Gerton and co-workers [24]. It is therefore likely that many of them indicate non-hot spot background recombination events, since these have been found to occur with low frequency between hot spots [3-5]. For PPTs of at least 15 bp, the mean frequency per kb was 1.70 in DSB-containing IGRs outside hot spots reported by Gerton *et al.* [24]. This was significantly greater than the mean per kb frequency of 0.925 found in IGRs without a DSB ($p = 0.00262$).

PPTs are associated with broad hot spot-containing regions

Because PPTs have been shown to stimulate formation of recombination intermediates at distances as great as 4000 bp [54], we wished to know if they were associated with broad hot spot-containing regions. In yeast, we compared frequencies of PPTs of at least 12 bp between IGRs and ORFs flanking hot spots, and IGRs and ORFs within cold spots. Comparisons were made up to a maximum distance of 4 IGRs from hot spots. IGRs one ORF removed from hot spots and IGRs two ORFs removed from hot

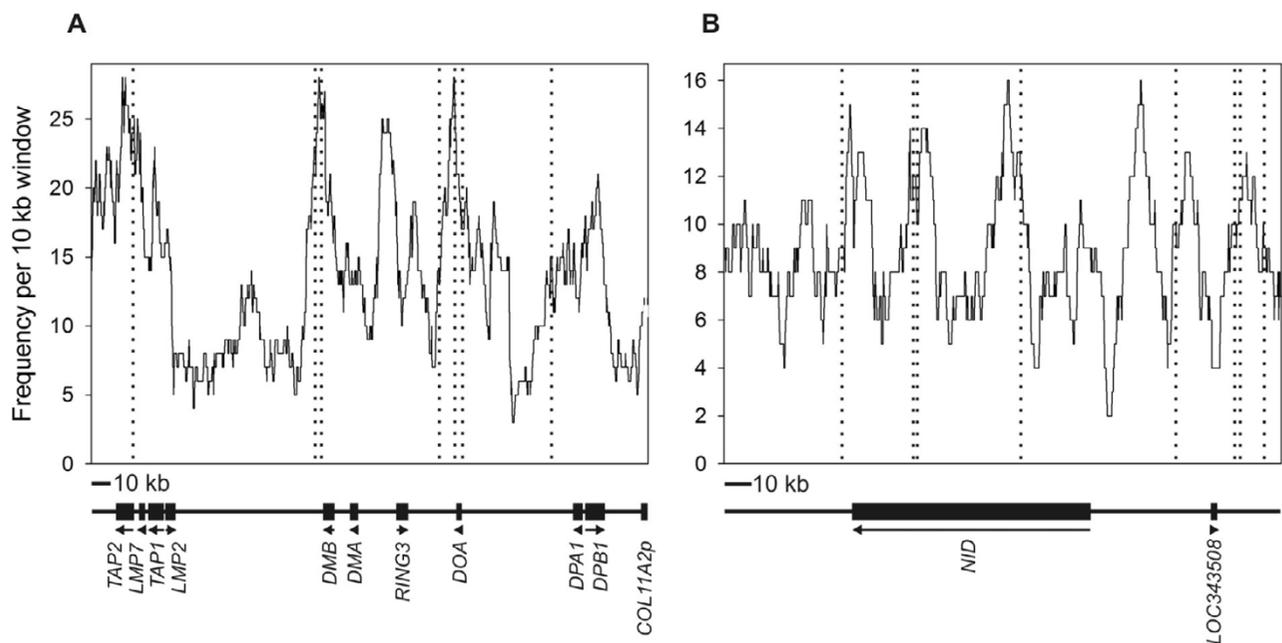


Figure 2
Densities of high GC-content PPTs relative to human hot spot locations. Sliding window plots of the densities of PPTs of at least 12 bp, with GC-contents above the mean for PPTs over the regions we investigated, relative to hot spot locations in A): a 292 kb region of the human MHC Class II region in which 7 hot spots are known and B): a 206 kb region of human chromosome I in which 8 hot spots are known. Vertical dotted lines represent hot spot mid point locations. Sliding windows were 10 kb wide and moved in steps of 100 bp. Locations of genes in the regions are shown below the plots with arrows indicating direction of transcription.

spots showed significantly higher PPT frequencies than cold spot IGRs, indicating a regional association between PPT frequency and hot spot-containing areas. The mean PPT frequency of 2.90 per kb in the PPT-enriched hot spot-flanking IGRs was significantly less, however, than the mean per kb frequency of 3.93 in hot spot IGRs. The mean distance encompassed by the hot spot-containing regions in which PPTs were enriched was just over 11.5 kb. We found no significant differences between ORFs flanking hot spots and cold ORFs.

In humans, we investigated PPT frequencies for all 17 hot spots in windows of increasing size centred on hot spot mid points, which we will refer to as hot regions. PPT frequencies in hot regions 3 to 40 kb wide were compared with remaining cold regions, which were defined as experimentally mapped cold regions lying outside these windows. We found no significant differences for PPTs of at least 12 bp, except when low GC-content tracts were excluded. PPTs of at least 12 bp with greater than the mean PPT GC-content had consistently higher frequencies in hot regions than remaining cold regions with the differences statistically significant for window sizes 3, 4, 9, 10, 12, 14 through 22, and 24 kb. Excluding the hot spot sequences themselves from the analysis weakened the

associations, but they remained significant for window sizes 4, 10, and 15 through 20 kb. Sliding window plots of the densities of high GC-content PPTs in the two regions in which multiple human hot spots have been mapped showed that peaks in density often occur within a few kb of hot spots. For PPTs of at least 12 bp, this was most striking with window sizes of about 10 kb (Figure 2). No significant regional associations were found for high GC-content PPTs in yeast.

The associations are not primarily due to microsatellite PPTs

One cause of the associations we observed might be a mutation bias resulting from recombination, or other properties of hot spot regions, acting to cause expansion of PPTs. With regard to this possibility, the degree to which the hot spot-associated PPTs consist of short, direct tandem repeats (STRs, or microsatellites) is relevant, because microsatellites have well-described mutational dynamics [55], but it is unclear how a mutation bias could act on non-repetitive PPTs. We therefore asked whether hot spots contained an elevated proportion of microsatellite PPTs. Using a separate search for PPTs of at least 12 bp consisting of short tandem repeats with a repeated unit 6 bp or less (PP-STRs), we found mean per kb frequencies in

yeast of 0.41 in hot spots and 0.13 in cold spots. Poly-A/T made up a large proportion of these, with mean per kb frequencies of 0.27 in hot spots and 0.059 in cold spots. After subtracting the number of PP-STRs of at least 12 bp from the number of PPTs of all kinds of at least 12 bp for each hot spot and cold spot, we found that the difference in mean per kb frequencies was still significant (1.51 in hot spots and 0.84 in cold spots, $p = 1.19 \times 10^{-6}$). In humans, PP-STRs of at least 12 bp were more frequent in cold regions than hot spots. This was also the case for poly-A/T considered separately. While not ruling out the possibility of a mutation bias, these results suggest that any bias that may be operating to cause an excess of PPTs in hot spot regions is probably not primarily due to insertions causing duplications of adjacent sequence, which is how microsatellites expand [55].

Sequence polymorphisms suggesting a function in hot spots for high GC-content poly-pu/py-rich sequences

At the time of writing, there are three known cases in humans of single nucleotide changes associated with altered recombination levels in hot spots [16,18,19]. All three polymorphisms were associated with several-fold reductions in recombination frequency and were located close to the estimated hot spot mid points. We found that each of these polymorphisms occurs within 3 bp of the end of a sequence 14 bp or longer consisting of 85% or more poly-pu/py and at least 70% G/C (Table 1).

Known features of hot spot recombination are similar across very diverse taxa (reviewed in [6]), so comparison of PPTs between humans and chimpanzees in the regions of human hot spots where there is no hot spot in chimpanzees might provide interesting data. To date, there are six cases in which the recombination activity of chimpanzee chromosomal regions orthologous to experimentally characterized human hot spots has been investigated, and in all these cases little or no evidence for hot spot activity was found in chimpanzees [15]. Because the human hot spot activity-associated polymorphisms all occurred within 166 bp of the hot spot mid points [16,18,19], we aligned the human and chimpanzee sequences 166 bp on either side of the human hot spot mid points. We found that all five of the hot spot central regions for which alignments are possible contain differences within sequences

14 bp or longer consisting of more than 80% poly-pu/py and at least 50% G/C (Table 2). In three out of the six sequences, there were differences within 3 bp of their ends, and four out of six contained a difference in the terminal 4 bp. The analysis did not include the DNA2 hot spot, because the publicly available chimpanzee sequence was missing 206 of the 332 bp region centred on the hot spot mid point, including the mid point itself. The DMB2 hot spot was included in the analysis, but the chimpanzee sequence was missing 34 bp from the central region. Within these limitations, 11 out of the total of 43 differences within 166 bp of the human hot spot mid points occurred within high GC-content, poly-pu/py-rich sequences as defined above. Over all the central regions we found 9 such sequences, 6 of which contained at least one difference between humans and chimpanzees. Of the 33 sequences of this type located within hot spots but outside the central regions, 10 contained differences. The differences were probably not due to recombination or other features of hot spots preferentially causing mutations in this type of sequence, because similarity between the two species in poly-pu/py-rich sequences as defined was 96.73% in hot spots, which was only slightly lower than the 97.27% overall sequence similarity for the 5 hot spots.

The observation that all three human hot spot recombination frequency-associated polymorphisms known at the time of writing occur in poly-pu/py-rich high GC-content sequences suggests that certain sequences of this type may be involved in determining hot spot activity. This is somewhat supported by the similar sequence contexts found for the human/chimpanzee differences. The particular PPTs that we have identified as possibly having functional roles in hot spots (Tables 1 and 2) might prove to be useful targets for experimental studies. Secondary structure formation and protein binding by the sequences are consistent with the data, and could be investigated further.

Although the sequence requirements for poly-pu/py-rich sequences to form secondary structures are not fully understood [52,53,56], the structures can tolerate a substantial proportion of interruptions to the homopurine/homopyrimidine motif [34,36,53], and can also be sensitive to single nucleotide changes [38,39]. They have been shown to stimulate inter-plasmid recombination [33] and

Table 1: Poly-pu/py-rich sequence contexts of polymorphisms associated with hot spot activity in humans. Sequence contexts of the three polymorphisms associated with reduced recombination frequencies in human hot spots. The recombination-suppressing alleles are shown in lower case.

Hot Spot	References	Sequence context	Distance from hot spot mid point (bp)
MS32	[19]	(G/c)GTGGGAAGGGTGG	151
NID1	[3, 18]	CC(C/t)CCCACCCCCACCCC	64
DNA2	[4, 16]	AGGGGGCAGCAACAGGG(A/g)GG	166

Table 2: Poly-pu/py-rich sequence contexts of differences between humans and chimpanzees in human hot spot central regions. Differences between humans and chimpanzees within 166 bp of human hot spot mid points occurring in sequences at least 14 bp long consisting of over 80% poly-pu/py and at least 50% G/C. In each case, the hot spots are absent in chimpanzees. The chimpanzee sequence is shown in lower case.

Hot Spot	References	Sequence context	Distance of difference from human hot spot mid point (bp)
Beta-Globin	[15, 45]	CCTTTTCCCCTCCTACCCTACTTT(C/t)T	133
DMB1	[4, 15]	GGGGGTGAGCAACAG(A/g)AA	35
DMB2	[4, 15]	GAGAGGAGAGAAGTGGAGTGGGAGAAAGAAGGAGAG(G/a)GACATAGAGAGAGGGGA(G/a)(G/a)GAGGGGAG(C/-)GG(-/a) GGGAAG/AGAGAGAGAGCTGGA	115, 98, 97, 88, 85
DMB2	[4, 15]	GGGAGGGGAG(G/t)TGGA	51
DNA3	[4, 15]	GGGAGGCC(A/g)AGG(C/a)GGG	23,27
TAP2	[4, 15]	(C/g)CTTCATTCCATCC	83

dimerization [57], as well as recombination hot spot activity in *E. coli* [58]. Immunocytological evidence shows that PPTs do form secondary structures on human chromosomes *in vivo* [59], but the possibility that this occurs in meiotic recombination hot spots has not to our knowledge been tested.

The additional, non-exclusive, possibility that certain poly-pu/py-rich sequences could contain binding motifs for proteins that act to stimulate recombination is supported by evidence that PPTs can bind transcription factors [50,60], since transcription factor binding is a determining factor in some yeast hot spots [20,21,26-28]. Binding of the nuclear matrix-associated type III intermediate filament proteins is also a possibility. Intramolecular quadruplex secondary structures, which can be formed by poly-pu/py-rich, GC-rich sequences [53], have been shown preferentially to bind these proteins *in vitro* [61].

Speculation on possible mechanisms by which PPTs could influence hot spot recombination

There are several possible mechanisms by which PPTs could act as functional components in recombination hot spots. It is conceivable that PPTs could mediate recombination by being themselves the sites of DSB formation, possibly as a result of nuclease action on secondary structures. This seems unlikely to be a common mechanism, however, since DSBs are found at many positions over 100–500 bp regions in yeast hot spots, and are probably position-specific rather than sequence-specific (reviewed in [2]). PPTs may also function in hot spots by altering local chromatin structure. This could result from their previously demonstrated involvement in the formation of DNase I hypersensitive sites, which are nucleosome-free regions of chromatin [62,63]. An open chromatin structure is one important factor in hot spots, presumably allowing access to the recombination machinery [7,8]. Binding of transcription factor proteins may help to achieve this in at least some hot spots, since opening up of chromatin structure is one function of transcription fac-

tors, and this has been proposed as the most likely reason for the requirement shown by some yeast hot spots for transcription factor binding (reviewed in [2]). The possible role of PPTs in hot spots may also be mediated by transcription factors, since PPTs in promoter regions can affect transcription, and evidence suggests that this is due to opening of the chromatin structure by the PPT *via* secondary structure formation and/or binding of transcription factor proteins (reviewed in [60,64]). Not all promoter regions are hot spots, however, so factors other than an open chromatin structure are clearly involved.

The potential of PPTs to cause replication pausing (reviewed in [65]) may also be relevant to their possible function in hot spots. A presently unexplained spatial and temporal coupling between DNA replication and meiotic recombination has been demonstrated in yeast [66], and it has been suggested that replication pausing could promote DSB formation *via* localized modification of histones [2]. Another property of PPTs suggesting a mechanism by which they could promote recombination is their ability to stick together, forming multi-stranded aggregates [41,67]. This suggests the possibility that they may help homologous chromosomes to align prior to meiotic recombination, and it has been proposed that this could be mediated by Hoogsteen base pairing interactions [67].

Conclusion

Our results, along with previous evidence for the unique biochemical properties and recombination-stimulating potential of poly-pu/py-rich sequences, suggest that the possible functional involvement of this type of sequence in meiotic recombination hot spots deserves further experimental exploration. Relevant tests include deleting PPTs from hot spot regions in model organisms and assaying the effects on recombination. Another possible approach would be to test the structure-forming and protein-binding capabilities of hot spot-associated PPTs *in vitro*.

Methods

We used pattern-matching algorithms programmed in C to search for PPTs and PP-STRs in genomic DNA sequence. Where possible, we used the sequence versions used in the studies that reported the hot spot locations. Yeast sequences and ORF locations were downloaded from the Stanford website [68]. The GenBank accession numbers for the 16 yeast chromosomes are [NC_001133](#) through [NC_001148](#). The GenBank accession numbers for the human hot spot sequences are: Beta-Globin hot spot: [GI:37541814](#), chromosome 1 hot spots: [GI:37549514](#), and *SHOX* hot spot: [U182668](#). For the MHC hot spots we used the 28 October 1999 version of the MHC class II region sequence, since that was the version to which the reported hot spot locations corresponded [4]. This version is available at the Sanger Centre website [69]. The GenBank accession numbers for the chimpanzee sequences were as follows: *DNA3*: [NW_108387.1](#), Beta-Globin: [NW_113864.1](#), *DNA2*, *DMB1* *DMB2* and *TAP2*: [NW_107937.1](#).

PPTs overlapping hot and cold regions were excluded. When we investigated PPT frequencies in windows with increasing length centred on human hot spots, some of the windows included areas for which recombination rates are unknown, and these were excluded from the analysis. Areas where windows overlapped were combined. When we searched for sequences with mismatches to the poly-pu/py motif we disallowed mismatches in the terminal two bp of tracts. When we searched for sequences at least 80% poly-pu/py rich, we imposed the additional restriction that no 6 bp segment of any sequence was allowed more than two mismatches. Alignments were performed using the BLAST algorithm [70].

Statistical comparison of means (Student's T-test and Mann-Whitney U Test, 2-tailed tests in call cases) and correlation analyses (Spearman's Rank Test) were done using SPSS with significance inferred where $p < 0.05$ in all cases. All samples were initially tested for normality (Shapiro-Wilk Test) and significantly non-normal samples were subjected only to the non-parametric tests. In cases where yeast hot spots contained multiple ORFs, we used the mean reported ORF rank [24] for the correlation analysis of hot spot intensity and PPT frequency.

Authors' contributions

ATMB conceived and designed the experiments, analyzed the data and wrote the paper. JPWP wrote the computer programmes. NJG contributed to the interpretation of the data and the writing of the paper.

Acknowledgements

ATMB has been supported by a University of Canterbury PhD scholarship and a Royal Society of New Zealand Marsden grant (UOC 202). NJG's

work is supported by a Royal Society of New Zealand Marsden grant (UOC 202).

References

1. Kauppi L, Jeffreys AJ, Keeney S: **Where the crossovers are: recombination distributions in mammals.** *Nat Rev Genet* 2004, **5(6)**:413-424.
2. Petes TD: **Meiotic recombination hot spots and cold spots.** *Nat Rev Genet* 2001, **2(5)**:360-369.
3. Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P: **Human recombination hot spots hidden in regions of strong marker association.** *Nat Genet* 2005, **37(6)**:601-606.
4. Jeffreys AJ, Kauppi L, Neumann R: **Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex.** *Nat Genet* 2001, **29(2)**:217-222.
5. Kauppi L, Stumpf MP, Jeffreys AJ: **Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region.** *Genomics* 2005, **86(1)**:13-24.
6. Nishant KT, Rao MR: **Molecular features of meiotic recombination hot spots.** *Bioessays* 2006, **28(1)**:45-56.
7. Wu TC, Lichten M: **Meiosis-induced double-strand break sites determined by yeast chromatin structure.** *Science* 1994, **263(5146)**:515-518.
8. Shenkar R, Shen MH, Arnheim N: **DNase I-hypersensitive sites and transcription factor-binding motifs within the mouse E beta meiotic recombination hot spot.** *Mol Cell Biol* 1991, **11(4)**:1813-1819.
9. Sun H, Treco D, Schultes NP, Szostak JW: **Double-strand breaks at an initiation site for meiotic gene conversion.** *Nature* 1989, **338(6210)**:87-90.
10. Steiner WW, Schreckhise RW, Smith GR: **Meiotic DNA breaks at the *S. pombe* recombination hot spot M26.** *Mol Cell* 2002, **9(4)**:847-855.
11. Haring SJ, Halley GR, Jones AJ, Malone RE: **Properties of natural double-strand-break sites at a recombination hotspot in *Saccharomyces cerevisiae*.** *Genetics* 2003, **165(1)**:101-114.
12. Wu TC, Lichten M: **Factors that affect the location and frequency of meiosis-induced double-strand breaks in *Saccharomyces cerevisiae*.** *Genetics* 1995, **140(1)**:55-66.
13. Borde V, Wu TC, Lichten M: **Use of a recombination reporter insert to define meiotic recombination domains on chromosome III of *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1999, **19(7)**:4832-4842.
14. Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Paabo S: **Fine-scale recombination patterns differ between chimpanzees and humans.** *Nat Genet* 2005, **37(4)**:429-434.
15. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrou P, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D: **Comparison of fine-scale recombination rates in humans and chimpanzees.** *Science* 2005, **308(5718)**:107-111.
16. Jeffreys AJ, Neumann R: **Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot.** *Nat Genet* 2002, **31(3)**:267-271.
17. Schuchert P, Langsford M, Kaslin E, Kohli J: **A specific DNA sequence is required for high frequency of recombination in the *ade6* gene of fission yeast.** *Embo J* 1991, **10(8)**:2157-2163.
18. Jeffreys AJ, Neumann R: **Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot.** *Hum Mol Genet* 2005, **14(15)**:2277-2287.
19. Jeffreys AJ, Murray J, Neumann R: **High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot.** *Mol Cell* 1998, **2(2)**:267-273.
20. Kon N, Krawchuk MD, Warren BG, Smith GR, Wahls WP: **Transcription factor Mts1/Mts2 (Afl1/Pcr1, Gad7/Pcr1) activates the M26 meiotic recombination hotspot in *Schizosaccharomyces pombe*.** *Proc Natl Acad Sci U S A* 1997, **94(25)**:13765-13770.
21. Wahls WP, Smith GR: **A heteromeric protein that binds to a meiotic homologous recombination hot spot: correlation of binding and hot spot activity.** *Genes Dev* 1994, **8(14)**:1693-1702.
22. Steiner WW, Smith GR: **Optimizing the Nucleotide Sequence of a Meiotic Recombination Hotspot in *Schizosaccharomyces pombe*.** *Genetics* 2005.

23. Zahn-Zabal M, Lehmann E, Kohli J: **Hot spots of recombination in fission yeast: inactivation of the M26 hot spot by deletion of the ade6 promoter and the novel hotspot ura4-aim.** *Genetics* 1995, **140(2)**:469-478.
24. Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD: **Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci U S A* 2000, **97(21)**:11383-11390.
25. Kobori JA, Strauss E, Minard K, Hood L: **Molecular analysis of the hotspot of recombination in the murine major histocompatibility complex.** *Science* 1986, **234(4773)**:173-179.
26. White MA, Dominska M, Petes TD: **Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci U S A* 1993, **90(14)**:6621-6625.
27. Mieczkowski PA, Dominska M, Buck MJ, Gerton JL, Lieb JD, Petes TD: **Global analysis of the relationship between the binding of the Bas1p transcription factor and meiosis-specific double-strand DNA breaks in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 2006, **26(3)**:1014-1027.
28. Steiner WW, Smith GR: **Natural meiotic recombination hot spots in the *Schizosaccharomyces pombe* genome successfully predicted from the simple sequence motif M26.** *Mol Cell Biol* 2005, **25(20)**:9054-9062.
29. Schultes NP, Szostak JW: **A poly(dA.dT) tract is a component of the recombination initiation site at the ARG4 locus in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1991, **11(1)**:322-328.
30. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P: **A fine-scale map of recombination rates and hotspots across the human genome.** *Science* 2005, **310(5746)**:321-324.
31. Kong A, Gudbjartsson DF, Sainz J, Jonsson GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K: **A high-resolution recombination map of the human genome.** *Nat Genet* 2002, **31(3)**:241-247.
32. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ: **Comparative recombination rates in the rat, mouse, and human genomes.** *Genome Res* 2004, **14(4)**:528-538.
33. Rooney SM, Moore PD: **Antiparallel, intramolecular triplex DNA stimulates homologous recombination in human cells.** *Proc Natl Acad Sci U S A* 1995, **92(6)**:2141-2144.
34. Dayn A, Samadashwily GM, Mirkin SM: **Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization.** *Proc Natl Acad Sci U S A* 1992, **89(23)**:11406-11410.
35. Maiti AK, Brahmachari SK: **Poly purine.pyrimidine sequences upstream of the beta-galactosidase gene affect gene expression in *Saccharomyces cerevisiae*.** *BMC Mol Biol* 2001, **2(1)**:11.
36. Raghavan SC, Chastain P, Lee JS, Hegde BG, Houston S, Langen R, Hsieh CL, Haworth IS, Lieber MR: **Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation.** *J Biol Chem* 2005, **280(24)**:22749-22760.
37. Kohwi Y, Kohwi-Shigematsu T: **Altered gene expression correlates with DNA structure.** *Genes Dev* 1991, **5(12B)**:2547-2554.
38. Bacolla A, Ulrich MJ, Larson JE, Ley TJ, Wells RD: **An intramolecular triplex in the human gamma-globin 5'-flanking region is altered by point mutations associated with hereditary persistence of fetal hemoglobin.** *J Biol Chem* 1995, **270(41)**:24556-24563.
39. Boles TC, Hogan ME: **DNA structure equilibria in the human c-myc gene.** *Biochemistry* 1987, **26(2)**:367-376.
40. Hanvey JC, Klysik J, Wells RD: **Influence of DNA sequence on the formation of non-B right-handed helices in oligopyrimidine.oligopyrimidine inserts in plasmids.** *J Biol Chem* 1988, **263(15)**:7386-7396.
41. Kiyama R, Nishikawa N, Oishi M: **Enrichment of human DNAs that flank poly(dA).poly(dT) tracts by triplex DNA formation.** *J Mol Biol* 1994, **237(2)**:193-200.
42. Baudat F, Nicolas A: **Clustering of meiotic double-strand breaks on yeast chromosome III.** *Proc Natl Acad Sci U S A* 1997, **94(10)**:5213-5218.
43. Jeffreys AJ, Ritchie A, Neumann R: **High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot.** *Hum Mol Genet* 2000, **9(5)**:725-733.
44. Schneider JA, Peto TE, Boone RA, Boyce AJ, Clegg JB: **Direct measurement of the male recombination fraction in the human beta-globin hot spot.** *Hum Mol Genet* 2002, **11(3)**:207-215.
45. Wall JD, Frisse LA, Hudson RR, Di Rienzo A: **Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates.** *Am J Hum Genet* 2003, **73(6)**:1330-1340.
46. May CA, Shone AC, Kalaydjieva L, Sajantila A, Jeffreys AJ: **Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene SHOX.** *Nat Genet* 2002, **31(3)**:272-275.
47. Yi S, Li WH: **Molecular Evolution of Recombination Hotspots and Highly Recombining Pseudoautosomal Regions in Homioids.** *Mol Biol Evol* 2005.
48. Verhoeven KJ, Simonsen KL: **Genomic Haplotype Blocks May Not Accurately Reflect Spatial Variation in Historic Recombination Intensity.** *Mol Biol Evol* 2005, **22(3)**:735-740.
49. Matsugami A, Ohashi K, Kanagawa M, Liu H, Kanagawa S, Uesugi S, Katahira M: **New quadruplex structure of GGA triplet repeat DNA--an intramolecular quadruplex composed of a G:G:G tetrad and G:(A):G:(A):G:(A):G heptad, and its dimerization.** *Nucleic Acids Res Suppl* 2001:271-272.
50. Sandaltzopoulos R, Mitchelmore C, Bonte E, Wall G, Becker PB: **Dual regulation of the *Drosophila* hsp26 promoter in vitro.** *Nucleic Acids Res* 1995, **23(13)**:2479-2487.
51. Hanvey JC, Shimizu M, Wells RD: **Intramolecular DNA triplexes in supercoiled plasmids. II. Effect of base composition and noncentral interruptions on formation and stability.** *J Biol Chem* 1989, **264(10)**:5950-5956.
52. Simonsson T: **G-quadruplex DNA structures--variations on a theme.** *Biol Chem* 2001, **382(4)**:621-628.
53. Dai J, Dexheimer TS, Chen D, Carver M, Ambrus A, Jones RA, Yang D: **An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution.** *J Am Chem Soc* 2006, **128(4)**:1096-1098.
54. Biet E, Sun JS, Dutreix M: **Stimulation of D-loop formation by polypurine/polypyrimidine sequences.** *Nucleic Acids Res* 2003, **31(3)**:1006-1012.
55. Ellegren H: **Microsatellites: simple sequences with complex evolution.** *Nat Rev Genet* 2004, **5(6)**:435-445.
56. Radhakrishnan I, Patel DJ: **DNA triplexes: solution structures, hydration sites, energetics, interactions, and function.** *Biochemistry* 1994, **33(38)**:11405-11416.
57. Hampel KJ, Burkholder GD, Lee JS: **Plasmid dimerization mediated by triplex formation between polypyrimidine-polypurine repeats.** *Biochemistry* 1993, **32(4)**:1072-1077.
58. Napierala M, Dere R, Vetcher A, Wells RD: **Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron I of the Friedreich's ataxia gene.** *J Biol Chem* 2004, **279(8)**:6444-6454.
59. Ohno M, Fukagawa T, Lee JS, Ikemura T: **Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies.** *Chromosoma* 2002, **111(3)**:201-213.
60. Lu Q, Teare JM, Granok H, Swede MJ, Xu J, Elgin SC: **The capacity to form H-DNA cannot substitute for GAGA factor binding to a (CT)n*(GA)n regulatory site.** *Nucleic Acids Res* 2003, **31(10)**:2483-2494.
61. Tolstogonov GV, Li G, Shoeman RL, Traub P: **Interaction in vitro of type III intermediate filament proteins with higher order structures of single-stranded DNA, particularly with G-quadruplex DNA.** *DNA Cell Biol* 2005, **24(2)**:85-110.
62. Elgin SC: **The formation and function of DNase I hypersensitive sites in the process of gene activation.** *J Biol Chem* 1988, **263(36)**:19259-19262.
63. Gross DS, Garrard WT: **Nuclease hypersensitive sites in chromatin.** *Annu Rev Biochem* 1988, **57**:159-197.
64. Raghavan S, Burma PK, Brahmachari SK: **Positional preferences of polypurine/polypyrimidine tracts in *Saccharomyces cerevisiae* genome: implications for cis regulation of gene expression.** *J Mol Evol* 1997, **45(5)**:485-498.
65. Rao BS: **Regulation of DNA replication by homopurine/homopyrimidine sequences.** *Mol Cell Biochem* 1996, **156(2)**:163-168.
66. Baudat F, Keeney S: **Meiotic recombination: Making and breaking go hand in hand.** *Curr Biol* 2001, **11(2)**:R45-8.

67. Sen D, Gilbert W: **Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis.** *Nature* 1988, **334(6180)**:364-366.
68. <ftp://genome-ftp.stanford.edu/pub/yeast/>.
69. http://www.sanger.ac.uk/HGP/Chr6/MHC_991021.fasta.
70. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

