Recombination rates in most organisms do not vary uniformly with physical distance. This heterogeneity in the recombination rates is largely a consequence of different densities and intensities of recombination hot spots, which are distributed nonrandomly in the genome (Yu et al., 2001). In mice and humans, regions are generally termed as hot spots if their recombination frequency significantly exceeds 1 cM / Mb. Identifying the location and intensity of such hot spots is important to get a better understanding of meiotic recombination in mammals. Recombination hot spots also influence haplotype structure and diversity and therefore their characterization is important for the design and analysis of association studies between markers and loci contributing to phenotypic traits. Most of our understanding of recombination hot spots have come from studies of lower eukaryotes, in particular *Saccharomyces cerevisiae*. The identification of recombination hot spots in mammals has lagged behind for a long time due to experimental difficulties. Factors influencing mammalian recombination hot spot activity have also remained largely elusive. In the mouse, a cluster of only four recombination hot spots in the MHC regions have been characterized through genetic crosses. These include the *Psmb9* hot spot (Mizuno, 1996), the hot spot located in the *Eb* second intron (Zimmerer and Passmore, 1991), the *Pb* hot spot (Isobe et al., 2002) and the hot spot located in the *Ea* fourth intron (Khambata et al., 1996). A few more hot spots are known in humans through sperm typing and population genetics based approaches and based on the distribution of hot spots in the human MHC region and from recent genome wide scale analysis of hot spot activity (Crawford et al., 2004; McVean et al., 2004) it has been speculated that there could be a hot spot in every 60 – 90 kb of DNA giving rise to almost 50,000 recombination hot spots in the human genome (Kauppi et al., 2004). Earlier work from our laboratory while characterizing the DNA repair sites of pachytene spermatocytes, had resulted in the isolation of a 1.3 kb DNA fragment from the rat genome (Ramachandra and Rao, 1994) which contained several recombination potentiating sequences, that is highly conserved in most mammalian species. This 1.3 kb meiotic repair site has been used a probe to identify homologous sequences from the mouse and human genomes which have subsequently been analyzed for meiotic recombination activity. We decided to use the mouse and humans as model systems since
only a few meiotic hot spots are known and they offer a robust genetic system which facilitates the detection of recombination events.

**Characterization of a mouse meiotic recombination hot spot**

In the mouse, probes derived from the 1.3 kb sequence were used to screen a mouse genomic library which identified a larger 17.2 kb fragment flanking the mouse 1.3 kb sequence. This was localized to chromosome 8 C-D using mouse–hamster radiation hybrid panels. Localized SNP diversity was observed in a 1 kb interval from 13.3 to 14.3 kb in the 17.2 kb fragment from which 31 parsimony informative polymorphic markers were identified. Using these polymorphic markers, Linkage disequilibrium (LD) analysis was performed with 40 different mice strains of the *Mus.musculus* species. D’ values were seen to drop to 0.77 to 0.79 for 15 of the statistically significant 239 pairwise associations showing presence of recombinational activity. These corresponded to a region between 13584 bp to 13847 bp. The region showing localized break down of LD also showed peak levels of gene conversion indicating the active role played by conversion in shaping patterns of LD in this region. Twenty fragment pairs showing gene conversion were detected between different strains with the size of the conversion tract ranging from 255 bp to 627 bp. The average size of the conversion tract was around 371 bp which is comparable to the average conversion tract length of 480 bp reported earlier for the *Psmb 9* locus by Guillon and de Massy, (2002). In order to quantify the extent of recombination we calculated the recombination parameter R which is related to LD, for this 1 kb interval. Initially, a model free approach involving the four gamete test was used and the minimum number of recombination events ($R_M$) for the 1 kb interval was estimated to be two. Since $R_M$ is an underestimate (Hudson and Kaplan, 1985), a coalescent model based approach involving simulations of the history of the sample for different values of $4Nr$ that give the observed $R_M$ of 2 was used to calculate R. Such simulations performed a 1000 times for different values of $4Nr$ gave a value of $R = 0.00257$ / bp. Since no estimates for N for the mouse population were available in the literature, using two most plausible values of $10^3$ and $10^4$, the recombination intensity value for this locus was calculated to be around 10-100 fold higher than the genome
Chapter 4: Summary

average. The 17.2 kb locus at chromosome 8 C-D therefore, defines a new meiotic recombination hot spot in the mouse. This is also the first meiotic hot spot to be identified in the mouse outside the MHC complex.

Comparison of the properties of this hot spot with the mouse Ea, Eb, Psmb9 and Pb MHC recombination hot spots showed the presence of shared sequence motifs which include an LTR element and HVMS sequence. It was also interesting to see during in silico analysis that the crossover sites in the 17.2 kb locus along with the Ea, Psmb9 and Pb MHC recombination hot spots, correlated very well with a region showing MAR (matrix associated region) potential. In order to experimentally verify such an association between crossover sites and the nuclear matrix, we prepared matrix and loop DNA from mice testis and determined the relative enrichment of these sites in MAR DNA as compared to loop DNA by a PCR based assay (Kramer and Krawetz, 1997). This identified MARs as a novel feature of mouse meiotic recombination hot spots by demonstrating their presence at the cross over regions for the Ea, Psmb9, Pb and the hot spot in the 17.2 kb locus. Many recombination hot spots require the binding of transcription factors or transcriptional activity, which creates an open chromatin structure for access to the recombination machinery. A novel 2.4 kb non protein-coding RNA transcript was observed in the 17.2 kb hot spot locus which was expressed in kidney, liver, spleen and testis but not in brain, heart, lung and muscle. The presence of transcriptional activity close to a recombination hot spot region suggests that the 17.2 kb hot spot has properties of α hot spots (hot spots associated with transcription factor binding / transcribed regions). The hot spot in the 17.2 kb locus is also seen to be at the 3' end of the transcribed region as observed for the MHC hot spots which are found at the 3' end (Psmb9) or in the introns (Ea, Eb, Pb), but not at the 5' end of the genes as in Saccharomyces cerevisiae. The presence of a MAR domain having an intrinsically open chromatin structure in proximity to the crossover site shows that the 17.2 kb hot spot also has properties of β hot spots (hot spots associated with nucleosome excluding sequences). However the moderate percentage (38.57%) of G + C content rules out a γ hot spot (hot spots found in GC rich regions) like nature. It was also seen that the recombination / gene
conversion site is around 10 kb downstream of the 1.3 kb region which was identified earlier as a meiotic repair site (Ramachandra and Rao, 1994). It is likely that the 1.3 kb sequence probably acts as a cis acting sequence to stimulate recombination activity in the surrounding genomic regions which is another feature that this hot spot shares with the \textit{Psmb9} hot spot where hot spot activity is similarly influenced by cis acting sequences.

**Analysis of a human subtelomeric meiotic recombination hot spot**

In humans, the 1.3 kb sequence was mapped to the subtelomeric region of chromosome 3p 26.3 using human – hamster radiation hybrid panels, which was seen to be a recombination jungle that may comprise several recombination hot spots based on the recombination maps for chromosome 3. A 10.5 kb sequence showing homology to the 1.3 kb sequence was cloned from this region. Fifty parsimony informative polymorphic markers were identified in the 10.5 kb locus by screening a South Indian population sample of 49 individuals. LD analysis using these markers showed 58.8 % pairwise marker combinations in the South Indian population where D' was less than 0.9 signifying the presence of recombination exchanges. Further analysis, through estimation of the recombination parameter R, identified a meiotic recombination hot spot around 7.5 kb sequence interval in the 10.5 kb locus with intensity 7016 fold higher than the genome average. The hot spot was also seen to be conserved in a global sample of 32 populations at a similar position (7-8 kb). Significant association with hot spot activity was observed for the 18 bp human replication origin consensus sequence, MAR elements and LTR elements. The hot spot region also showed high nucleotide diversity both in the South Indian and global populations, and reduced diversity and more of low frequency singleton variants outside the hot spot. The background recombination rates at the subtelomeric 10.5 kb locus were seen to be 575 fold higher than the genome average values. This is in contrast to the background recombination rates that have been calculated for the MHC region (0.49 cM/Mb, Cullen \textit{et al.}, 2002) in humans which are lower than the genome average values (1.13 cM/Mb).
The study also examined whether there is a concordance of haplotype block structure and haplotype diversity with recombination rate variation (Greenwood et al., 2004). For the total South Indian population comprising 49 individuals we observed 98 haplotypes, which could be grouped into three major blocks with a predominant block boundary around 7 kb overlapping with the hot spot region and a minor one around 2.2 kb. When the haplotype structure in the global population sample for the 7 – 8 kb region corresponding to the hot spot was examined, 61 haplotypes for 32 individuals were seen with absence of any blocks in the region corresponding to the hot spot around 7.5 kb. However, the region outside the hot spot domain from 9 – 10 kb had a single block for the entire sequence interval with very little haplotype diversity. Only seven haplotypes could be seen for the 32 individuals out of which a single cosmopolitan haplotype had a chromosome coverage of 81.2%. This high degree of correlation between recombination rate heterogeneity and haplotype structure both within and across populations at the 10.5 kb locus shows the dominant role of recombination hot spots in shaping haplotype blocks and diversity through its overriding effects on population history in shaping LD patterns and localized SNP diversity (Reich et al., 2002; Kauppi et al., 2003). Establishing the generality of haplotype blocks constructed using one population is an important problem. Results from this study suggests that if recombination hot spots are shared across populations, then haplotype blocks flanked by these hot spots are likely to be common.

Subtelomeric regions of all human chromosomes generally show higher recombination rates in males than in females (Broman et al., 1998) and this is also true for the 3p26.3 region. However, the population genetics based approach used in the present study gives only sex-averaged recombination rate estimates for the 10.5 kb locus, since it involves sampling of both male and female chromosomes. The few sex-specific hot spots identified in humans so far by pedigree analysis, that allows determination of sex-specific recombination frequencies, include the ones observed at the Xp / Yp pseudoautosomal region (Rouyer et al., 1986) the IGF2-H19 region on chromosome 11p15.5 (Paldi et al., 1995) and on chromosome 15q11 – q13 (Robinson and Lalande,
1995). It is interesting to see that the only other subtelomeric hot spot known in humans on chromosome 16p13.3 is also sex-specific (Badge *et al.*, 2000), but at present it cannot be determined whether the same is true for the subtelomeric hot spot described in this paper and it needs to be analyzed through pedigree studies. We are presently addressing this question in more detail.