INTRODUCTION

Zinc finger (ZNF) genes represent one of the largest gene families in the human genome with an estimated 500-600 members (Becker et al. 1995; Hoovers et al. 1992; Klug and Schwabe 1995). Although the specific function of the majority of ZNF genes remains largely unknown, as a class they are believed to encode transcriptional regulators which in a few instances have been shown to play critical roles in cellular and developmental differentiation processes (Pieler and Bellefroid 1994). Zinc finger proteins have been implicated in many diverse eukaryotic developmental processes, such as segment pattern formation in the *Drosophila* embryo (Rosenberg et al. 1986); cellular proliferation in the cerebellar hindbrain of the mouse (Wilkinson et al. 1989); and hematopoietic differentiation among human myeloid precursor cells (Hromas et al. 1991). DNA binding of the encoded proteins is typically mediated by a zinc finger motif that consists either of two cysteines and two histidines (Krüppel family or C_2/H_2 type) or four cysteines alone (steroid receptor or C_2/C_2 type). The conserved cysteines and/or histidines form a tetrahedral complex around a zinc metal ion, generating a folded loop or "finger" of 30 amino-acids which is capable of making contact with DNA (Miller et al. 1985). The number of zinc finger motifs is highly variable, especially among the C_2/H_2 type, ranging from two to forty copies in different members of this family (Bellefroid et al. 1991).

The estimated 500 ZNF genes map to a variety of human chromosomes**.** Fluorescence *in situ* hybridization (FISH) on metaphase chromosomes using various zinc finger cDNAs as probes revealed a clustered organization of these genes on chromosomes 7, 9, 10, 12, 16 and 19 (Huebner et al. 1991; Jackson et al. 1996; Rousseau-Merck et al. 1993; Tommerup and Vissing 1995). This clustered organization suggests that tandem duplications were primarily responsible for increasing copy number of this gene family (Thiesen et al. 1991). Chromosome 19 appears to be particularly enriched for zinc finger genes with ZNF loci distributed within three gene clusters corresponding to cytogenetic band locations 19p12, q13.2 and q34 (Bellefroid et al. 1993; Shannon et al. 1996). A survey of chromosome 19 C_2/H_2 type zinc finger genes reveals that the majority encode an evolutionarily conserved protein motif of 75 aa, termed the Krüppel-associated box or KRAB domain (Bellefroid et al. 1993; Thiesen et al. 1991). The KRAB domain has recently been shown to function as a critical domain for protein-protein interaction (Friedman et al. 1996; Kim et al. 1996)**.** Although these genes may theoretically be involved in transcriptional activation or repression, all KRAB ZNF genes studied to date have been shown to act only as potent transcriptional repressors (Margolin et al. 1994; Pengue et al. 1995; Vissing et al. 1995; Witzgall et al. 1994)**.**

Based on relatively few examples, the exon/intron structure of the KRAB subfamily of C_2/H_2

type ZNF genes appears to be highly conserved (Baban et al. 1996; Bellefroid et al. 1993; Derry et al. 1995; Grondin et al. 1996; Villa et al. 1993). Members of this gene subfamily characteristically consist of four exons. The first three exons are relatively small and contain the 5' UTR and the KRAB domain, which is split by a single small intron. The last exon contains the spacer region, the tandemly repeated zinc finger motif, and the untranslated portion of the KRAB ZNF gene. Recent reports have suggested the existence of an additional 5' UTR exon consisting of human endogenous retroviral sequences which occasionally appear in alternatively spliced KRAB ZNF transcripts (Baban et al. 1996; Di Cristofano et al. 1995). The spacer region of the KRAB ZNF gene is highly variable and is often used to further categorize different members of this subfamily (Bellefroid et al. 1993). There are an estimated 40 closely related KRAB zinc finger loci clustered on 19p12 (Bellefroid et al. 1995; Bellefroid et al. 1993), spanning a genomic distance of about four megabases. To date, only eight ZNF cDNAs that map generally to the 19p12 band location (ZNF20, 43, 85, 90, 91, 92, 93 and 94) have been identified and characterized (Bellefroid et al. 1993; Bellefroid et al. 1991; Lichter et al. 1992). Northern blot analysis with a few 19p12 cDNAs indicates low levels of transcription in multiple tissues, including undifferentiated myeloid cells as well as embryonal carcinoma cell lines (Bellefroid et al. 1993).

Southern and Northern "zooblot" comparisons using 19p12 ZNF spacer-encoding probes detect no cross-hybridization signals in either mouse or hamster (Bellefroid et al. 1995; Bellefroid et al. 1993). These observations have been confirmed by FISH analysis, using ZNF91 YAC pools as probes, on chromosomal metaphase spreads from a variety of primate species, which indicates that ZNF91 synteny in this region of chromosome 19 does not extend to prosimians and rodents (Bellefroid et al. 1995). These data suggest that the emergence and expansion of the 19p12 cluster of KRAB ZNF genes have occurred relatively recently during primate evolution (~64 mya) (Bellefroid et al. 1995). This has led to the supposition that the function of the ZNF91 gene family can not be involved in fundamental developmental processes common to the anthropoid and prosimian primate lineages, rather that their function may be related more to chromatin modulation as opposed to gene-specific transcriptional regulation (Bellefroid et al. 1995). The recent identification of a KRAB corepressor protein (KAP-1 or KRIP-1) that contains protein domains (PHD, bromo and ring-finger motifs) common to eukaryotic chromatin-modulating genes gives some support to this hypothesis (Friedman et al. 1996; Kim et al. 1996). There are relatively few examples of recent evolutionary expansions of gene families in the primate lineage (Eichler et al. 1997; Eichler et al. 1996; Regnier et al. 1997; Teglund et al. 1994; Zimonjic et al. 1997). It is likely that specific molecular mechanisms exist for the propagation and expansion of selectively favored gene families in any genome. In order to investigate more directly the molecular basis of the recent expansion of the ZNF91 gene cluster in 19p12, a detailed BAC/cosmid physical map was constructed in a 700 kb interval of this cluster and large-scale sequence analysis was performed within two selected regions. Our analysis provides the first detailed insight into the genomic architecture of the KRAB ZNF gene family and implicates higher-order beta-satellite repeat structures in the expansion of this gene family in the anthropoid genome.

RESULTS

A. Construction of a (700kb) physical map between D19S454 and D19S269.

We developed a highly integrated physical map of cosmid and BAC clones between markers D19S454 and D19S269 (Mohrenweiser et al. 1996) using a combined approach of FISH, STS markers and conversion of overlapping sets of YAC genomic clones to cosmid and later BAC clones within this interval (Figure 1). Four overlapping YAC clones (784E8, 411H1, 138D1 and 60D10) spanning an estimated 1.75 MB of this region were initially used as probes to screen a flow-sorted chromosome 19 arrayed cosmid library. A total of 85 cosmids were identified, assigned to bins and assembled into overlapping sets of cosmid clones based on fluorescent fingerprinting methods (Carrano et al. 1989; Trask et al. 1993). The location and relative position of each cosmid contig were confirmed using an STS screening strategy and FISH analysis of sperm pronuclei with representative cosmid clones from each contig as probes (Brandriff et al. 1994; Brandriff et al. 1991) (Figure 1b). Five sets of contiguous cosmid clones were initially identified within the D19S454/ D19S269 interval (Figure 1c), representing ~550 kb of the 700 kb interval (Figure 1). Using existing chromosome-19 cosmid clones as walking probes, larger-insert BAC clones which bridged adjacent but non-overlapping contigs were identified (see Materials and Methods and Figure 1). Comparative *EcoR*1 digests and subsequent BAC to cosmid hybridizations confirmed the position and orientation of each BAC clone within the region. Since the BAC and chromosome-19 specific cosmid libraries originate from two different genomic sources, comparative *EcoR*1 digests between these two sources served as an important check-and-balance in confirming the integrity of the genomic clones used in the construction of a physical map of this region. Two BAC/cosmid contigs were generated in this region (representing 300 and 500 kb of chromosome 19 sequence), separated by a gap of ~ 25 kb (Figure 1c), as determined by distance estimates using high-resolution FISH two-color FISH.

B. ZNF Orientation and Duplicon Size within 19p12.

Due to the reported highly conserved nature of ZNF genes in 19p12 (Bellefroid et al. 1993; Bellefroid et al. 1991), exon-specific PCR products corresponding to the KRABA, KRABB, spacer and ZNF domains of ZNF91 were developed and used as probes to screen nylon-transferred *EcoR*1 digests of portions of the D19S454/S269 interval (Figure 1c). All probes hybridized strongly to paralogous *EcoR*1 fragments within this region with the exception of the KRABB probe, which demonstrated variable degrees of hybridization signal intensity (data not shown). Using this cross-hybridization approach, it was possible to determine the position of ZNF genes in the region, their orientation and the distance separating paralogous ZNF genes. Within the D19S454/S269 region we identified at least four genes that are organized in a head-to-tail fashion. The direction of putative transcription of each gene is telomeric to centromeric. The average duplicon size for each gene in this region was approximately 180 kb (Figure 1c). In comparison, a second region was analysed which is located ~1 MB proximal to the D19S454/S269 region, abutting the alpha-satellite repeat region of the chromosome 19 centromere. Here, the distance between two adjacent ZNF genes was found to be smaller (~100 kb) suggesting variable compaction of ZNF genes across the 4 MB cluster (data not shown).

C. Comparative Sequence Analysis.

Two genomic clones within the D19S454/S269 region were selected for large scale sequence analysis: BAC 33152 (approximately 150 kb length) and cosmid 32532 (approximately 40 kb in length). Based on the physical map and earlier hybridization experiments, each clone contained a putative ZNF gene. These genes were separated by a distance of approximately 450 kb (Figure 1c). Random shotgun M13 libraries were prepared for each clone, subclones were sequenced and sequence data were assembled using PHRAP software. The actual inserts of BAC 33152 and cosmid 32532 were determined to be 165,199 bp and 43,768 bp, respectively for a total of 208,967 bp of 19p12 ZNF genomic sequence. The finished sequences for BAC33152 and cosmid 32532 have been deposited in GenBank under the accession numbers AC003973 and AC004004, respectively.

An overview of the comparative genomic organization and the ZNF gene structures of BAC33152 and cosmid 32532 is presented below (Table 1 and Figure 2). A combination of GRAIL analysis (version 1.3) and BLAST sequence similarity comparisons against cDNA sequence from ZNF gene family members was used to determine the most likely position of intron-exon boundaries within each clone. Cosmid 32532, due to the length of its insert, contained only three (KRABA, KRABB and spacer-ZNF exons) of the four exons (the fourth exon, predominantly 5’UTR, was not present in this clone). BAC 33152 contained a complete complement of ZNF exons. The two putative ZNF genes showed the greatest nucleotide sequence similarity (~83%) with two previously identified members of the ZNF 19p12 gene family, ZNF43 and ZNF91, and were designated ZNF208 (BAC 33152) and ZNF209 (cosmid 32532) in accordance with the GDB nomenclature committee. The gene structures of ZNF208 and ZNF209 are generally conserved and are similar to the ZNF91 model (Bellefroid et al. 1993), consisting of separate exons for KRABA and KRABB protein binding domains and a single large exon that incorporates the ZNF DNA-binding domain and spacer region. The sizes of the KRABA and KRABB exons (exons 2 and 3) are highly conserved between ZNF208 and 209 and are predicted to be both 126 and 95 bp in length respectively (Table 1). With the exception of exons 2 and 3, the gene structure of ZNF208 is more compact than ZNF209. The size of the last putative exon, for example, differs significantly. Exon 4 of ZNF209 contains 15 identifiable ZNF (28 amino-acid) repeats over 2.8 kb while the same exon from ZNF209 (from BAC 33152) is almost twice as large, 4.9 kb, with 41 ZNF repeat motifs. Similarly, the lengths of introns 2 and 3 are almost twice as long in ZNF208 as in ZNF209.

Using Miropeats, RepeatMasker and Dotter software (Parsons et al. 1993; Sonnhammer and Durbin 1995), we compared the non-genic organization of BAC33152 and cosmid 32532. Two 43 kb segments which corresponded to the complete insert of cosmid 32532 (43,768 bp, GenBank AC004004) and the paralogous segment of BAC33152 (64401-108,079 bp, GenBank AC003973) were masked for low-copy repeat sequences and the two sequences were compared using miropeats (Figure 2). This analysis identified six regions (~10.5 kb) of relatively high sequence similarity (77.4 - 82.0%) between the two duplicated segments of 19p12 (Figure 2). Not surprisingly, three of these regions corresponded to the positions of exon/introns of the putative ZNF genes. The other three regions, however, were located distal to the final exon and were not associated with any known genic or repeat sequences. An analysis of LINE and SINE retroposons between these two segments found very little conservation in the organization and subfamily identity of these repeat elements in the region, suggesting that the majority of Alu and L1 retroposition invasions occurred independently within the genomic context of these two ZNF cassettes. Only a single, fragmentary L1 element, (726 bp in length located in intron 3), showed complete conservation in orientation, position and identity between the two sequences. Sequence analysis of this LINE element show that it belongs to a relatively ancient subfamily (L1MB), which was predominantly active during the time of the mammalian radiation. BESTFIT (GCG software) alignment of non-coding paralogous segments between BAC33152 and cosmid 32532 revealed 79.4% sequence identity over a compared region of 8,041 bp (conserved segments I, II, IV, V and VI; Figure 2). This degree of sequence similarity, however, does not persist throughout the entire 43 kb segment, largely due to the differential organization of L1 and Alu repeat elements between the two regions. 83.1% sequence similarity was observed between the putative coding portions of ZNF208 and ZNF209.

D) Expression Analysis

BLAST sequence similarity searches with the putative ZNF genic portion of BAC33152 against the NCBI dbEST database identified a single EST from a human pregnant uterus cDNA library

(I.M.A.G.E. clone 501492) with 99.9% identity over 500 bp. Subsequent complete sequence analysis of the cDNA insert (809 bp) showed nearly 100% sequence identity between the cDNA clone and the genomic sequence of BAC33152. The cDNA sequence extends from the 5' UTR region (exon 1) through the KRABA and KRABB portions of ZNF208 (exons 2 and 3), terminating 470 bp distal to the transcriptional splice donor of exon 3a near an alternative poly A addition signal (AATATA) (exon 3b; Table 1). Translation of the ORF of EST501492 predicts a 77 amino-acid protein, consisting only of KRABA and B protein-interacting domains. Conceptual translation of exon 4 from the BAC clone, however, suggests an additional 3.7 kb of ZNF motifs that is not part of cDNA clone 501492. Three strong polyadenylation signals were identified near the terminal portion of the ZNF repeat motifs predicting a transcript of ~5.0 kb. Based on this analysis, two different ORFs are expected for ZNF208: a short peptide (77 aa) completely devoid of ZNF repeat motifs and a longer isoform (ORF= 3,951 bp/1317 aa) consisting of 41 ZNF (28 amino acid) repeats.

Northern analysis of 16 human tissues using EST 501492 PCR-amplified insert did not show strong signal hybridization to a single transcript. Instead, a high level of background (multiple weak hybridization signals of ~1.0 kb, 4.2, 4.4 and 5.0 kb) was observed in most tissues examined. Such weak background hybridization signals have been reported previously for several ZNF genes (Baban et al. 1996; Derry et al. 1995; Ogawa et al. 1998) and, as has been suggested, likely represent a combination of low level expression and cross-hybridization from multiple members of the KRAB ZNF gene family. To eliminate the problem of background hybridization and to specifically test for each of the two ZNF208 isoforms, a specific RT-PCR assay was developed for each of the two alternative transcripts of the ZNF208 gene (Figure 3). RT-PCR products consistent with the two different mRNA isoforms for ZNF208 were identified in most tissues examined (Figure 3). Sequence analysis of the RT-PCR products confirmed that these amplification products represented expression from ZNF208. These data confirm that ZNF 208 is a *bona fide* gene with two different transcripts resulting from alternative splicing and utilization of different polyadenylation signal sequences

Database searches with the putative ZNF genic portion of cosmid 32532 identified no ESTs with sequence similarity greater than the background level of homology for the ZNF91 gene family cluster (85-90%). Conceptual translation of ZNF209 predicted multiple stops (~6 stop codons) in-frame with normal ZNF translation. RT-PCR using two different primer pairs failed to detect expression of ZNF209 in any of the tissues examined. These data argue that ZNF 209 is a non-processed pseudogene.

E) Beta-satellite Repeat Structures flanking 19p12 ZNF Genes

Examination of the genomic organization of BAC 33152 using Miropeats, Repeatmasker and dot-matrix analysis software revealed the presence of large blocks (~24- 45 kb in size) of inverted repeat structures flanking the ZNF208 transcription unit (Figure 4). Sequence similarity searches indicated that large portions of these repeat structures showed limited homology (75-84%) to previously characterized beta-satellite consensus motifs (GenBank #M81228) (Greig and Willard 1992). In addition to inverted beta-satellite structures, human endogenous retrovirus sequences were also observed in opposite orientation flanking the ZNF gene within BAC33152. These observations suggest an inverted symmetry to the organization of repetitive elements in the vicinity of ZNF208.

Analysis of the beta-satellite blocks flanking ZNF208 revealed a remarkable higher-order superstructure. Three beta-satellite superstructures ranging in length from 21-23 kb were defined within BAC33152 (Figure 5, Table 2). Each of these structures was found to consist of three portions: 1) a 5' beta-satellite segment of 5-10 kb; 2) a middle Alu/LTR portion of 2-3 kb; and 3) a 3' beta-satellite segment of 5-10 kb. Two of these are located adjacent to one another, suggesting that they may compound to form a larger 40 kb structure (Figure 4). The marked symmetry in both length and orientation of repeat elements is shown below (Figure 5). If interspersed repeat elements (LINES and SINES) are excluded from the calculation, the length of the 5' and 3' beta-satellite portions with respect to the LTR/Alu complex segment appear highly conserved for each beta-satellite superstructure (Figure 5). Interestingly, in all three beta-satellite blocks examined, an Alu repeat element was observed symmetrically located within the center of the beta-satellite repeat element. In two of the three blocks examined, the Alu repeat element is conspicuously oriented in an inverted orientation with respect to other repeat elements within the superstructure.

During the analysis of the beta-satellite repeat structures, it was found that the individual beta-satellite repeat units were not simply organized as tandem reiterations of the 68 bp consensus motif within each beta-satellite segment (Agresti et al. 1989; Waye and Willard 1989; Willard 1990). Instead, the repeats are distributed as clusters of tandem arrays of variable length with intervening sequence separating each cluster. A total of 322 beta-satellite repeat sequences were found to be distributed in 26 clusters comprising 26,573 bp of the total sequence of BAC33152. Using MEME (multiple EM motif elicitation) software (Bailey and Elkan 1994), a 71mer consensus motif was generated from these 322 repeats (Table 2). The most favored consensus motif demonstrates 78.9% sequence identity with the previously identified "beta-satellite" consensus sequence (Agresti et al. 1989; Waye and Willard 1989; Willard 1990) (Figure 6). Sequence similarity searches with each intervening sequence between each cluster of beta-satellite repeats showed no significant homology to known repeat sequences in the NCBI database. Dot-matrix analysis of these regions, however, indicated that different intervening segments exhibit low level sequence similarity to each other. This suggests that the intervening segments themselves are repetitive. MEME analysis further revealed that each intervening segment is composed of tandem reiterations of a degenerate 38mer repeat motif (A total of 389 repeat units were identified within 22 clusters embedded within 21,036 bp of sequence). The most favored consensus motif of this 38mer repeat shows 65.9% sequence similarity to a core segment of the beta-satellite consensus motif (Figure 6). Thus, the beta-satellite segments are composed of alternating clusters of repeats demonstrating high sequence similarity (the 71mer repeat) and low sequence similarity (the 38mer repeat) to the beta-satellite consensus motif.

In order to investigate whether these beta-satellite repeat structures were a general property of the architecture of the ZNF cluster in 19p12, a 1.5 kb beta-satellite probe (M13 clone afb69e9, corresponding to positions 142121-153021 of GenBank accession#: AC003973) was hybridized against an arrayed (10X coverage) chromosome 19 cosmid library. A total of 95 chromosome 19-specific cosmids were identified that hybridized intensely with the beta-satellite repeat probe. 52 of these cosmids were distributed among five contigs whose location within 19p12 had been previously determined using high-resolution two-color FISH and STS content mapping (see Methods). Analysis of the locations of the beta-satellite structures within the *EcoR*1 by Southern hybridization as well as inference of positive/negative hybridizing cosmids within each contig predict that there are at least 7 blocks of beta-satellites (~40 kb in size), spanning 1.5 MB of the 4 MB cluster of ZNF genes. Comparison of the positions of the beta-satellite blocks with the position of putative ZNF genes indicates that these repeat structures generally bracket ZNF genes (see Figure 1). These large blocks of beta-satellites occur with a periodicity of once every 150-200 kb, a pattern of reiteration consistent with the tandem duplication of the ZNF genes (Figure 1). The predicted "beads-on-a-string" architecture of beta-satellite repeats in this region of 19p12 was subsequently confirmed by FISH analysis using beta-satellite repeats as probes against alkaline-borate preparations of metaphase nuclei (data not shown).

DISCUSSION

We have investigated the genomic organization of the zinc finger gene cluster located in cytogenetic band interval 19p12 at several levels of scrutiny. In order to obtain a general overview of the ZNF gene cluster organization we constructed an integrated physical map (~ 700 kb) of overlapping cosmid and BAC clones between genetic markers D19S454 and D19S269. Based on hybridization experiments with ZNF91 exon-specific probes, we identified four potential ZNF genes arranged in a head-to-tail fashion in this region with an average periodicity of one ZNF gene every 150 kb. This ZNF gene density in 19p12 is in general agreement with earlier estimates of the size of this gene cluster which predicted approximately 40 different genes within the 4-5 MB interval of 19p12 (Bellefroid et al. 1991). A recent study into the organization of KRAB ZNF genes in a different ZNF cluster located in 19q13.2 found a much greater density of genes, with as many as 15 different ZNF genes duplicated over a distance of 350-450 kb (Shannon et al. 1996). Interestingly, the average size of the 19q13.2 ZNF duplicon is more than 5 times smaller than the spacing between ZNF genes in 19p12. The discovery of large (25-40 kb) beta-satellite structures located on either side of some ZNF genes in 19p12 indicates that these repeat elements may account for some of the differences in spacing between these clusters. Similar to the 19p12 cluster, however, 19q13.2 ZNF genes were found to be arranged in a head-to-tail fashion. This suggests that both ZNF gene clusters have most likely arisen by a common evolutionary mechanism involving endoduplication of an ancestral “seed” ZNF cassette to generate a tandem array of genes followed by subsequent divergence of individual family members.

The general model for the structural organization of 19p12 ZNF genes consists of four exons: a 5’-UTR exon which includes the translational initiation codon, two exons encoding KRABA and KRABB protein interacting domains and a fourth exon which contains the spacer region, the DNA-binding ZNF repeat domain and the 3’-UTR (Bellefroid et al. 1991). Our analysis of 208,967 bp of 19p12 identified two potential ZNF genes (Bellefroid et al. 1993; Bellefroid et al. 1991) whose intron/exon structures were in complete agreement with the ZNF91 model (Table 1). The fact that a nearly identical exon/intron structure has been observed in other chromosome 19 ZNF clusters as well as KRAB ZNF genes from other chromosomal locations (Constantinou-Deltas et al. 1996; Derry et al. 1995; Grondin et al. 1996; Villa et al. 1993) suggests that this modular organization is a general property of most KRAB ZNF genes in the human genome. Conceptual translation, cDNA sequencing and RT-PCR expression analysis indicated that ZNF208 is a functional gene which is expressed in most tissues (Figure 2). Two distinct splice variants were identified for ZNF208, one of which is comprised of only the KRABA and KRABB protein domains, and results from the utilization of an alternative suboptimal polyadenylation signal (AATATA) prior to splicing of the fourth exon. Although the functional significance of these ZNF “tailless” transcripts remains to be determined, one hypothesis is that KRABA and KRABB peptides devoid of their DNA/RNA binding motif function to sequester proteins which normally interact with full-length ZNF proteins in order to co-repress transcription of target genes (Baban et al. 1996; Friedman et al. 1996; Kim et al. 1996). Such a competition for interaction with KRAB corepressors such as KAP-1 may prevent the association of these repression protein complexes with

DNA/RNA.

In order to evaluate the evolutionary age of the expansion of the ZNF cluster, we compared 43 kb of sequence from two different non-adjacent ZNF gene cassettes from the 19p12 cluster (Figure 3). Sequence conservation was identified by Miropeat and dot-matrix analysis in six genomic regions (totalling ~10.2 kb) between the ZNF208 and ZNF209 duplicons. BESTFIT alignment of the non-coding portions of these duplicons (8,041 bp) showed 79.5% nucleotide identity. Based on the neutral mutation rate (5 X 10^{-9} mutations per site per year), we estimate that the two duplicated segments diverged from a common ZNF ancestral sequence approximately 40 mya. The remaining ~32.8 kb, which showed virtually no sequence homology, consisted almost entirely of different short interspersed repeat elements. It may be noteworthy that the majority of retroposons identified in these two duplicated segments belong to subfamilies that were active before the divergence of the Old World and New World monkeys (35-44 mya) (Batzer et al. 1996; Shen et al. 1991; Smit 1993; Smit and Riggs 1995; Smit et al. 1995). These include the L1PA7, L1PA16, HERV3 and AluS retroelements (Figure 3). The fact that these occur in non-paralogous positions between the ZNF208 and ZNF209 suggest the duplicated ZNF structure already existed prior to the divergence of these anthropoid clades. Although more detailed analysis of other ZNF91 duplicated genomic segments from 19p12 is required, both the molecular clock and molecular fossil data would indicate that the expansion of the ZNF91 gene cluster in 19p12 occurred approximately 40-50 mya. These findings are in general agreement with other studies which showed no association of ZNF91 genes with the syntenic region of 19p12 among prosimians, such as tarsier and squirrel monkey, but identified 19p12 orthologous ZNF91 genes in all anthropoids studied (Bellefroid et al. 1991).

Sequence analysis of the intergenic regions of ZNF208 uncovered a complex genomic architecture of beta-satellite repeats (Figure 4) bracketing this zinc-finger gene. Three hierarchial levels of organization were identified. First, large blocks (25-40 kb) of beta-satellite repeat sequences flanking a core segment harboring a ZNF gene appear to define the basic unit of ZNF duplication for a significant portion of the 19p12 gene cluster. Secondly, peculiar super-repeat structures were identified ranging in length from 20-23 kb within each "block" of beta-satellites. Each of the three beta-satellite superstructures of BAC33152 showed a similar tripartite organization: a 5' beta-satellite repeat portion, a middle portion consisting of a complex of Alu and LTR retroelements and a 3' beta-satellite repeat portion (Figure 5). Remarkable symmetry was observed for each of these structures in which the Alu/LTR complex was centrally located flanked by beta-satellite "arms" of nearly identical length (Figure 5 and Table 2). Such a conservation in length of genomic segments harboring beta-satellites is particularly surprising since the number of beta-satellite repeats, especially among acrocentric chromosomes, is known to be unstable and highly variable within the human population (Willard 1990). Finally, the basic chromosome 19 beta-satellite units (71 bp) were organized into clusters of tandem repeats ranging from 3 to 34 monomers (Table 2). These clusters were separated by intervening segments (approximately 800 bp in length) which were themselves a 38mer degenerate repeat of the beta-satellite repeats (Figure 6).

Previous studies into the organization of beta-satellite repeats have not indicated such a complicated higher-order repeat organization (Willard 1990; Worton et al. 1988). Most investigations have suggested that beta-satellites are organized as large tandem arrays ranging in length from 70-400 kb abutting other satellite DNAs in the vicinity of the centromere (Cooper et al. 1992; Shiels et al. 1997). Although the chromosome 19 organization may be exceptional, large-scale sequence analysis will be required to determine whether similar higher-order repeat structures are present in other beta-satellite chromosomal regions. It is intriguing that large palindromic structures have recently been identified for a different pericentromeric repeat multisequence family, termed chAB4 (Assum et al. 1991; Wohr et al. 1996). ChAB4 repeat units are organized as inverted duplications of 90 kb flanking a “non duplicated” core sequence estimated to be approximately 60 kb in length. This is similar to the ZNF-beta satellite organization observed in this study in which a 40 kb inverted duplication of beta satellite repeats flanks a 90 kb core sequence harboring the ZNF gene. Interestingly, both of these palindromic structures appear to be localized exclusively within the pericentromeric regions of chromosomes and are organized as clusters within each of these regions. As has been suggested, such inverted structures may be an inherent property in the dispersal and proliferation of these repeat sequences (Wohr et al. 1996).

In the human genome, beta-satellite repeats have been identified in the pericentromeric regions of chromosomes 1, 9 and Y, as well as all acrocentric chromosomes (13,14,15, 21 and 22) (Agresti et al. 1989; Agresti et al. 1987; Cooper et al. 1992; Greig and Willard 1992; Waye and Willard 1989; Willard 1990). To our knowledge, this is the first report describing the presence of beta-satellite repeats in 19p12. Several features of the chromosome 19 beta-satellite repeats, however, are anomalous with respect to the classically defined beta-satellite (68 mer) repeats of other chromosomes. The chromosome 19 beta-satellite repeats reiterate with a periodicity of once every 71 bp instead of 68 bp and they lack the highly conserved nucleotide block (GATCAGTGC) which has been proposed to function as a protein-binding site for this repeat (Agresti et al. 1989; Vogt 1990). Although the overall consensus motif exhibits 78.9% identity to the standard” beta-satellite repeat consensus, out of the 322 repeat units examined in BAC33152 not a single repeat showed greater than 75% identity to previously characterized beta-satellite repeat units. The interposed 38mer beta-satellite-like clusters between the 71mer repeats showed substantially less sequence similarity (Figure 6). It is not surprising, then, that previous fluorescent *in situ* experiments with other beta-satellite probes failed to identify the presence of beta-satellites on 19p12. Indeed, reciprocal experiments with probes derived from the beta-satellite repeat structures of BAC33152 hybridized exclusively to 19p12 (data not shown) indicating that these particular repeats are specific to chromosome 19. This suggests that other large satellite repeat sequences distributed in the pericentromeric region of the human genome may yet remain to be discovered.

Both pericentromeric and telomeric regions of human chromosomes have recently been shown to demonstrate an unusual proclivity to duplicate gene-containing genomic segments (Eichler et al. 1997; Eichler et al. 1996; Regnier et al. 1997; Trask et al. 1998; Winokur et al. 1996; Zimonjic et al. 1997). It has been suggested that various repeat sequences in these regions may be involved in promoting duplication. The fact that many of the KRAB ZNF genes are located in close proximity to subtelomeric and pericentromeric regions may explain their rapid proliferation in the human genome (Hoffman et al. 1996; Jackson et al. 1996; Lichter et al. 1992; Tommerup et al. 1993; Tommerup and Vissing 1995; Tunnacliffe et al. 1993). Since beta-satellites appear to have duplicated in concert with the ZNF genes in this region of 19p12, the most prosaic explanation is that they were part of the original ZNF gene cassette that became duplicated. We propose a model in which a single ancestral ZNF progenitor gene became associated with beta-satellite repeat sequences located in proximal 19p12 (Figure 7). This may have occurred by a process of pericentromeric-directed transposition as has been described for other human chromosomes or by chromosomal rearrangements which are common during speciation events. In this regard, it is interesting that FISH experiments with ZNF91 cDNA probes against prosimian chromosomal metaphase spreads have identified putative orthologues in subtelomeric regions of chromosomes which are not syntenic to 19p12 (Bellefroid et al. 1995). Centromeric repeat sequences, such as beta-satellites, are known to be capable of rapid expansion and contraction, presumably by mechanisms involving saltatory replication or unequal crossing-over events (Willard 1990). Once the ancestral ZNF gene integrated near such a heteromorphic beta-satellite repeat, it began to be duplicated, becoming effectively carried within the beta-satellite matrix which was in a state of flux. The presence of inverted beta-satellite blocks flanking the ZNF gene may have promoted further duplication events. Such large palindromes have been shown to promote gene amplification somatically (Hyrien et al. 1988; Windle and Wahl 1992) and are found associated with other duplicated gene family clusters (Bishop et al. 1985; Gao et al. 1997; Groot et al. 1990). Due to the potential selective advantage of expressed ZNF genes within the repeat structure, evolutionary pressure may have favored expansion over contraction of this region, leading to the generation of a large cluster of tandemly duplicated genes in the human genome (Figure 7). Such a model, although speculative, would help explain the rapid expansion of the ZNF91 gene family over a relatively short period of time during primate evolution (Bellefroid, 1995). It should be emphasized that an association between ZNF genes and beta-satellites has only been documented for a 1.5 MB region of 19p12 located in proximity to the centromere. It will be interesting to determine whether beta-satellites or perhaps other pericentromeric repeat sequences have been involved in the expansion of the remaining ~3.0 MB of the ZNF gene cluster in this region.

METHODS

Physical Map Construction

A foundation physical map of sets of overlapping cosmids between STS markers D19S454 and D19S269 was constructed as previously described (Ashworth et al. 1995). More than 85 cosmid clones were identified from chromosome-19 specific libraries (LLN19C02 “F”, LLN19C03 “R” (de Jong et al. 1989) which hybridized to YAC clone probes from this region. The organization of the framework cosmid map was confirmed by fluorescence *in situ* hybridization in sperm pronuclei to estimate distance between selected cosmid clones in the map (Figure 1), the assignment of known chromosome 19 STS (sequence tagged site) markers to the region, an automated fluorescence-based restriction fingerprinting technique to confirm the order and overlap of *EcoR*1 fragments and hybridization of known YAC insert probes (CEPH YAC library) that had been assigned to the region (Ashworth et al. 1995). A human genomic BAC library (5 X coverage) (Research Genetics) was screened in order to identify larger insert clones which would bridge adjacent but non-overlapping cosmid contigs. A previously described protocol involving long-range inter-Alu PCR in conjunction with T7-Alu and Sp6-Alu PCR (Parrish et al. 1995) was used to generate probes from terminal cosmids of each set of overlapping clones. Hybridization against total human genomic BAC libraries identified a set of candidate BAC clones which were each, in turn, subjected to long-range inter-Alu PCR amplification and the fragments were used as probes back against the chromosome 19 specific cosmid libraries. This cosmid-to-BAC and BAC-to-cosmid approach was used to insure the isolation of *bona fide* chromosome 19 BACs and to identify additional cosmid contigs that mapped to the region. Comparative *EcoR*1 fluorescent fingerprinting between cosmid contigs and BACs as well as FISH hybridization on chromosomal preparations from human metaphase and sperm pronuclei (Trask et al. 1993) were used as final criteria for assigning BACs to the 19p12 physical map.

Library Preparation and Sequencing

Random shotgun libraries from BAC33152 and cosmid32532 were prepared in the M13mp18 vector using slight modifications of a previously described protocol (Lamerdin et al. 1996). Cosmid and BAC DNA was isolated using Quiawell 8 DNA Isolation Systems (Qiagen) and the DNA was sheared using a TDL Nebulizer (constructed at the Washington University School of Medicine) for 4 min at 30 psi to generate DNA fragments with an average insert size of 1.5 kb. Size-selected and end-repaired fragments were blunt-end ligated and subcloned into the M13mp18 vector. Well-separated M13 plaques were arrayed into 96-well microtiter plates using an automated colony picker designed by Lawrence Berkeley Laboratory. After incubation at 37 °C for 7-8 hrs, single-stranded DNA for each subclone was isolated using either Qiagen 96-well format M13 kits (per manufacturer's specifications) or a previously described PEG precipitation protocol (Kristensen et al. 1987). Fluorescent dye-primer sequence reactions were prepared using a Catalyst 800 Molecular Biology labstation, a PE9600 thermocycler and ABD Taq thermosequenase cycle sequencing kits (Perkin-Elmer Applied Biosystems). A modified assymetric PCR protocol (Munzy et al. 1993) was used in the directed reverse sequencing phase of the project. Sequencing reaction products were analysed on both ABD 373A and ABD 377 sequencers (Applied Biosystems). The sequence data was analysed using PHRED/PHRAP software and the assembled sets of overlapping sequence reads were edited using Consed v.3.0 (software available from Phil Green and David Gordon; http://genome.wustl.edu.). Regions lacking double-stranded continuity or areas of poor sequence quality within the sequence assembly were identified by SWEDISH software (available from Matt Nolan, Lawrence Livermore National Laboratory). Additional subclones and sequence reads were generated within these regions. One region (2.2 kb) was directly subcloned from BAC33152 and a sublibrary, using the ABI PRISM Primer Island Transposition Kit (PE Applied Biosystems), was prepared in order to increase the number of high quality double-stranded sequencing reads in the region (Devine and Boeke 1994). A total of 590 sequence reads were analysed and assembled for the 43 kb insert of cosmid 32532, (11.7-fold sequence redundancy, 99.5% double-stranded). A total of 3419 sequence reads were analysed and assembled for the 163 kb insert of BAC33152 (10.7-fold sequence redundancy, 98.5% double-stranded). The sequence of BAC33152 and cosmid 32532 was deposited in GenBank under the accession nos. AC003937 and AC004004.

RT-PCR Analysis

Alternative splicing of ZNF208 was analysed by RT-PCR using three oligonucleotide amplification primers: afb51 (5'-TCCTTACTGCTGTGTGTCCTCTGCTCC-3'), afb52 (5'-CTACTTCT TTTGGAACACAG CTTCCAG-3') and afb62 (5'-TTCTATGCCCTGCTCTGGCCAAAG-3'). High stringency PCR conditions were optimized to eliminate cross-amplification from other ZNF loci and insure ZNF208 specificity. Afb51/afb52 amplification conditions consisted of an initial denaturation of 5 minutes at 95 °C, followed by

35 cycles of 30 seconds at 95 °C, 30 seconds at 65 °C and a 45 second extension at 72 °C. A final extension of 5 minutes was carried out at 72 °C. Afb51/afb62 RT-PCR reactions were similar, with the exception that both extension and annealing profiles were combined into a single amplification step of 45 seconds at 75 °C. All cycling conditions were performed in a 9600 thermocycler (Perkin-Elmer Applied Biosystems). The cDNA was prepared with the Superscript cDNA synthesis kit following manufacturer's suggested protocol (Gibco BRL). 1ug of polyA mRNA (Clontech) isolated from twelve human tissue sources (adult brain, liver, spleen, heart, lung, muscle, kidney, pancreas, testis, uterus, placenta and fetal brain) was used as template in cDNA preparation. For each cDNA synthesis reaction, a negative control without the reverse transcriptase was included (Figure 3). PCR products from brain, testis and uterus were cloned for each RT-PCR reaction using pGEMT-Easy (Promega) and plasmid DNA was isolated (Qiagen) and sequenced using standard dye-primer fluorescent chemistry (Applied Biosystems). A minimum of three clones were sequenced for each RT-PCR reaction.

Computer Software

The location and identity of various repeat elements in the sequences were determined using RepeatMasker software (http://ftp/genome/washington.edu/cgi-bin/RepeatMasker). Miropeat software (Parsons 1995) was used to identify other blocks of internal repeat sequences that were not contained within the RepMask database. DOTTER, a dot-matrix sequence alignment program (Sonnhammer and Durbin 1995), in conjunction with Miropeats determined the extent of paralogy between duplicated segments of BAC33152 and cosmid 32532. In addition, both software programs were used to determine the general architecture of the beta-satellite repeat motifs flanking ZNF208 (Figure 4). The locations of putative exons in the sequences were determined with the GRAIL-2 gene-recognition tool (Uberbacher and Mural 1991) and by comparisons of known ZNF91 cDNA sequences to the genomic sequence. All sequence alignments were performed using BESTFIT software (GCG). A GeneWorks software package (v.2.1, Intelligenetics) provided conceptual translation of putative coding regions in the sequences. In order to identify repeat consensus motifs within the beta-satellite repeat regions, the software tool MEME (multiple expectation-maximization for motif elicitation) (Bailey and Elkan 1994) was employed. Regions showing sequence similarity to the beta-satellite consensus (>70-80%) were extracted from BAC33152 (total basepairs=26,573) and subjected to MEME analysis. A similar analysis was performed on intervening repeat-masked sequences (total basepairs=21,306) located between regions of beta-satellite homology. The most probable consensus motifs were generated from comparison of 322 sequence motifs for the 71mer repeat and 389 sequence motifs for the 38mer repeat. MAST software (Motif alignment search tool) was used to evaluate the significance of each consensus motif (Bailey and Elkan 1994).

Acknowledgements

We would like to thank A.F. Smit and T.L. Bailey for assistance in the analysis of repeats in this region and M. Coleman for helpful suggestions in the preparation of this manuscript. We are grateful to M. Christensen, L. Woo, A. Kobayashi, D. Ow for excellent technical and informatics assistance. The chromosome specific gene libraries LL019NC02 and LL019NC03 used in this work were constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550. This research was supported, in part, by a US Dept. of Energy Human Genome Postdoctoral fellowship to EEE, by a Howard Hughes Medical Institute grant to Case Western Reserve University School of Medicine and by the Lawrence Livemore National Laboratory under the auspices of DOE contract W-7045-Eng48.

LEGENDS

Figure 1: Physical Map of the ZNF gene cluster in 19p12. The organization of the region under study is depicted at three levels of resolution. a) An ideogram of chromosome 19 delineates the ~1.5 Mb portion from the 4-5 Mb region of the ZNF cluster that has been characterized. b) Cytogenetic distances between representative cosmids are shown based on estimates of physical distance from two-color FISH analysis on decondensed chromatin from sperm pronuclei. Cosmids used in the construction of the framework cytogenetic map are indicated below each contig (shown as open horizontal bars) in association with genetic markers for the chromosome 19 map. c) Overlapping BAC and cosmid genomic clones for 700 kb of this region are shown as horizontal lines. The orientation and order of clones in the contig were determined based on an *EcoR* I fingerprinting strategy of multiple overlapping clones. Only a subset of the clones in the total tiling path of these clones is indicated. The positions of cross-hybridizing probes specific for the KRABA exon (A), spacer region (S) and zinc finger gene (Z) of ZNF91 are indicated within the physical map. Vertical shaded bars indicate the position and extent of regions that cross-hybridize to beta-satellite probes derived from BAC33152. The two clones that have been sequenced are indicated with an asterisk.

Figure 2: Comparative Sequence Analysis of ZNF Duplicons. A genomic segment, corresponding to the entire insert (43,378 bp) of cosmid clone 32532 (GenBank HUAC003973) and paralogous positions 64,401-108,079 from BAC33152 (GenBank HUAC004004), was compared between the two ZNF 19p12 clones using Miropeats software (threshold score s=25; setting=onlyinter). Prior to analysis, common repeat elements were removed from the sequence using RepeatMasker software. Regions of sequence conservation which do not carry LINE or SINE sequences are delineated by “joining” lines between the two sequences. Six regions of paralogy were identified (indicated by Roman numerals) and each region was aligned using the BESTFIT alignment program (GCG software package). The position and identity of exons and other repeat elements are illustrated schematically above the miropeat alignment. Similar results were obtained using dot-matrix alignment software with unmasked sequence. Note that the identity and position of most repeat elements are not conserved between the duplicated segments.

Figure 3: RT-PCR Analysis of ZNF208. RT-PCR analysis of ZNF208 from twelve human tissue sources confirmed the presence of two transcripts that result from the usage of alternative polyadenylation signals. The position of the primers and concomittant PCR product size are shown with respect to the ZNF208 gene structure. cDNA synthesis reactions without reverse-transcriptase for each tissue served as a negative control (indicated by a minus sign).

Figure 4: Genomic Organization of BAC33152. A schematic diagram depicting the general organization of BAC33152, including the positions of exons as well as the human endogeneous retroviral elements and beta-satellite repeat regions (~20-25 kb in length) that flank the ZNF208 gene. The beta-satellite repeat sequences flanking the gene are inverted in orienation with respect to one another. The genomic organization is placed in the context of a dot-matrix alignment of BAC33152 sequence (AC004004) against itself (DOTTER). The ZNF repeats appear as a black square symmetrically located in the center of the figure, while the three beta-satellite repeat superstructures appear as a “patchwork crosses” on either side of the ZNF gene.

Figure 5: Higher-order structure of beta-satellite repeats. Beta-satellites are organized into super-repeat structures consisting of two beta-satellite segments flanking an Alu/LTR middle portion. The organization of each of the three “superstructures” is drawn to scale with reference to the positions of the repeats in GenBank AC004004). The total length of each of the three segments in bp is indicated beside each segment. Calculation of length of the beta-satellite flanks did not include Alu and LINE elements. This is especially evident for the structure located at positions 140-160, in which an L1PA5 element has integrated.

Figure 6: Beta-satellite Consensus motifs. a) A multi-level consensus sequence was generated using MEME software analysis of 322 motifs of 71 bp in length over 26,573 bp of BAC33152 sequence. These regions were analysed together based on sequence similarity to beta-satellite consensus sequence. The information content (described in bits) provides a relative measure of the degree of conservation for each basepair position in the consensus motif. The most favored consensus is shown in bold with less favored bases shown below each position. A given basepair is only included in the multi-level consensus if it occurs with a frequency of greater than 0.2 in the consensus. b) A multi-level consensus sequence was similarly constructed based on MEME analysis of 389 motifs of 35 bp in length over 21,036 bp. This analysis was performed on those sequences which were located between regions showing sequence similarity to beta-satellites. c) BESTFIT alignment of the most-favored consensus motifs for the 38mer and 71mer repeats against the beta-satellite consensus (Vogt 1990) is shown. Sequence which is conserved among all three repeat elements is shaded and boxed. Underlined sequence indicates regions highly conserved among previously identified beta-satellite repeat units. d) Percentage pairwise sequence similarity (BESTFIT software, GCG) is shown for the three consensus motifs.

Figure 7: Model for the Expansion of the ZNF gene cluster in 19p12. A hypothetical model is proposed in which the pericentromeric region of 19p12 is in a state of expansion and contraction due to saltatory amplification and/or unequal crossing-over of beta-satellite repeats

in this region. A functional ZNF progenitor gene associates with beta-satellite repeats in 19p12 approximately 50 mya. The region continues to expand and contract, effectively carrying the inserted ZNF gene as part of its heteromorphism. Expansion becomes favored over contraction among the beta-satellites due to the placement of a functional gene within its context which confers a selective advantage. This leads to the formation of a large cluster of tandemly duplicated ZNF genes in the anthropoid ancestor.

Table 1: Gene Structure. The exon/intron structures of gene ZNF208 and pseudogene ZNF209 are shown.   Exon boundaries were determined based on cDNA comparisons of other ZNF91 family members to the genomic sequence and by GRAIL analysis. Positions of exons refer to GenBank HUAC003987 for ZNF208 and GenBank HUAC004004 for ZNF209.

Table 2: Beta-satellite substructure of BAC33152. The beta-satellite repeat segments consist of alternating clusters of a 71 bp repeat motif and a 38 bp repeat motif as determined by MEME analysis. The length of each cluster, the number of repetitive units and their orientation with respect to the beta-satellite consensus sequence are summarized. Alu and LINE elements were excluded when estimating the size of each cluster of repetitive motifs.

Agresti A., Meneveri R., Siccardi A., Marozzi A., Corneio G., Gudi S., and Ginelli E. 1989. Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequence (HaeIII family). *J. Mol. Biol.* **205**:625-631

Agresti A., Rainaldi G., Lobbiani A., Magnani I., Di Lernia R., menerveri R., Siccardi A., and Ginelli E. 1987. Chromosomal location by in situ hybridization of the human Sau3A family of DNA repeats. *Hum Genet.* **75**:326-332

Ashworth L., Batzer M., Brandriff B., Branscomb E., de Jong P., Garcia E., Garnes J., Gordon L., Lamerdin J., Lennon G., et al. 1995. An integrated metric physical map of human chromosome 19. *Nat. Genet.* **11**:422-427

Assum G., Fink T., Klett C., Lengl B., Schanbacher M., Uhl S., and Wohr G. 1991. A new multisequence family in human. *Genomics* **11**:34-41

Baban S., Freeman J., and Mager D. 1996. Transcripts from a novel human KRAB zinc finger gene contain spliced Alu and endogeneous retroviral segments. *Genomics* **33**:463-472

Bailey T., and Elkan C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, California, pp 28-36

Batzer M., Arcot S., Phinney J., Alegria-Hartman M., Kass D., Milligan S., Kimpton C., Gill P., Hochmeister M., Panayiotis A., et al. 1996. Genetic variation of recent Alu insertions in the human populations. *J. Mol. Evol.* **42**:22-29

Becker K., Nagle J., Canning R., Biddison W., Ozato K., and Drew P. 1995. Rapid isolation and characterization of 118 novel C2H2-type zinc finger cDNAs expressed in human brain. *Hum. Molec. Genet.* **4**:685-691

Bellefroid E., Marine J., Matera A., Bourguignon C., Desai T., Healy K., Bray-Ward P., Martial J., Ihle J., and Ward D. 1995. Emergence of the ZNF91 Kruppel-associated box-containing zinc finger gene family in the last common ancestor of anthropoidea. *Proc. Natl. Acad. Sci. USA* **92**:10757-10761

Bellefroid E., Marine J.-C., Ried T., Lecocq P., Riviere M., Amemiya C., Poncelet D., Coulie P., deJong P., Szpirer C., et al. 1993. Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymophoid cells. *EMBO J.* **12**:1363-1374

Bellefroid E., Poncelet D., Lecocq P., Relevant O., and Martial J. 1991. The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc. Nat. Acad. Sci. USA* **88**:3608-3612

Bishop J., Selman G., Hickman J., Black L., Saunders R., and Clark A. 1985. The 45-kb unit of major urinary protein gene organization is a gigantic imperfect palindrome. *Mol. Cell. Biol.* **5**:1591-1600

Brandriff B., Gordon L., Fertitta A., Olsen A., Christensen M., Ashworth L., Nelson D., Carrano A., and Mohrenweiser H. 1994. Human chromosome 19p: a fluorescence in situ hybridization map with genomic distance estimates for 79 intervals spanning 20 Mb. *Genomics* **23**:582-591

Brandriff B., Gordon L., and Trask B. 1991. A new system for high-resolution DNA sequence mapping interphase pronuclei. *Genomics* **10**:75-82

Carrano A., de Jong P., Branscomb E., Slezak T., and Watkins B. 1989. Constructing chromosome- and region-specific cosmid maps of the human genome. *Genome* **31**:1059-1065

Constantinou-Deltas C., Bashiardes E., Patsalis P., Hadjimarcou M., Kroisel P., Ioannou P., Roses A., and Lee J. 1996. Complete coding sequence, exon/intron arrangement and chromosome location of ZNF45, a KRAB-domain-containing gene. *Cytogenet. Cell Genet.* **75**:230-233

Cooper K., Fisher R., and Tyler-Smith C. 1992. Structure of the pericentric long arm region of the human Y chromosome. *J. Mol. Biol.* **228**:421-432

de Jong P., Yokabata K., Chen C., Lohman F., Pederson L., McNinch J., and Van Dilla M. 1989. Human chromosome-specific partial digest libraries in lamda and cosmid vectors. *Cytogenet. Cell Genet.* **51**:985

Derry J., Jess U., and Francke U. 1995. Cloning and characterization of a novel zinc finger gene in Xp11.2. *Genomics* **30**:361-365

Devine S., and Boeke J. 1994. Efficient integration of artificial transposons into plasmid targets *in vitro*: a useful tool for DNA mapping, sequencing and genetic analysis. *Nucl. Acids. Res.* **22**:3765-3772

Di Cristofano A., Strazzullo M., Longo L., and La Mantia G. 1995. Characterization and genomic mapping of the ZNF80 locus: expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucl. Acids Res.* **23**:2823-2830

Eichler E., Budarf M., Rocchi M., Deaven L., Doggett N., Baldini A., Nelson D., and Mohrenweiser H. 1997. Interchromosomal duplications of the adrenoleukodystrophy locus: a phenomenon of pericentromeric plasticity. *Hum. Molec. Genet.* **6**:991-1002

Eichler E., Lu F., Shen Y., Antonacci R., Jurecic V., Doggett N., Moyzis R., Baldini A., Gibbs R., and Nelson D. 1996. Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Molec. Genet.* **5**:899-912

Friedman J., Fredericks W., Jensen D., Speicher D., Huang X., Neilson E., and Rauscher F. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Devel.* **10**:2067-2078

Gao L., Frey M., and Matera A. 1997. Human genes encoding U3 snRNA associate with coiled bodies in interphase cells and are clustered on chromosome 17p11.2 in a complex inverted repeat structure. *Nucl. Acids Res.* **25**:4740-4747

Greig G., and Willard H. 1992. Beta satellite DNA: characterization and localization of two subfamilies from the distal and proximal short arms of human acrocentric chromosomes. *Genomics* **12**:573-580

Grondin B., Bazinet M., and Aubry M. 1996. The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix. *J. Biol. Chem.* **271**:15458-15467

Groot P., Mager W., Henriquez N., Pronk J., Arwert F., Planta R., Eriksson A., and Frants R. 1990. Evolution of the human alpha-amylase multigene family through unequal, homologous and inter- and intrachromosomal crossover. *Genomics* **8**:97-105

Hoffman S., Hromas R., Amemiya C., and Mohrenweiser H. 1996. The location of MZF-1 at the telomere of human chromosome 19q makes it vulnerable to degeneration in aging cells. *Leuk Res* **20**:281-283

Hoovers J., Mannens M., John R., Bliek J., van Heynigingen V., Poreous D., Leschot N., Westerveld A., and Little P. 1992. High-resolution localization of 69 potential human zinc finger protein genes: a number are clustered. *Genomics* **12**:254-263

Hromas R., Collins S., Hickstein D., Raskind W., Deaven L., O'Hara P., Hagen F., and Kaushansky K. 1991. A retinoic acid-responsive human zinc finger gene, MZF-1, preferentially expressed in myeloid cells. *J. Biol. Chem.* **266**:14183-14187

Huebner K., Druck T., Croce C., and Thiesen H. 1991. Twenty-seven nonoverlapping zinc finger cDNAs from human T cells map to nine different chromosomes with apparent clustering. *Am. J. Hum. Genet.* **48**:726-740

Hyrien O., Debatisse M., Buttin G., and de Saint Vincent B. 1988. The multicopy appearance of a large inverted duplication and the sequence at the inversion joint suggest a new model for gene amplification. *EMBO J* **7**:407-417

Jackson M., See C., Mulligan L., and Lauffart B. 1996. A 9.75-Mb map across the centromere of human chromosome 10. *Genomics* **33**:258-270

Kim S., Chen Y., O'Leary E., Witzgall R., Vidal M., and Bonventre J. 1996. A novel member of the RING finger family, KRIP-a, associates wtih the KRAB-A transcriptional repressor domain of zinc finger proteins. *Proc. Natl. Acad. Sci. USA* **93**:15299-15304

Klug A., and Schwabe J. 1995. Zinc fingers. *FASEB J.* **9**:597-604

Kristensen T., Voss H., and Ansorge W. 1987. A simple and rapid preparation of M13 templates for manual and automated dideoxy sequencing. *Nuc. Acids. Res.* **15**:5507-5516

Lamerdin J., Stilwagen S., Ramierez M., Stubbs L., and Carrano A. 1996. Sequence analysis of the ERCC2 gene region in human, mouse and hamster reveals three linked genes. *Genomics* **34**:399-409

Lichter P., Bray P., Ried T., Dawid I., and Ward D. 1992. Clustering of C2-H2 zinc finger motif sequences within telomeric and fragile sites of human chromosomes. *Genomics* **13**:999-1007

Margolin J., Friedman J., Meyere W., Vissing H., Thiesen H., and Rauscher F. 1994. Kruppel-associated boxes are potent transcriptional repression domains. *Proc. Natl. Acad. Sci. USA* **91**:4509-4513

Miller J., McLachan A., and Klug A. 1985. Repetitive zinc-binding doains in the protein transcription factor IIA from Xenopus oocytes. *EMBO Journal* **4**:1609-1614

Mohrenweiser H., Olsen A., Archibald A., beattie C., Burmeister M., Lamerdin J., Lennon G., Stewart E., Stubbs L., Weber J., et al. 1996. Report on abstracts of the third international workshop on human chromosome 19 mapping 1996. *Cytogenet. Cell Genet.* **74**:161-186

Munzy D., Richards S., Shen Y., and Gibbs R. (1993) PCR based strategies for gap closure in large scale sequencing projects. In: Venter C (ed) Automated DNA Sequencing and Analysis Techniques. Harcourt, Brace and Jovanovich, London

Ogawa T., Poncelet D., Kinoshita Y., Noce T., Takeda M., Kawamoto K., Udagawa K., Lecocqu P., Marine J., Martial J., et al. 1998. Enhanced expression in seminoma of human zinc finger genes located on chromosome 19. *Cancer Genet. Cytogenet.* **100**:36-42

Parrish J., Eichler E., Shofield T., Chinault A., Graves M., Arenson A., Lee C., and Nelson D. 1995. Cosmid binning and cDNA identification in Xq28. *Am. J. Hum. Genet.* **57**:A267

Parsons J. 1995. Miropeats: graphical DNA sequence comparisons. *Comput Appl Biosci* **11**:615-619

Parsons R., Li G.-M., Longley M.J., Fang W.-h., Papadopoulos N., Jen J., de la Chapelle A., Kinzler K.W., Vogelstein B., and Modrich P. 1993. Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**:1227-1236

Pengue G., Caputo A., Rossi C., Barbanti-Brodano G., and Lania L. 1995. Transcriptional silencing of human immunodeficieny virus type 1 long terminal repeat-driven gene expression by the Kruppel-associated box repressor domain targeted to the transactivating response element. *J. Virol* **69**:6577-6580

Pieler T., and Bellefroid E. 1994. Perspectives on zinc finger protein function and evolution--an update. *Mol. Biol Rep.* **20**:1-8

Regnier V., Meddeb M., Lecointre G., Richard F., Duverger A., Nguyen V., Dutrillaux B., Bernheim A., and Danglot G. 1997. Emergence and scattering of multiple neurofibromatosis (NF1)-related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition. *Hum. Molec. Genet.* **6**:9-16

Rosenberg U.B., Schroeder C., Kienlin A., Cote S., Riede I., and Jaeckle H. 1986. Molecular genetics of Kruppel, a gene required for segmentation of the Drosophila embryo. *Nature* **319**:336-339

Rousseau-Merck M., Hillion J., Jonveaux P., Couillin P., Seite P., Thiesen H., and Berger R. 1993. Chromosomal localization of 9 KOX zinc finger genes: physical linkages suggest clustering of KOX genes on chromosomes 12, 16 and 19. *Hum. Genet.* **92**:583-587

Shannon M., Ashworth L., Mucenski M., Lamerdin J., Branscomb E., and Stubbs L. 1996. Comparative analysis of a conserved zinc finger gene cluster on human chromosome 19q and mouse chromosome 7. *Genomics* **33**:112-120

Shen M.-R., Batzer M., and Deininger P. 1991. Evolution of the master Alu gene(s). *J. Mol. Evol.* **33**:311-320

Shiels C., Coutelle C., and Huxley C. 1997. Contiguous arrays of satellites 1, 3, and beta form a 1.5-Mb domain on chromosome 22p. *Genomics* **44**:35-44

Smit A. 1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucl. Acids Res.* **21**:1863-1872

Smit A., and Riggs A. 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucl. Acids Res.* **23**:98-102

Smit A., Toth G., Riggs A., and Jurka J. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246**:401-417

Sonnhammer E., and Durbin R. 1995. A dot matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**:GC1-10

Teglund S., Olsen A., Khan W., Frangsmyr L., and Hammarstrom S. 1994. The pregnancy-specific glycoprotein (PSG) gene cluster on human chromosome 19: fine structure of the 11 PSG genes and identification of 6 new genes forming a third subgroup within the carcinoembryonic antigen (CEA) family. *Genomics* **23**:669-84

Thiesen H., Bellefroid E., Relevant O., and Martial J. 1991. Conserved KRAB protein domain identified upstream from the zinc finger region of Kox 8. *Nucleic Acids Res* **19**:3996

Tommerup N., Aagaard L., Lund C., boel E., Baxendale S., Bates G., Lehrach H., and Vissing H. 1993. A zinc-finger gene ZNF141 mapping at 4p16.3/D4S90 is a candidate gene for Wolf-Hirschhorn (4p-) syndrome. *Hum. Molec. Genet.* **2**:1571-75

Tommerup N., and Vissing H. 1995. Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics* **27**:259-264

Trask B., Fertitta A., Christensen M., Youngblom J., Bergmann A., Copeland A., de Jong P., Mohrenweiser H., Olsen A., Carrano A., et al. 1993. Fluorescence in situ hybridization mapping of human chromosome 19: cytogenetic band location of 540 cosmids and 70 genes or DNA markers. *Genomics* **15**:133-145

Trask B., Friedman C., Martin-Gallardo A., Rowen L., Akinbami C., Blankenship J., Collins C., Giorgi D., Iadonato S., Johnson F., et al. 1998. Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Molec. Genet.* **1998**:13-26

Tunnacliffe A., Liu L., Moore J., Leversha M., Jackson M., Papi L., Ferguson-Smith M., Thiesen H., and Ponder B. 1993. Duplicated KOX zinc finger gene clusters flank the centromere of human chromosome 10: evidence for a pericentric inversion during primate evolution. *Nucl. Acids. Res.* **21**:1409-1417

Uberbacher E., and Mural R. 1991. Locating protein-coding regions in human DNA sequences by multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88**:11261-11265

Villa A., Zucchi I., Pilia G., Strina D., Susani L., Morali F., Patrosso C., Frattini A., Lucchini F., and Repetto M. 1993. ZNF75: isoaltion of a cDNA clone of the KRAB zinc finger gene subfamily mapped in YACs 1 Mb telomeric of HPRT. *Genomics* **18**:223-229

Vissing H., Meyere W., Aagaard L., Tommerup N., and Thiesen H. 1995. Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Lett* **369**:153-157

Vogt P. 1990. Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". *Hum. Genet.* **84**:301-336

Waye J., and Willard H. 1989. Human beta satellite DNA: genomic organization and sequence efiintion of a class of highly repetitive tandem DNA. *Proc. Natl. Acad. Sci. USA* **86**:6250-6254

Wilkinson D., Bhatt S., Chavrier P., Bravo R., and Charnay P. 1989. Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. *Nature* **337**:461-464

Willard H. 1990. Alpha and beta satellite sequences on chromosome 21: The possible role of centromere and chromosome structure in nondisjunction. *Prog. Clin. Biol. Res.* **360**:39-52

Windle B., and Wahl G. 1992. Molecular dissection of mammalian gene amplification: new mechanistic insights revealed by analyses of very early events. *Mutat Res.* **276**:199-224

Winokur S., Bengtsson U., Vargas J., Wasmuth J., Altherr M., Weiffenbach B., and Jacobsen S. 1996. The evolutionary distribution and structural organization of the homebox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum. Mol. Genet.* **5**:1567-1575

Witzgall R., O'Leary E., Leaf A., Onaldi D., and Bonventre J. 1994. The Krueppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci. USA* **91**:4514-4518

Wohr G., Fink T., and Assum G. 1996. A palindromic structure in the pericentromeric region of various human chromosomes. *Genome Res.* **6**:267-279
Worton R., Sutherland J., Sylvester J., Willard H., Bodrug S., Dube I., Duff C., Kean V., Ray P., and Schmickel R. 1988. Human ribosomal RNA genes: orientationof the tandem array and conservation of the 5' end. *Science* **239**:64-68
Zimonjic D., Kelley M., Rubin J., Aaronson S., and Popescu N. 1997. Fluorescence in situ hybridization analysis of keratinocyte growth factor gene amplification and dispersion in evolution of great apes and humans. *Proc. Natl. Acad. Sci. U.S.A.* **94**