Mammalian Genome 10, 768–770 (1999).

Mammalian Genome
Incorporating Mouse Genome


# The genomic organization of the histone clusters on human 6p21.3

**Jung Ahn, Jeffrey R. Gruen**

Yale Child Health Research Center, Department of Pediatrics, Yale University School of Medicine, 464 Congress Avenue, New Haven, Connecticut 06520-8081, USA

Received: 22 December 1998 / Accepted: 26 February 1999

Histone proteins are essential for stabilizing and folding DNA in the cell nucleus. In higher eukaryotes, four core histone proteins, H2A, H2B, H3 and H4, form the octamer of the nucleosome, the smallest basic unit of chromatin structure (Stein et al. 1984). Another histone, H1, interacts with linker DNA between nucleosomes and is involved in making compact structures of chromatin. While all the histone genes are defined by their highly conserved coding sequences, expression pattern and genomic structure can further classify them. Cell cycle-independent histone genes, or replacement histone genes, are constitutively expressed in most tissues and cells and especially in terminally differentiated or quiescent cells; they have introns and polyadenylation signals. In contrast, the expression of cell cycle-dependent histone genes is tightly linked to DNA synthesis and abundant in rapidly dividing cells; they lack polyadenylation signals and introns.

In humans, most cell cycle-dependent histone genes are located on the short arm of Chromosome (Chr) 6 (6p21.3-22), with a few mapping to Chr 1 (Tripputi et al. 1986). They are grouped into two clusters separated by 2 Mb. According to Albig and associates, the telomeric cluster near marker D6S2239 contains 34 genes, and the centromeric cluster near D6S105 contains 18 genes. Both clusters include two pseudogenes (Albig et al. 1997; Albig and Doenecke 1997).

The genomic organization of histone genes has been determined in only two other vertebrates: chicken and mouse. Chicken has a single cluster of 39 histone genes. Mouse has three distinct histone clusters totaling 45 genes distributed within 2 Mb on Chr 13 (syntenic to human Chr 6p). While total histone gene number increases from chicken (39 genes) to mouse (45 genes) to human (52 genes), the number of distinct clusters appears random at one, three, and two respectively.

While developing a transcript map of the human 6p21.3-22 region, we identified additional histone genes that do not map to either described cluster. Here, we report the results of a 2 Mb search for additional human histone genes and describe the genomic organization of the histone clusters on 6p21.3-22, including a new microcluster. The organization of the histone clusters is contrasted in human, mouse, and chicken.

We converted a YAC contig of human 6p21.3-22 (Bray-Ward et al. 1996) into a contig of smaller bacterial clones consisting of PACs, BACs, and cosmids. Human sequences extracted by inter-IRS PCR (Stein 1994) from three overlapping YACs[1], 753H12, 905G1, and 947F6, were used to screen a human Chr 6 cosmid library (LANL Life Science Division, Los Alamos, NM; http://www-bio.llnl.gov/genome/html/cosmid-html) and a human PAC library (Roswell Park Cancer Institute, Buffalo, N.Y.; http://bacpac.med.buffalo.edu). Cosmid and PAC clones identified from the screening were grouped into several bins by dot blot hybridization and by shared STSs that were obtained from public databases[2] and from clone end sequences. The bins were then connected by rescreening the libraries with clone end fragments isolated by the inverse PCR method (Silver 1991). This approach yielded a contig of PACs and cosmids spanning 2 Mb, from D6S2239 (pter) through D6S105 (cen), with one gap in the middle. We filled the gap with BAC clones, 61E19 and 24O18, and extended the telomeric end of the contig with PAC clones described by Lauer and colleagues (1997; Fig. 1).

To eliminate any possible uncertainties in the contig, partial and complete *Xba*I restriction digests of individual PAC, BAC, and cosmid clones were compared to confirm overlaps. Both conventional and field inversion agarose gel electrophoresis (FIGE) were used to maximize the resolution of the 0.5–40 kb fragments that were generated by the *Xba*I digests (Ota and Amemiya 1996). The resulting restriction patterns eliminated false positives and provided an accurate estimate of the physical distances between clones, genes, and markers (Fig. 1). We also searched the entire minimum tiling path consisting of 35 overlapping PACs, two BACs, and five cosmids for additional histone genes. These clones were sent to the Sanger Centre (Hinxton, Cambridge, CB10 1SA, UK) for sequencing as a part of the Chr 6 sequencing project.

Histone genes were detected by hybridizing five unique histone probes from H1, H2A, H2B, H3, and H4 genes to Southern blot filters made from PACs, BACs, and cosmids of our contig. These hybridizations detected clusters of histone genes in three unique locations (Fig. 1). The centromeric cluster contained 18 genes (one H1, five H2A and H2B, three H3, and four H4) distributed over 120 kb near D6S105. The telomeric cluster contained 34 genes (five H1, six H2A, eight H2B, seven H3, and eight H4) distributed over 300 kb around D6S2239. These results were consistent with previous reports by Albig and coworkers (Albig et al. 1997; Albig and Doenecke 1997).

Between the two known clusters, a new microcluster was detected containing five histone genes mapping within three *Xba*I fragments from four overlapping clones: p13H15, p86C11, p48N19, and b61E19 (Fig. 1). Two fragments were positive for both H2A and H2B, while the third contained only H4. All four clones contained D6S2252, which we mapped between D6S1260 (cen) and D6S1558 (pter). The relative order and orientation of the five histone genes were determined from the sequence of PAC clone p86C11 (AL021807). p86C11 is an 89,016-bp clone identified by several groups (Lauer et al. 1997; Volz and Ziegler 1996) including our own and sequenced by the Chr 6 group at the Sanger Centre (www.sanger.ac.uk/HGP/Chr6). Three genes in this new middle cluster were previously reported and named H4/m (AB000905), H2A/p (L19778), and H2B/r (X00088; Albig and Doenecke 1997). The other two genes, an H2A and an H2B, were not previously described, and we name them H2A/s and H2B/s. All five histones have adequate open reading frames and typical

*Correspondence to:* J.R. Gruen

[1] From CEPH mega YAC library.

[2] GDB: The Genome Database, http://gdbwww.gdb.org. Whitehead database: Center for genome research at Whitehead Institute for Biomedical Research, http://www-genome.wi.mit.edu.

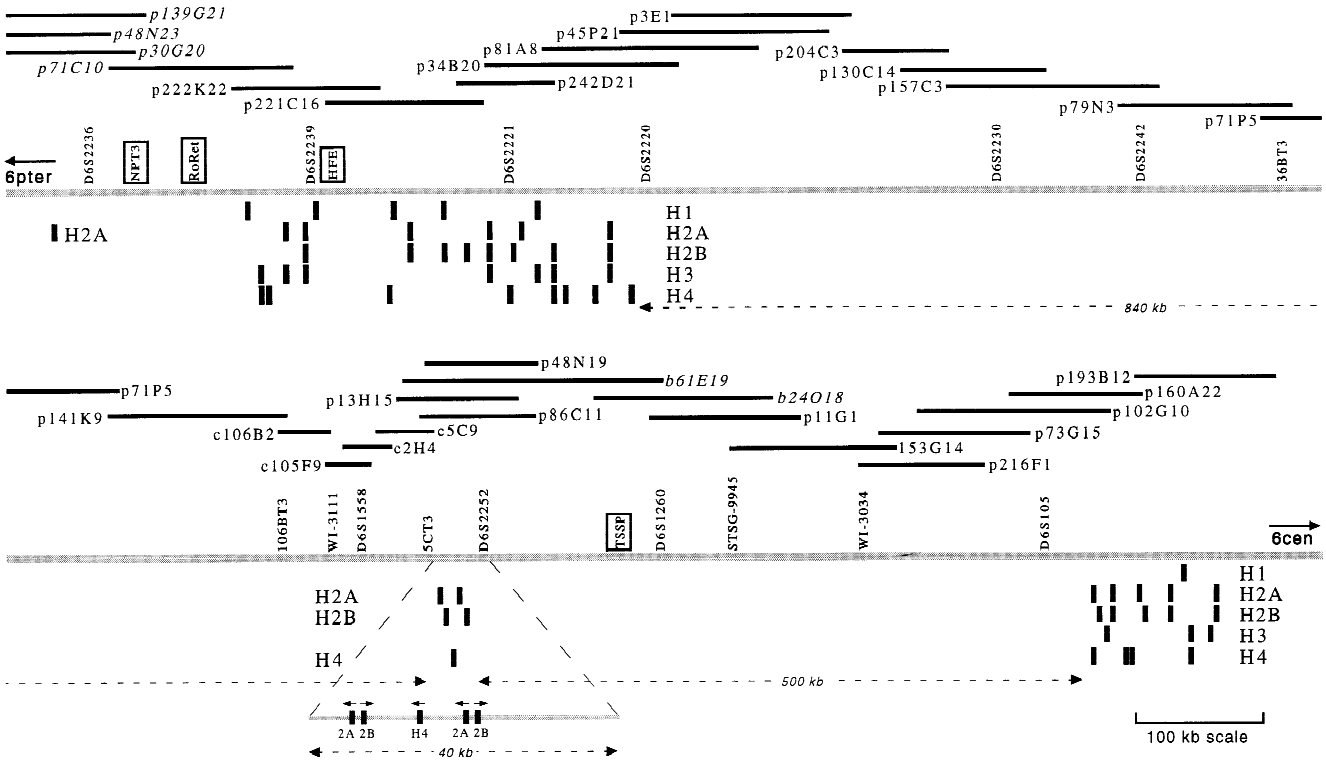


**Fig. 1.** Physical map and overall view of the organization of the three histone clusters on human Chr 6p. The continuous gray line represents Chr 6 with 6pter at the left end in the upper panel and 6cen at the right end in the lower panel. Upper and lower panels overlap in the region of clone p71P5. Overlapping PACs "p____", BACs "b____", and cosmids "c____" are ordered above the chromosome. Physical distances between markers and genes and the extent of overlap between genomic clones were derived from partial and complete restriction digests and Southern hybridizations as described in the text. The clones obtained from previous reports (Lauer et al. 1997) are italicized. STSs and known genes (boxed in vertical rectangles) are positioned above the chromosome. The primer sequences of STSs from the clone ends are: 36BT3—5′AAACTAGTGTCCTAAACTGTAGC3°, 5′CTACTTCTGATATTATCAGGCC3′; 106BT3—5′GCTTTGTGAAAGCATCTAG3′, 5′CTAGTGAACTAAATCAGTCTCC3′; 5CT3—5′TATCATCTGATGACCTTGGG3′, 5′GGGCCAACATGTACCAGA3′. Histone genes are represented as vertical bars under the chromosome. The arrows above the histone genes in the blow-up of the middle cluster in the lower panel indicate the orientation of histone ORFs based on the DNA sequence of dj86C11 (AL021807). Histone genes were identified by hybridizing individual histone probes to Southern blots of *Xba*I-restricted and biotinylated PAC, BAC, and cosmid DNAs.

promoters and stem-loop termination structures common to all cell cycle-dependent histone genes, suggesting they are likely to be transcribed.

We also found an H2A gene located at least 100 kb telomeric to the most telomeric cluster. It is contained in a 1.6-kb *Xba*I fragment common to three overlapping PACs: p139G21, p30G20, and p48N23 (Fig. 1). The sequence of this lone H2A gene shows that it lacks a promoter and the first 64 nucleotides of the open reading frame, suggesting it is probably a pseudogene (AF098457). Interestingly, the 5′-truncated ORF is flanked by direct pentanucleotide repeats (CCAGC) that are also found in the flanking regions of a *Drosophila* histone pseudogene (Akhmanova and Hennig 1998). The presence of the repeats suggests that the incomplete H2A pseudogene was inserted by a process involving transposable elements.

Detection of the middle microcluster and a lone pseudogene at the telomeric end complete the search of histone genes on 6p21.3-22. In this region, histone genes are clustered in three locations separated by 500- and 840-kb intervals (Figs. 1, 2). The clusters differ from each other in size and number of histone genes. Clues to their origin and history are suggested by the comparative studies of chicken and mouse (Wang et al. 1996; Takami et al. 1996). In chicken, 39 cell cycle-dependent histone genes are organized compactly into a single cluster of 110 kb. No other genes are present between the chicken histone genes or in the vicinity of the cluster. Histone genes on mouse Chr 13 (syntenic to human 6p21.3-22) are organized into three clusters with an average size of approximately 100 kb (Wang et al. 1996).

Even though the intercluster distances in mouse are not well delineated, two consecutive clusters are believed to be close to each other. The location of the hemochromatosis gene (HFE) within the telomeric cluster in human and between the two closest clusters in mouse suggests that the organization of histone genes in this region is similar in both species (Albig et al. 1998). Furthermore, five H1 genes found in the two closest clusters in mouse are all contained in the telomeric cluster in human. This suggests that the two mouse clusters correspond to the single telomeric cluster in human (Fig. 2).

The gene organization of the third mouse cluster coincides well with that of the centromeric cluster in human. Interestingly, the number of H1 genes in all three species is fixed at six, while the number of all other histone genes increases from chicken to human. Despite the significant similarities within clusters, differences are apparent in the overall genomic organizations in all three species (Fig. 2). While no interruptions are found in the chicken histone cluster, the distance between the two larger clusters in mouse is 500–650 kb, which is significantly smaller than the 1,350 kb in human. Furthermore, the middle microcluster identified in human was not detected in mouse.

This study suggests that the major histone clusters, including the conserved HFE gene, are orthogonal in human and mouse. The human histone region is expanded in total size and gene number

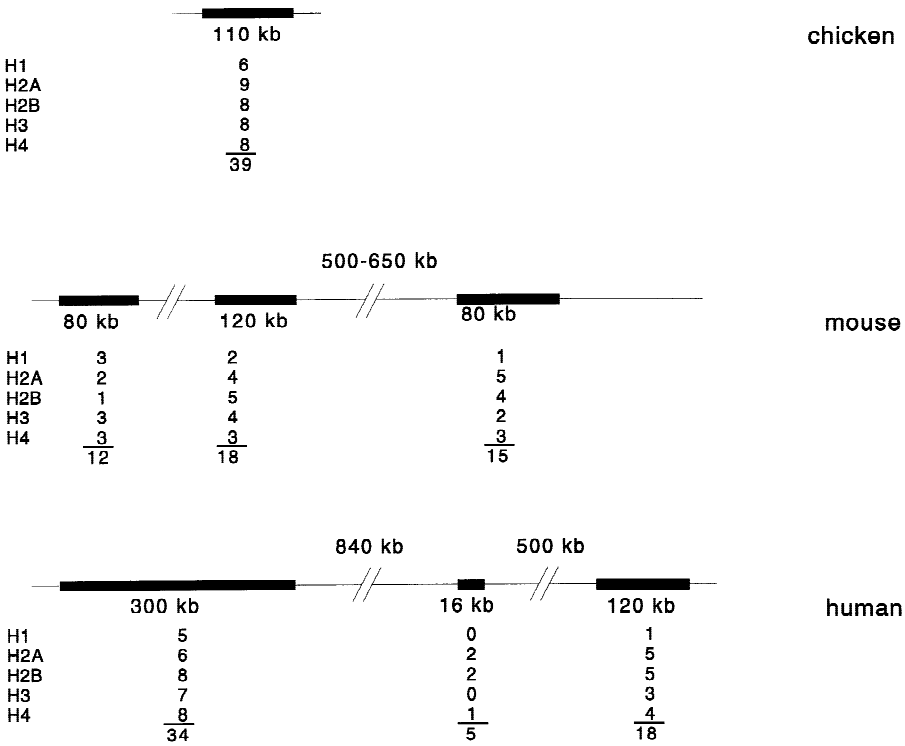


**Fig. 2.** Schematic diagram of major histone clusters in chicken, mouse, and human.

compared with mouse with the curious exception of the 6 H1 genes found in chicken, mouse, and human. Conclusions about the evolutionary events that forged this region await further transcriptional and genomic mapping studies of other non-human genomes.

*Acknowledgments.* This research was supported by National Institutes of Health Grant 5R29DK45819 (J.R. Gruen) and March of Dimes Clinical Research Grant 6-FY97-0539 (J.R. Gruen).

## References

Akhmanova A, Hennig W (1998) *Drosophila melanogaster* histone H2b retropseudogene is inserted into a region rich in transposable elements. Genome 41, 396–401

Albig W, Doenecke D (1997) The human histone gene cluster at the D6S105 locus. Hum Genet 101, 284–294

Albig W, Kioschis P, Poustka A, Meergans K, Doenecke D (1997) Human histone gene organization: nonregular arrangement within a large cluster. Genomics 40, 314–322

Albig W, Drabent B, Burmester N, Bode C, Doenecke D (1998) The haemochromatosis candidate gene HFE (HLA-H) of man and mouse is located in syntenic regions within the histone gene cluster. J Cell Biochem 69, 117–126

Bray-Ward P, Bowlus C, Choi J, Paslier DL, Weissenbach J et al. (1996) FISH-mapped CEPH YACs spanning 0 to 46 cM on human chromosome 6. Genomics 36, 104–111

Lauer P, Meyer NC, Prass CE, Starnes SM, Wolff RK et al. (1997) Clone-contig and STS maps of the hereditary hemochromatosis region on human chromosome 6p21.3-p22. Genome Res 7, 457–470

Ota T, Amemiya CT (1996) A nonradioactive method for improved restriction analysis and fingerprinting of large P1 artificial chromosome clones. Genet Anal 12, 173–178

Silver J (1991) Inverse polymerase chain reaction. In *PCR: A Practical Approach*, MJ McPherson, P Quirke, GR Taylor, eds. (New York: IRL Press), pp 137–146

Stein GS, Stein JL, Marzluff WF (1984) *Histone Genes: Structure, Organization, and Regulation.* (New York: Wiley)

Stein LD (1994) Large-insert cloning and analysis. In *Current Protocols in Human Genetics*, AL Boyle, ed. (New York: John Wiley and Sons), Vol 1, pp 5.9.1–5.9.4

Takami Y, Higashio M, Fukuoka T, Takechi S, Nakayama T (1996) Organization of the chicken histone genes in a major gene cluster and generation of an almost complete set of the core histone protein sequences. DNA Res 3, 95–99

Tripputi P, Emanuel BS, Croce CM, Green LG, Stein GS, et al. (1986) Human histone genes map to multiple chromosomes. Proc Natl Acad Sci USA 83, 3185–3188

Volz A, Zeigler A (1996) Physical mapping of a 6 Mbp region directly telomeric of the HLA-complex. DNA Seq. 7, 61–62

Wang ZF, Krasikov T, Frey MR, Wang J, Matera AG et al. (1996) Characterization of the mouse histone gene cluster on chromosome 13: 45 histone genes in three patches spread over 1 Mb. Genome Res 6, 688–701