



## Comparative Analysis of the $\alpha$ -Like Globin Clusters in Mouse, Rat, and Human Chromosomes Indicates a Mechanism Underlying Breaks in Conserved Synteny

Cristina Tufarelli, Ross Hardison, Webb Miller, et al.

*Genome Res.* 2004 14: 623-630

Access the most recent version at doi:[10.1101/gr.2143604](https://doi.org/10.1101/gr.2143604)

---

### References

This article cites 32 articles, 8 of which can be accessed free at:  
<http://genome.cshlp.org/content/14/4/623.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

A banner advertisement with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the words "LEARN MORE". On the right, there is a photograph of a woman wearing a red mask and a red cape, and the Cellecta logo, which consists of a green molecular structure and the word "CELLECTA" in white capital letters.

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# Comparative Analysis of the $\alpha$ -Like Globin Clusters in Mouse, Rat, and Human Chromosomes Indicates a Mechanism Underlying Breaks in Conserved Synteny

Cristina Tufarelli,<sup>1,6</sup> Ross Hardison,<sup>2,4</sup> Webb Miller,<sup>3,4</sup> Jim Hughes,<sup>1</sup> Kevin Clark,<sup>1</sup> Nicki Ventress,<sup>1</sup> Anna Maria Frischauf,<sup>5</sup> and Douglas R. Higgs<sup>1,7</sup>

<sup>1</sup>MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK; <sup>2</sup>Department of Biochemistry and Molecular Biology, <sup>3</sup>Department of Biology, and <sup>4</sup>Huck Institutes of Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA; <sup>5</sup>Institut fuer Genetik und Allgemeine Biologie, Universitaet Salzburg, A5020 Salzburg, Austria

We have sequenced and fully annotated a 65,871-bp region of mouse Chromosome 17 including the *Hba-ps4*  $\alpha$ -globin pseudogene. Comparative sequence analysis with the functional  $\alpha$ -globin loci at human Chromosome 16p13.3 and mouse Chromosome II shows that this segment of mouse Chromosome 17 contains a group of three  $\alpha$ -like pseudogenes (*Hba-psm-Hba-ps4-Hba-q3*), similar to the duplicated sets found at the functional mouse cluster on Chromosome II. In addition, exons 7 to 12 of the *mLuc7L* gene are present just downstream from the pseudogene cluster, indicating that this clone contains the region in which human 16p13.3 switches in synteny between mouse Chromosomes II and 17. Comparison of the sequences around the  $\alpha$ -like clusters on the two mouse chromosomes reveals the presence of conserved tandem repeats. We propose that these repetitive elements have played a role in the fragmentation of the mouse  $\alpha$  cluster during evolution.

[Supplemental material is available online at [www.genome.org](http://www.genome.org). The sequence data from this study have been submitted to GenBank/EMBL under accession no. AY016022.]

Genome sequencing projects and comparative genome analyses provide powerful tools to gain information on various aspects of gene function and regulation (Hardison and Miller 1993). Multiple alignments of large genomic segments containing genes with evolutionarily conserved patterns of expression have identified potential regulatory elements that have been maintained following speciation (see, e.g., Hardison et al. 1997; Oeltjen et al. 1997; Gottgens et al. 2000; Flint et al. 2001; Waterston et al. 2002). Such analyses have also proven to be valuable by increasing our understanding of the mechanisms underlying genome evolution (Hardison 2000).

The erythroid-specific  $\alpha$ -globin gene cluster represents one of the best-characterized models for studying mammalian gene regulation and genome evolution. We have previously compared long-range sequences of the  $\alpha$ -globin gene clusters from human, mouse, chicken, and pufferfish and identified a well-defined chromosomal unit of conserved synteny that may contain all the *cis*-acting sequences required for full regulation of  $\alpha$ -globin gene expression (Flint et al. 1997, 2001; Daniels et al. 2001). This region of conserved synteny spans ~112 kb of the mouse  $\alpha$ -globin cluster and the surrounding sequences, including all known long-range regulatory elements and several ubiquitously expressed genes lying on the telomeric side of the human  $\alpha$ -globin genes (Fig. 1; Flint et al. 2001).

The 5' break in synteny lies at least 126 kb upstream of the human  $\alpha$  cluster corresponding to a region ~87 kb upstream of the mouse  $\alpha$  cluster. In both species, this breakpoint lies beyond

the *IL9R* gene, which is a functional gene in mouse but a pseudogene in human (Kermouni et al. 1995; Vermeesch et al. 1997). In human, the *IL9R* pseudogene lies only 16 kb from the telomere of Chromosome 16 (Flint et al. 1997), whereas in mouse, the entire  $\alpha$  cluster lies at an interstitial chromosomal location (Leder et al. 1981, 1985; Tan and Whitney 1993). Therefore, as previously described, the breakpoint in synteny between these two species is delimited by the initial position of the human 16p subtelomeric region (Flint et al. 2001). It is of interest that whereas in most mammalian species the  $\alpha$  cluster appears to lie close to a telomere, the  $\alpha$  cluster in the rat, like mouse, also lies at an interstitial position (<http://genome-test.cse.ucsc.edu/>; see below), suggesting that the transition from a telomeric to an interstitial location may be a rearrangement that is specific to the rodent lineage. Beyond the *IL9R* gene, both mouse and rat sequences are homologous to human Chromosome 5 (<http://genome-test.cse.ucsc.edu/>).

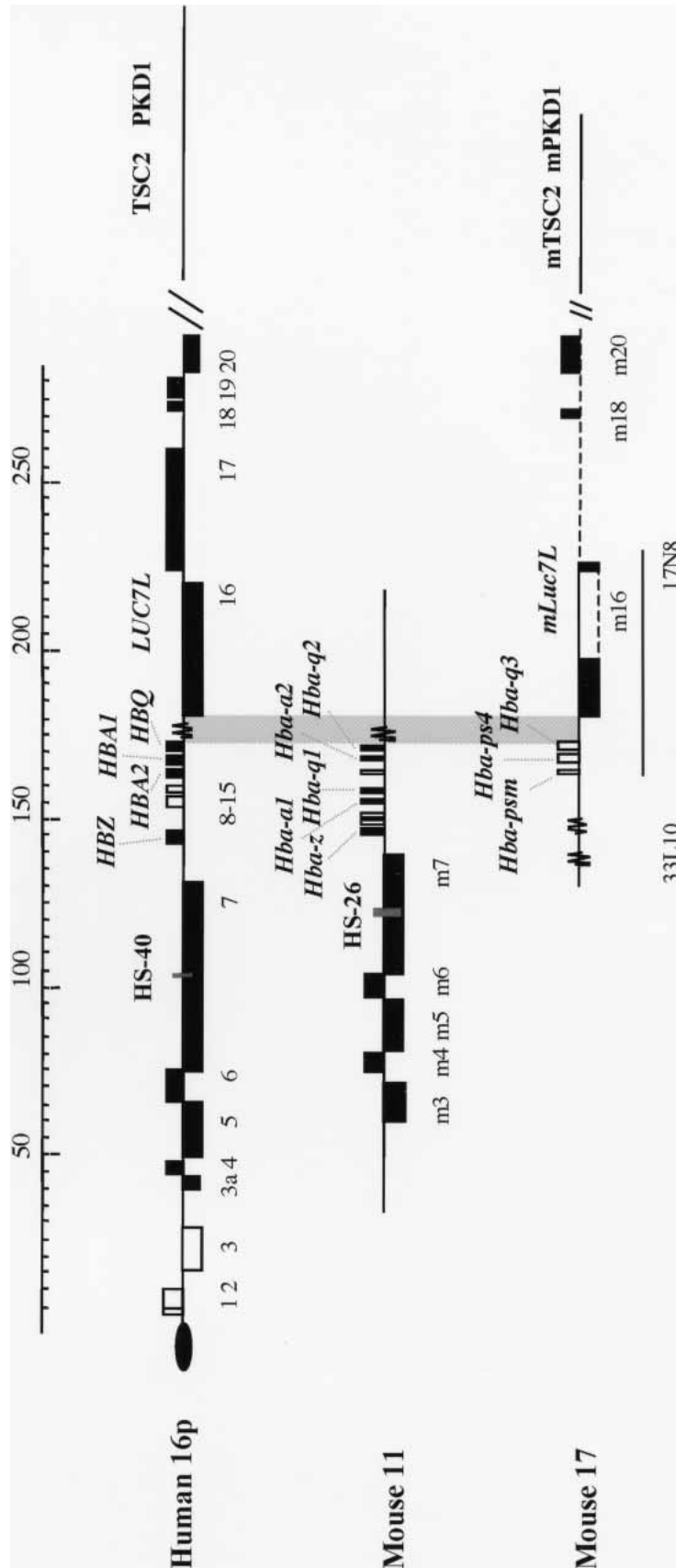
At the 3'-end of the cluster, sequence conservation between human, mouse, chicken, and pufferfish is abruptly lost just downstream from the  $\alpha$  cluster, immediately centromeric to the *HBQ1* ( $\theta$ ) gene on human Chromosome 16 (Flint et al. 2001). Remarkably, this synteny breakpoint lies close to the 3' limit of the chromatin domain that becomes hyperacetylated in human and mouse erythroid cells (Anguita et al. 2001), adding further evidence that this region might represent the minimal chromosomal domain required to fully regulate the  $\alpha$ -globin genes. It has therefore been of interest to characterize this region in some detail. Unfortunately, this 3' break in synteny has been difficult to define precisely because the organization of the structural  $\alpha$ -globin genes within the cluster differs between mouse and human such that the sequence conservation within and downstream from the  $\alpha$ -globin gene cluster is limited to the exons of the structural genes.

<sup>6</sup>Present address: Department of Genetics, University of Leicester, Leicester LE1 7RH, UK.

<sup>7</sup>Corresponding author.

E-MAIL [doug.higgs@imm.ox.ac.uk](mailto:doug.higgs@imm.ox.ac.uk); FAX 44 1865 222 500.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.2143604>.



**Figure 1** Homology relationship between human 16p13.3 and mouse Chromosomes 11 and 17. At the top is a schematic representation of the 16p13.3 telomere. The oval represents the telomeric repeats (TTAGG)<sub>n</sub>. Genes are shown as boxes numbered as in Flint et al. (1997). Genes above the line are transcribed toward the centromere; those below the line are transcribed toward the telomere. Empty boxes identify pseudogenes. The  $\alpha$ -globin regulatory element (HS-40) is shown in dark gray. In the *middle* is the interstitial region of mouse Chromosome 11 containing the  $\alpha$ -globin cluster. Genes are numbered in similarity to the human genes. At the *bottom* is the region of mouse Chromosome 17 spanning the mouse *LUC7L* homolog and the *Hba-ps4* pseudo- $\alpha$  gene. The dashed segment of *mLuc7L* indicates the region that has not yet been sequenced. The light gray area shows the 8-kb interval of the human 16p13.3 sequence within which lie the centromeric syntenic breakpoint and the centromeric limit of the human  $\alpha$ -globin regulatory domain. The symbol  $\blacksquare$  indicates the direct repeats found at the syntenic breakpoint. The mouse Chromosome 17 P1 clones 33L10 and 17N8 are shown.

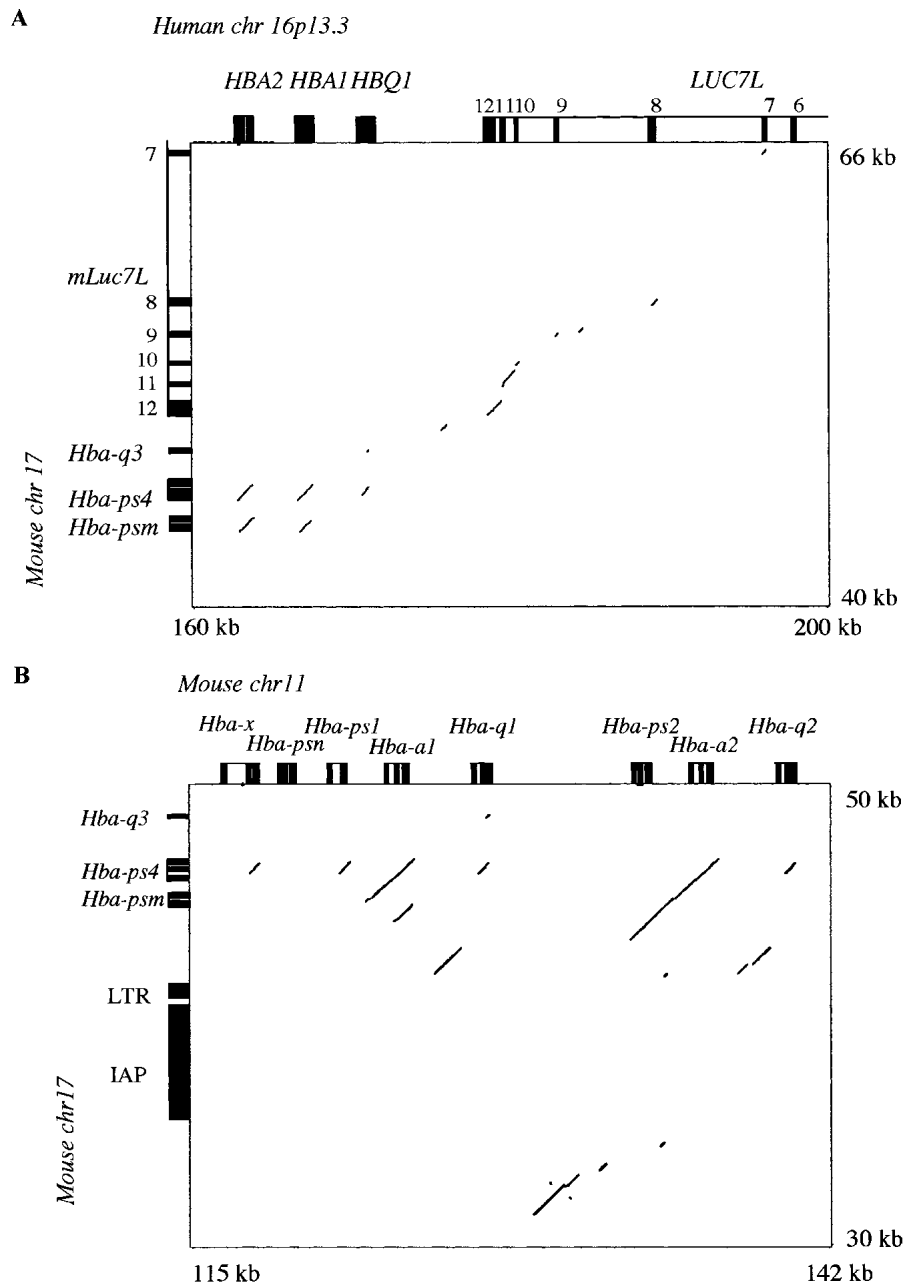
Recently, this analysis became more tractable when we identified the next gene lying centromeric to the human  $\alpha$ -globin cluster at 16p13.3 (*LUC7L*). This gene encodes a ubiquitous RNA-binding protein component of the U1 snRNP complex, which surprisingly maps to Chromosome 17 in the mouse, and is therefore separated from the functional  $\alpha$ -globin gene cluster located on Chromosome 11 (Fig. 1; Tufarelli et al. 2001). The region of human Chromosome 16p13.3 beyond *LUC7L* up to the mouse ortholog of the *PKD1* gene (~2 Mb) is also syntenic to mouse Chromosome 17 (Himmelbauer et al. 1992; Obermayr et al. 1995; Olsson et al. 1995, 1996; Daniels et al. 2001), indicating that the 3' breakpoint where human 16p13.3 switches in synteny between mouse Chromosomes 11 and 17 lies within the 8-kb segment separating the  $\theta$  and *LUC7L* genes (Fig. 1; Tufarelli et al. 2001).

A clue to the mechanism underlying this 11-to-17 translocation in mouse was provided by finding some remnants of the mouse  $\alpha$ -like globin genes on Chromosome 17. In addition to the active genes on Chromosome 11, two pseudo- $\alpha$  genes, *Hba-ps3* and *Hba-ps4*, have been identified and assigned to mouse Chromosomes 15 and 17, respectively (Leder et al. 1980, 1981; Vanin et al. 1980). *Hba-ps3* is a processed pseudogene, lacking the intervening sequences (Vanin et al. 1980), and might have moved to Chromosome 15 by transposition via an RNA intermediate (Leder et al. 1981). *Hba-ps4* is an  $\alpha$ -globin gene homolog having three exons and two introns (Leder et al. 1981). The mouse *Luc7L* (*mLuc7L*) gene is so closely linked to *Hba-ps4* that we identified two P1 clones spanning both genes and encompassing the breakpoint in synteny (Fig. 1; Tufarelli et al. 2001). To define this limit in more detail, we have now fully sequenced the P1 clone 33L10 from mouse Chromosome 17 including the *Hba-ps4* gene and the last six exons of *mLuc7L* (Fig. 1). Analysis of this sequence and comparisons of the human Chromosome 16, rat Chromosome 10, and mouse Chromosome 11 sequences have allowed us to propose a mechanism by which *Hba-ps4* and adjacent sequences may have been translocated from mouse Chromosome 11 to Chromosome 17.

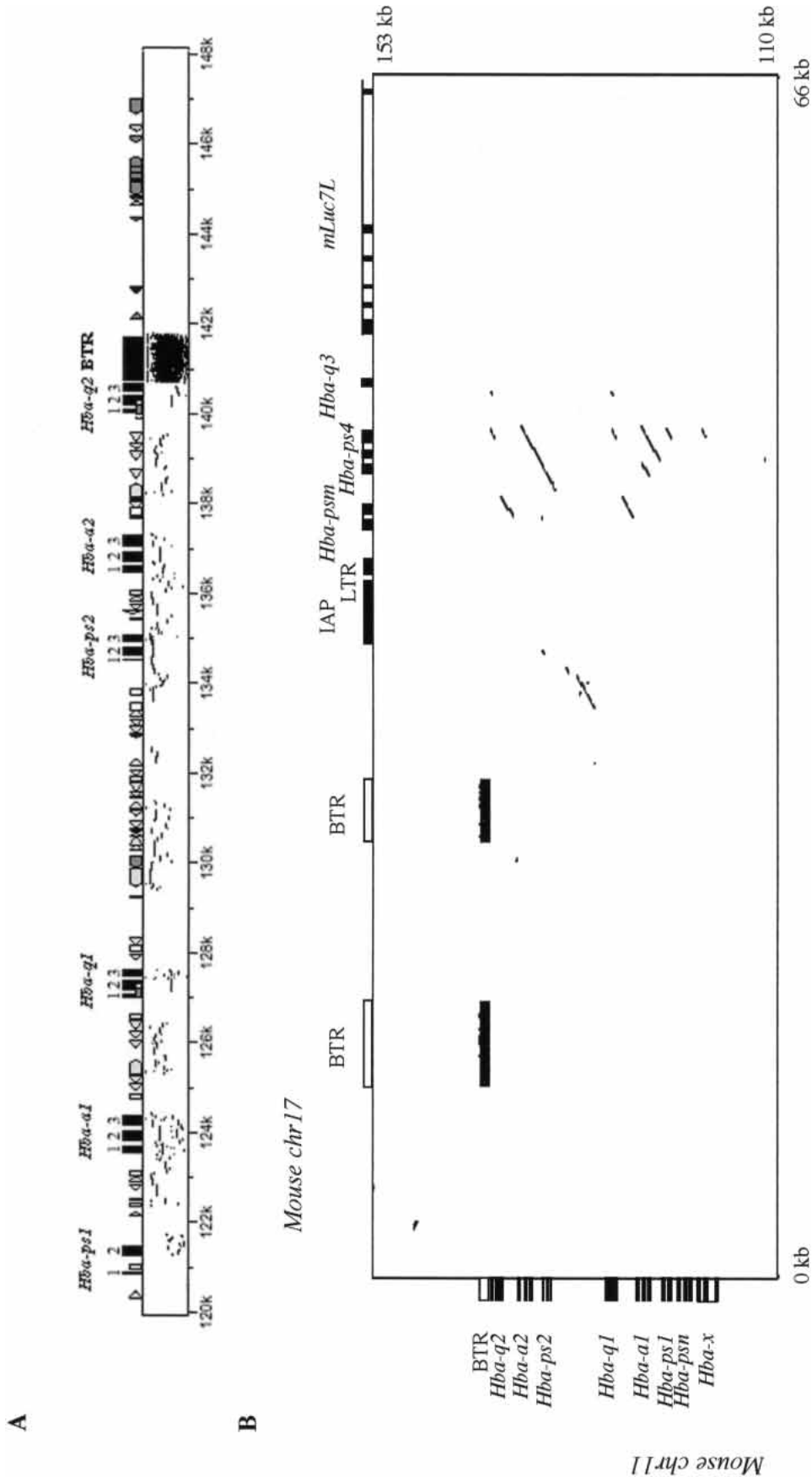
## RESULTS

Using an M13-based shotgun strategy (Wilson et al. 1994), we sequenced 65,871 bp of the P1 clone 33L10 from mouse Chromosome 17, known to contain *Hba-ps4* and the 3'-end of *mLuc7L* (Fig. 1; Sutherland et al. 1995; Tufarelli et al. 2001). Identification of the genes

in this sequence was mainly performed by comparison with the human 16p13.3 (Fig. 2A) and mouse Chromosome 11 (Figs. 2B and 3A) sequences. In Figure 2A is shown a dot plot of the mouse 33L10 sequence (coordinates 40 kb onward) against the human 16p13.3 sequence (coordinates 160–200 kb). This analysis reveals the presence of exons 7–12 of *mLuc7L* (spanning coordinates



**Figure 2** Dot plots of the mouse Chromosome 17 sequence against the human Chromosome 16 (A) and mouse Chromosome 11 (B) sequences. (A) Alignment with repeats masked of human and mouse sequences showing the presence of three  $\alpha$ -like pseudogenes and of *LUC7L* exons 7–12 in the 33L10 sequence (diagonals in the plot). Note that the three exons of *Hba-ps4* match human  $\alpha 1$  (*HBA1*) and  $\alpha 2$  (*HBA2*), and, to a lesser extent,  $\theta 1$  (*HBQ1*); *Hba-psm* contains two exons more related to human  $\alpha 1$  (*HBA1*) than  $\alpha 2$  (*HBA2*); *Hba-q3* consists of one exon only that matches, although poorly, human  $\theta 1$  (*HBQ1*). (B) Alignment of mouse 11 and 17 sequences with repeats not masked. Note that the pseudo- $\alpha$  gene set on Chromosome 17 shows more extensive similarity to the 3' set on Chromosome 11. The similarity extends from just 3' of the 5'- $\theta$  (*Hba-q1*) gene through to the 3'- $\theta$  (*Hba-q2*) gene on Chromosome 11, with a gap where an IAP element and an LTR have inserted on Chromosome 17 in the region of similarity to the Chromosome 11 segment just upstream of  $\psi\alpha 2$  (*Hba-ps2*).



**Figure 3** Blocks of tandem repeats (BTR) similar in sequence are found linked to the mouse  $\alpha$  and pseudo- $\alpha$  clusters on mouse Chromosomes 11 and 17, respectively. (A) Pip of the mouse  $\alpha$ -globin cluster on Chromosome 11 compared with mouse Chromosome 17 sequence with unmasked repeats. Along the top horizontal axis are shown genes and repeats in the mouse Chromosome 11. Gene exons and BTR are indicated and shown as black boxes. (B) Dot plot of mouse Chromosome 11 sequence (position 110–153 kb on the vertical axis) against mouse Chromosome 17 sequence (horizontal axis) with masked repeats. Note that 5' to the pseudo- $\alpha$  gene cluster on Chromosome 17 are found two BTRs similar in sequence to the BTR that immediately follows q2 (*Hba-q2*) on Chromosome 11 (two dark rectangles). The rectangles are generated by the clustering of a short 30-mer repeat for each 30-mer repeat unit of the BTRs.

51046 onward) and of the three exons of *Hba-ps4* (spanning coordinates 45,970–46,758). The transcriptional orientation of *mLuc7L* is opposite to that of *Hba-ps4*. This is the same orientation as the *LUC7L* and  $\alpha$ -globin genes at human 16p13.3 (Fig. 1; Flint et al. 1997).

Unexpectedly, the human–mouse comparison revealed that, in addition to *Hba-ps4*, which is similar to both human *HBA1* ( $\alpha 1$ ) and *HBA2* ( $\alpha 2$ ) genes, two additional  $\alpha$ -like globin pseudogenes are present on either side of *Hba-ps4* (Fig. 2A). The comparison reveals two exons with higher similarity to the human *HBA2* globin gene upstream of *Hba-ps4* ( $\psi\alpha$  or *Hba-psm*, spanning coordinates 44,293–44,742), and, downstream from *Hba-ps4*, one exon with matches to the human *HBQ1* ( $\theta$ ) gene ( $\psi\theta$  or *Hba-q3*, coordinates 48,724–48,735; Fig. 2A). This arrangement is interesting because previous analysis of the mouse  $\alpha$ -globin cluster suggested that during evolution there has been at least one duplication of a similar set of three genes ( $\psi\alpha$ – $\alpha$ – $\theta$ ) in the functional  $\alpha$ -globin cluster (Fig. 4). It therefore seems that *Hba-ps4* is part of a three-gene set (*Hba-psm*–*Hba-ps4*–*Hba-q3*) with a similar structure to the duplicated three-gene set on mouse Chromosome 11 (*Hba-ps*–*Hba-a*–*Hba-q*; Flint et al. 2001; R. Hardison and W. Miller, unpubl.). Comparison of mouse Chromosome 11 (coordinates 115–142 kb) and 17 (coordinates 30–50 kb) sequences confirms the presence of the three-gene set as a prominent block of homology encompassing the pseudogene and  $\alpha$ -globin genes (Fig. 2B). This analysis reveals that the degree of conservation is higher with the *Hba-ps2* and *Hba-a2* genes on Chromosome 11 (Fig. 2B). The dot plot shows that the homology block extends for ~5 kb 5' to the *Hba-ps2* on Chromosome 11, almost to the 3'-end of the *Hba-q1* gene on Chromosome 11 (Fig. 2B).

The gene set on Chromosome 17 has sustained insertions of an IAP (Intracisternal A Particle) element followed by an LTR (Long Terminal Repeat), and the  $\psi\theta$  (*Hba-q3*) gene, representing a  $\theta$ -like globin gene on Chromosome 17, has substantially diverged from its paralogs on Chromosome 11 (Fig. 2B). Therefore, the *Hba-psm*–*Hba-ps4*–*Hba-q3* set on mouse Chromosome 17 is more related to the 3' (*Hba-ps2*–*Hba-a2*–*Hba-q2*) set than to the 5' (*Hba-ps1*–*Hba-a1*–*Hba-q1*) set on mouse Chromosome 11, probably a consequence of the temporal order of the duplications of the three gene sets during evolution.

Another very interesting sequence feature is highlighted by comparing the sequences of mouse Chromosomes 11 and 17. The PIP (Percentage Identity Plot) of the alignment between these sequences in Figure 3A indicates that, in addition to the homologies between the  $\alpha$ -like globin genes, there is sequence conservation in the block of tandem repeats (BTR) found ~100 bp from the end of the *Hba-q2* gene on Chromosome 11, near the breakpoint in synteny (coordinates ~140.7–141.7 kb; R. Hardison and W. Miller, unpubl.). These repeats are composed mostly of 30-mers (e.g., TGTACATGTGCACCTCCTGTTTGATGCTGC), quite different in sequence from the 17-mers that constitute the human 3'-HVR (e.g., ACGGGGGGAACAGCGAC), which lies in a similar position downstream from the  $\alpha$ -globin cluster at 16p13.3 (Jarman et al. 1986). The dot plot of the mouse Chromosome 17 sequence against positions 110–153 kb in the mouse Chromosome 11 sequence clearly demonstrates two BTRs on mouse Chromosome 17, ~15 and 30 kb upstream of *Hba-psm* (coordinates ~10.4–15 kb and ~24–27.3 kb; Fig. 3B). These are seen as two dark rectangles created by a series of short parallel diagonals for each repeat unit (see Supplemental material). This indicates that the two blocks of tandem repeats on Chromosome 17 are composed of the same 30-mer units as the block of tandem repeats on Chromosome 11. However, whereas the BTR on Chromosome 11 includes 33 copies of the 30-mer, the two BTRs on Chromosome 17 are larger, containing 153 and 110 copies of the

repeat unit. A crossover mediated by alignment of some, but not necessarily all, of these repeats could provide a mechanism by which dispersal of the present day  $\alpha$ -globin clusters to mouse Chromosomes 11 and 17 may have occurred (Fig. 4; see Discussion).

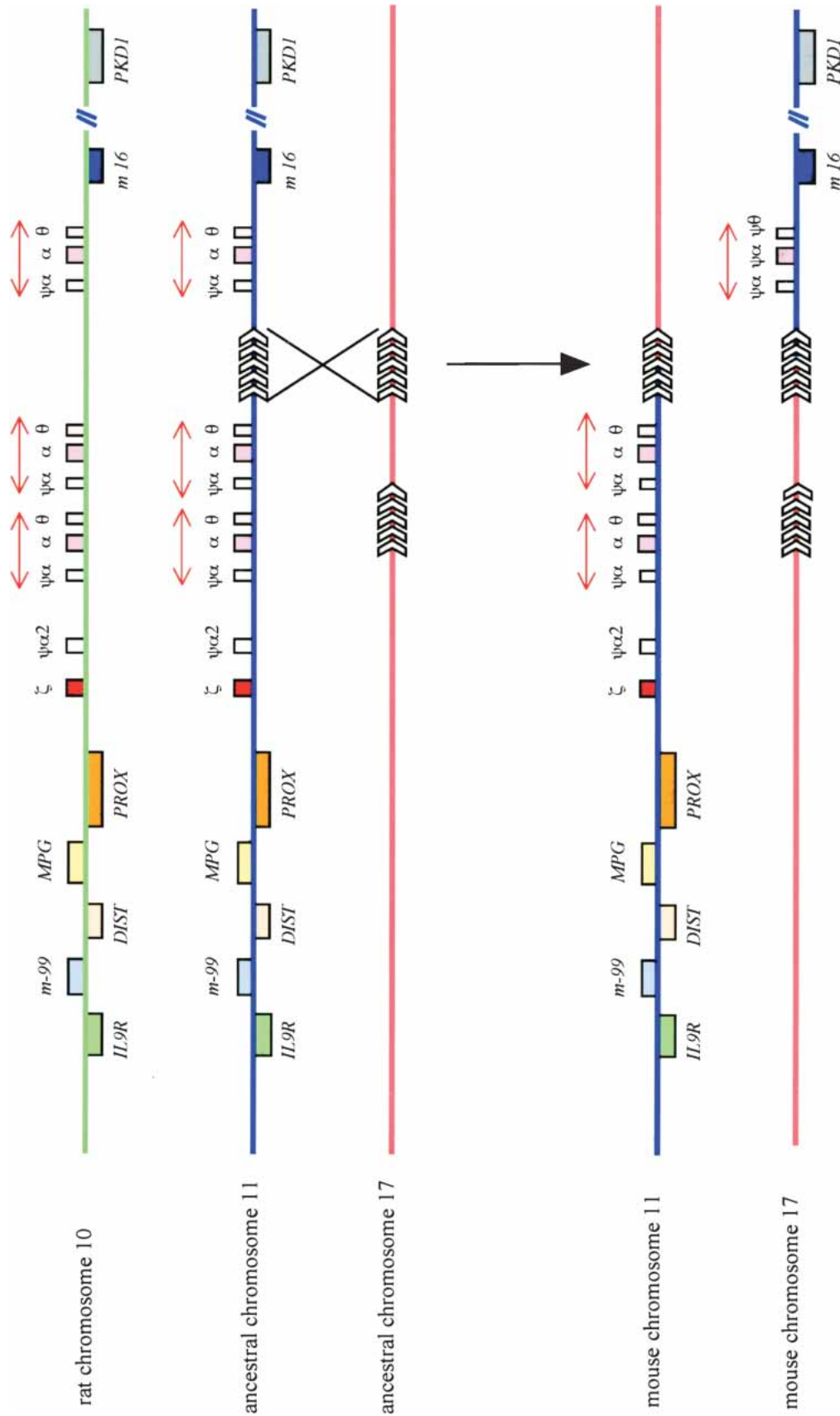
## DISCUSSION

We have previously reported a comparative sequence analysis of the human and mouse  $\alpha$ -globin gene clusters at human 16p13.3 and mouse Chromosome 11, respectively (Flint et al. 2001). We showed that although in both species the arrangement of the  $\alpha$ -like genes along the chromosome reflects their temporal order of expression (embryonic  $\zeta$  to adult  $\alpha$ ), the detailed organization of the globin genes varies between species. For example, in human the embryonic gene is duplicated ( $\zeta$ ,  $\psi\zeta$ ), whereas in mouse there is only one single  $\zeta$  gene. Similarly, in mouse the entire  $\psi\alpha$ – $\alpha$ – $\theta$  block has duplicated, whereas in human only the  $\alpha$  gene has duplicated. Therefore, these duplication events that generated the contemporary human and mouse clusters must have occurred after their diversion from a common ancestor. It is consequently difficult to propose a potential structure of the  $\alpha$ -globin cluster in the common mammalian ancestor, although it must have contained at least a copy of each  $\zeta$ ,  $\alpha$ , and  $\theta$  genes. In addition, the ancestral cluster was probably found in a GC-rich isochore at a telomeric location, like many of the extant mammalian  $\alpha$ -globin clusters (Hardison and Miller 1993; J. Hughes and D.R. Higgs, in prep.). In contrast, the mouse and rat  $\alpha$  clusters appear to have moved to an interstitial position when the rodent lineage formed (Hardison and Miller 1993). Consistent with this model, it has been shown that an interstitial telomere array is linked to the  $\alpha$ -globin cluster on mouse Chromosome 11 (Elliott and Pazik 1994).

The homology between human 16p13.3 and mouse Chromosome 11 ends at the 3'-end of the mouse 3'  $\theta$  (*Hba-q2*) gene on Chromosome 11 (Flint et al. 2001). We have now identified a cluster of three  $\alpha$ -like globin pseudogenes (*Hba-psm*–*Hba-ps4*–*Hba-q3*) on mouse Chromosome 17, highly related to the 3' *Hba-ps*–*Hba-a*–*Hba-q* gene set on Chromosome 11. Comparison of this sequence to the human sequence indicates that the homology to human 16p13.3 resumes at *Hba-psm*, within the sequenced region of Chromosome 17. An in-depth analysis of the sequence from mouse Chromosome 17 reveals the presence, upstream of the mouse  $\alpha$ -like pseudogenes, of duplicated blocks of tandem repeats similar to those found downstream from the  $\alpha$ -globin cluster on mouse Chromosome 11.

The organization and degree of similarity of the  $\alpha$ -like genes in the two mouse chromosomes and the location of the blocks of tandem repeats suggests a relatively simple model to explain the evolution of the mouse gene cluster. As shown in Figure 4, we propose that in the ancestral chromosome corresponding to Chromosome 11, three copies of the segment containing the three  $\alpha$ -like genes are present, with tandem repeats between the last two copies. Duplicated tandem repeats are present on the ancestral mouse Chromosome 17 and it is proposed that a recombination event mediated by a crossover between the tandem repeats may have partitioned the gene cluster on to the two chromosomes.

Circumstantial evidence supporting the structure of the hypothetical ancestral mouse Chromosome 11 comes from the analysis of rat chromosome 10 including the rat  $\alpha$ -globin cluster (<http://genome.ucsc.edu/>). The extent of the synteny between rat and human includes all the genes found telomeric to the  $\alpha$ -globin genes at human 16p13.3 and also the genes centromeric to *HBA*, including *LUC7L* and extending all the way to *AXIN1* and likely beyond (J. Hughes and D.R. Higgs, in prep.; R. Hardison, pers. comm.). Analysis of the publicly available rat genome



**Figure 4** A possible model to explain the evolution of the mouse  $\alpha$  cluster. In the top half are shown the structures of the present-day rat Chromosome 10 and of the hypothetical ancestral murine chromosomes. Genes are depicted as colored boxes with their corresponding names. The symbol indicates the blocks of tandem repeats. In the bottom half are shown the contemporary mouse Chromosome 11 and 17 gene structures. The double-headed arrows indicate the  $\psi\alpha$ - $\alpha$ - $\theta$  fragment that underwent duplication and divergence ( $\psi\alpha$ - $\alpha$ - $\theta$ ). The drawing is not to scale. See text for additional details.

sequence from Chromosome 10 (<http://genome.ucsc.edu/>) shows that the organization of the  $\alpha$ -globin genes in the rat cluster is the same as that in the mouse, but in the rat sequence, there are three copies of the three-gene set ( $\psi\alpha$ - $\alpha$ - $\theta$ ) found in mouse (J. Hughes and D.R. Higgs, in prep.). This organization bears striking similarity to the proposed structure of the ancestral mouse Chromosome 11 (Fig. 4), and indicates that a further duplication of the segment containing the  $\psi\alpha$ - $\alpha$ - $\theta$  genes occurred in the rodent lineage before rat and mouse diverged. The additional rearrangement, at the 3'-end of the  $\alpha$  cluster, described in this study is mouse-specific, that is, it happened after the divergence of mouse and rat, and resulted in the translocation of the third  $\psi\alpha$ - $\alpha$ - $\theta$  gene set from the ancestral mouse Chromosome 11 to mouse Chromosome 17.

It is interesting that this mouse-specific rearrangement separated the third gene set from the major  $\alpha$ -globin regulatory element ( $m\alpha$ RE or HS-26) found upstream of the  $\alpha$ -globin gene cluster on Chromosome 11. It is probably the lack of this and other important upstream regulatory elements that underlies the loss of function of the gene set on Chromosome 17, and this may ultimately have been responsible for these genes becoming pseudogenes. However, it is not possible to distinguish whether the loss of function happened at the same time or after the recombination event that separated the gene cluster to the two mouse chromosomes.

Sequence analysis has revealed that the breakpoints of the region of sequence similarity in and around the  $\alpha$ -globin gene clusters have been created independently in the various species (Flint et al. 2001). Intriguingly, blocks of tandem repeats, albeit different in sequence, are present in human, mouse, chicken, and dog (D.R. Higgs and J. Hughes, unpubl.), near or corresponding to the 3' breakpoint in synteny. Moreover, the 3' limit of the  $\alpha$ -globin acetylation domains of mouse and human lie close to or coincide with, these repetitive regions (Anguita et al. 2001). However, the repetitive elements are not conserved at the primary sequence level, indicating that the events that gave rise to the tandem repeats in the various species happened independently after speciation.

It is possible that the repeats represent a "signature" for the presence of a nearby chromatin boundary as observed in the chicken, mouse, and human  $\beta$ -globin clusters (Chung et al. 1993; Saitoh et al. 2000; Farrell et al. 2002), but such role in the regulation of the  $\alpha$ -globin genes or in the boundary formation is currently untested. Alternatively, their presence could result from an increased recombination rate in this region of the chromosome. In support of the presence of a putative recombination hot spot within the fragment containing the BTR, we have previously noted that in several deletions that cause  $\alpha$ -thalassaemia, the 3' breakpoints also cluster in this region (Horsley et al. 2001). Although the relationship between recombination hot spots and synteny breakpoints in the mouse genome has not yet been analyzed in detail, this is not the first time that tandem repeats have been identified at synteny breakpoints. For example, a previously reported comparison of human 19p13.3 to mouse Chromosome 10 has revealed an association between the presence of simple repeat sequences and chromosome breaks during evolution (Puttagunta et al. 2000).

## METHODS

The previously isolated (Sutherland et al. 1995) P1 clone 17N8 from the ICRF P703 C57BL/6 mouse genomic library was sequenced as described in Flint et al. (1997). Sequences were masked using RepeatMasker (<http://ftp.genome.washington.edu/RM/RepeatMasker.html>) and annotated by sequence homology as outlined in Flint et al. (2001). Sequence comparisons were performed using the PipMaker Web site (<http://bio.cse.psu.edu/>

pipmaker), which permits the comparison of two long DNA sequences. The alignments can be displayed either as a dot plot, allowing the identification of the homologous segments in both sequences, or as a percent identity plot (pip), showing both the position in one sequence and the degree of similarity for each aligning segment between the two sequences (Schwartz et al. 2000).

## ACKNOWLEDGMENTS

We thank Jonathan Flint and John Peden for the initial help with the sequence analysis, and Chris Talbot for helpful discussions. Support for this work was provided by the Medical Research Council (C.T., J.H., K.C., N.V., and D.R.H.), and the National Institutes of Health, USA Grant #DK27635 (R.H.) and Grant #HG02238 (W.M.). C.T. is a Royal Society Dorothy Hodgkin Research Fellow.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Anguita, E., Johnson, C.A., Wood, W.G., Turner, B.M., and Higgs, D.R. 2001. Identification of a conserved erythroid specific domain of histone acetylation across the  $\alpha$ -globin gene cluster. *Proc. Natl. Acad. Sci.* **98**: 12114–12119.
- Chung, J.H., Whiteley, M., and Felsenfeld, G. 1993. A 5' element of the chicken  $\beta$ -globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **74**: 505–514.
- Daniels, R.J., Peden, J.F., Lloyd, C., Horsley, S.W., Clark, K., Tufarelli, C., Kearney, L., Buckle, V.J., Doggett, N.A., Flint, J., et al. 2001. Sequence, structure and pathology of the fully annotated terminal 2 Mb of the short arm of human chromosome 16. *Hum. Mol. Genet.* **10**: 339–352.
- Elliott, R.W. and Pazik, J. 1994. An interstitial telomere array near Hba on mouse Chr 11 is a candidate for the homolog of the telomere at human 16p13.3. *Genomics* **27**: 217–218.
- Farrell, C.M., West, A.G., and Felsenfeld, G. 2002. Conserved CTCF insulator elements flank the mouse and human  $\beta$ -globin loci. *Mol. Cell. Biol.* **22**: 3820–3831.
- Flint, J., Thomas, K., Micklem, G., Raynham, H., Clark, K., Doggett, N.A., King, A., and Higgs, D.R. 1997. The relationship between chromosome structure and function at a human telomeric region. *Nat. Genet.* **15**: 252–257.
- Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R.J., Hardison, R., Miller, W., Philipsen, S., Tan-Un, K.C., McMurrow, T., et al. 2001. Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the  $\alpha$  globin cluster. *Hum. Mol. Genet.* **10**: 371–382.
- Gottgens, B., Barton, L.M., Gilbert, G.R., Bench, A.J., Sanchez, M.-J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M., et al. 2000. Analysis of the vertebrate *SCL* loci identifies conserved enhancers. *Nat. Biotech.* **18**: 181–186.
- Hardison, R. 2000. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet.* **16**: 369–372.
- Hardison, R. and Miller, W. 1993. Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol. Biol. Evol.* **10**: 73–102.
- Hardison, R.C., Oeltjen, J., and Miller, W. 1997. Long range human-mouse sequence alignments reveal novel regulatory elements: A reason to sequence the mouse genome. *Genome Res.* **7**: 959–966.
- Himmelbauer, H., Pohlschmidt, M., Snarey, A., Germino, G.G., Weinstat-Saslow, D., Somlo, S., Reeders, S.T., and Frischauf, A.-M. 1992. Human-mouse homologies in the region of the polycystic kidney disease gene (PKD1). *Genomics* **13**: 35–38.
- Horsley, S.W., Daniels, R.J., Anguita, E., Raynham, H.A., Peden, J.F., Villegas, A., Vickers, M.A., Green, S., Wayne, J.S., Chui, D.H., et al. 2001. Monosomy for the most telomeric, gene-rich region of the short arm of human chromosome 16 causes minimal phenotypic effects. *Eur. J. Hum. Genet.* **9**: 217–225.
- Jarman, A.P., Nicholls, R.D., Weatherall, D.J., Clegg, J.B., and Higgs, D.R. 1986. Molecular characterisation of a hypervariable region downstream of the human  $\alpha$ -globin gene cluster. *EMBO J.* **5**: 1857–1863.
- Kermouni, A., Van Roost, E., Arden, K.C., Vermeesch, J.R., Weiss, S.,

- Godelaine, D., Flint, J., Lurquin, C., Szikora, J.-P., Higgs, D.R., et al. 1995. The IL-9 receptor gene (IL9R): Genomic structure, chromosomal localization in the pseudoautosomal region of the long arm of the sex chromosomes, and identification of IL9R pseudogenes at 9qter, 10pter, 16pter, and 18pter. *Genomics* **29**: 371–382.
- Leder, P., Hansen, J.N., Konkel, D., Leder, A., Nishioka, Y., and Talkington, C. 1980. Mouse globin system: A functional and evolutionary analysis. *Science* **209**: 1336.
- Leder, A., Swan, D., Ruddle, F., D'Eustachio, P., and Leder, P. 1981. Dispersion of  $\alpha$ -like globin genes of the mouse to three different chromosomes. *Nature* **293**: 196–200.
- Leder, A., Weir, L., and Leder, P. 1985. Characterization, expression and evolution of the mouse embryonic  $\zeta$ -globin gene. *Mol. Cell. Biol.* **5**: 1025–1033.
- Obermayr, F., Sutherland, H.F., Kraus, B., and Frischauf, A.-M. 1995. Mouse cyclin F maps to a conserved linkage group on mouse chromosome 17. *Mamm. Genome* **6**: 149–150.
- Oeltjen, J.C., Malley, T.M., Muzny, D.M., Miller, W., Gibbs, R.A., and Belmont, J.W. 1997. Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. *Genome Res.* **7**: 315–329.
- Olsson, P.G., Sutherland, H.F., Nowika, U., Korn, B., Poustka, A., and Frischauf, A.-M. 1995. The mouse homologue of the tuberlin gene (TSC2) maps to a conserved linkage group between mouse chromosome 17 and human 16p13.3. *Genomics* **25**: 339–340.
- Olsson, P.G., Lohning, C., Horsley, S., Kearney, L., Harris, P.C., and Frischauf, A.-M. 1996. The mouse homologue of the polycystic kidney disease gene (Pkd1) is a single-copy gene. *Genomics* **34**: 233–235.
- Puttagunta, R., Gordon, L.A., Meyer, G.E., Kapfhamer, D., Lamerdin, J.E., Kantheti, P., Portman, K.M., Chung, W.K., Jenne, D.E., Olsen, A.S., et al. 2000. Comparative maps of human 19p13.3 and mouse chromosome 10 allow identification of sequences at evolutionary breakpoints. *Genome Res.* **10**: 1369–1380.
- Saitoh, N., Bell, A.C., Recillas-Targa, F., West, A.G., Simpson, M., Pikaart, M., and Felsenfeld, G. 2000. Structural and functional conservation at the boundaries of the chicken  $\beta$ -globin domain. *EMBO J.* **10**: 2315–2322.
- Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., and Miller, W. 2000. PipMaker—A Web server for aligning two genomic DNA sequences. *Genome Res.* **10**: 577–586.
- Sutherland, H.F., Pick, E., Francis, F., Lehrach, H., and Frischauf, A.-M. 1995. Mapping around the Fused locus on mouse Chromosome 17. *Mamm. Genome* **6**: 449–453.
- Tan, H. and Whitney III, J.B. 1993. Genomic rearrangement of the  $\alpha$ -globin gene complex during mammalian evolution. *Biochem. Genet.* **31**: 473–484.
- Tufarelli, C., Frischauf, A.-M., Hardison, R., Flint, J., and Higgs, D.R. 2001. Characterization of a widely expressed gene (LUC7-LIKE; LUC7L) defining the centromeric boundary of the human  $\alpha$ -globin domain. *Genomics* **71**: 307–314.
- Vanin, E.F., Goldberg, G.L., Tucker, P.W., and Smithies, O. 1980. A mouse  $\alpha$ -globin-related pseudogene lacking intervening sequences. *Nature* **286**: 222–226.
- Vermeesch, J.R., Petit, P., Kermouni, A., Renauld, J.-C., Van Den Berghe, H., and Marynen, P. 1997. The IL-9 receptor gene, located in the Xq/Yq pseudoautosomal region, has an autosomal origin, escapes X inactivation, and is expressed from the Y. *Hum. Mol. Genet.* **6**: 1–8.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* **22**: 4139–4147.

## WEB SITE REFERENCES

- <http://bio.cse.psu.edu/pipmaker>; PipMaker.  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>;  
 RepeatMasker Documentation page.  
<http://genome.ucsc.edu/>; UCSC Genome Browser.

Received November 4, 2003; accepted in revised form January 6, 2004.