Insights from Human/Mouse Genome Comparisons

Len A. Pennacchio

Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598

and

Genome Sciences Department
Lawrence Berkeley National Laboratory
One Cyclotron Road
Berkeley, CA 94720
Phone: 510-486-7498
Fax: 510-486-4229
Email: LAPennacchio@lbl.gov

Large-scale public genomic sequencing efforts have provided a wealth of vertebrate sequence data poised to provide insights into mammalian biology. These include deep genomic sequence coverage of human, mouse, rat, zebrafish, and two pufferfish (*Fugu rubripes* and *Tetraodon nigroviridis*) (Aparicio et al. 2002; Lander et al. 2001; Venter et al. 2001; Waterston et al. 2002). In addition, a high-priority has been placed on determining the genomic sequence of chimpanzee, dog, cow, frog, and chicken (Boguski 2002). While only recently available, whole genome sequence data have provided the unique opportunity to globally compare complete genome contents. Furthermore, the shared evolutionary ancestry of vertebrate species has allowed the development of comparative genomic approaches to identify ancient conserved sequences with functionality. Accordingly, this review focuses on the initial comparison of available mammalian genomes and describes various insights derived from such analysis.

Comparison of the Human and Mouse Genomes

The completion of a working draft assembly of the human genome in 2001 provided the first opportunity to catalog a mammalian genome and determine gene organization along its chromosomes (Lander et al. 2001; Venter et al. 2001). With the recent completion of a draft sequence for the mouse genome, similar analyses are possible (Waterston et al. 2002). While the mouse genome sequence received the distinction as the mammalian genome runner-up, its availability has provided the first chance for a comparison of two mammalian genomes. In this section, several examples of the large number of insights derived from human/mouse genome analysis are described.

One somewhat unexpected finding based on comparison of the human and mouse genomes was the difference in total genome size. While this wasn't obviously apparent based on previous gene by gene comparisons, a whole genome analysis revealed the mouse genome is ~14% smaller than humans (Waterston et al. 2002). It has been proposed that the genome size difference between human (2.9Gb) and mouse (2.5Gb) is due to higher rates of sequence deletion in the mouse genome. Though the mouse genome is slightly smaller, it was confirmed that the gene content and its linear organization along the chromosome are highly similar to that in humans (Waterston et al. 2002). Whole genome comparison revealed ~340 human/mouse syntenic segments, only slightly more than previous estimates of shared co-linear genomic blocks maintained since the last common ancestor of humans and mice (Hudson et al. 2001; Nadeau et al. 1984).

A second noted difference between the human and mouse genomes has been the overall rate of neutral substitution. By examination of ancient repeats which arose prior to the last common ancestor of humans and mice, it was found that mice have approximately twice as many nucleotide substitutions compared to humans (0.34 versus 0.17 substitutions per site) (Waterston et al. 2002). While it remains unclear why this has occurred, one possible explanation is the shorter generation time and consequent increased amount of selection in mice relative to humans (Li et al. 1996). Regardless, these findings have important implications for mammalian comparative genomics. As an example, the observation of increased rates of evolutionary sequence change in mammals with shorter generation times readily translates into decreased orthologous DNA sequence conservation. In Figure 1, the VISTA comparative sequence plot provides pair-wise comparison of the apolipoprotein A1 (*APOA1*) gene in human, rabbit and mouse (Mayor et al. 2000). In this analysis, it can be visualized that the highest degree of conservation exists between human/rabbit, followed by human/mouse and lastly mouse/rabbit. This is consistent with mice and rabbits having a shorter generation relative to humans. This phenomena support that a wide range of sequence divergence is present in *Mammalia* depending on the individual species selected for comparison. Thus, it is likely that as additional mammalian species are sequenced, future comparative genomic efforts will be even better poised to identify functionally conserved sequences.

In addition, human/mouse comparison confirmed that the mammalian genome has evolved at a non-uniform rate across the genome (Gottgens et al. 2001; Hood et al. 1995; Jimenez et al. 1992; Koop et al. 1994; Pennacchio et al. 2001b; Waterston et al. 2002). By examining 5Mb windows across the human/mouse genome, it was shown that nucleotide substitution rates vary in ancestral repeats greater than ten-fold more than expected under a uniform substitution process (Waterston et al. 2002). Again, the explanation for this observation is unknown but is hypothesized to be a function of local differences in DNA sequence mutation rates. In mammalian comparative genomics, this finding also has important implications since each region of the genome has evolved at varying rates. This emphasizes the importance of identifying the proper evolutionary distance for sequence comparisons to provide the ideal window for identifying conserved sequences with potential functionality. In response to this finding, the UC Santa Cruz Genome Browser has added a comparative sequence track that takes into account regional variation in DNA sequence change (Kent et al. 2002). For instance, human/mouse comparative data are presented as a track with running plots displaying "L-scores" to indicate the level of conservation (Figure 2). Regions of high conservation in otherwise non-conserved regions receive higher L-scores than regions of conservation in relatively highly conserved intervals. This strategy incorporates the assumption that conservation in regions with faster neutral rates of change is more likely to be functional than conservation in slowly evolving intervals. As an example, a known liver enhancer upstream of the *APOA1* gene is found to coincide with a conserved region receiving a high L-score in an interval devoid of other conserved noncoding sequences (Figure 2). This supports the notion that the identification of regions of the human/mouse genome with high L-scores can reveal functionally important sequences.

The Nature of Conserved Mammalian DNA

Comparison of the human and mouse genomes led to the unexpected result that ~40% of the human genome basepairs can be aligned to the mouse genome at the nucleotide level (Waterston et al. 2002). Similar analysis using a more stringent criterion ($\geq$70% nucleotide identity over $\geq$100bp) further supports the existence of greater than one million human/mouse conserved elements (Couronne et al. 2003). What explanations are there for this large amount of human/mouse conservation?

Might a significant fraction of the large number of human/mouse conserved elements be due to mis-alignments? Comparison of three different human/mouse whole genome alignments revealed that less than 80% of identified conserved noncoding sequences overlapped (Couronne et al. 2003; Waterston et al. 2002). This supports that current human/mouse alignments are not perfect and there are several explanations for this inconsistency. First, analysis of one human/mouse whole genome alignment found that ~2% of the human genome was covered by more than one mouse sequence. This may be due to significant problems with alignment programs (false-positive alignments) or real mouse duplication events. Indeed, analysis of coding regions indicates that approximately 80% of mouse genes have a direct 1:1-ortholog in the human genome but the remaining 20% of mouse genes are not unique to the mouse genome (Waterston et al. 2002). The large number of gene family expansions and deletions that have occurred independently in both humans and mice accounts for this discrepancy. Furthermore, separate analyses indicate that the human genome appears to have ~5% segmental duplication with blocks of 10kb or greater (Bailey et al. 2002). Thus, these genome peculiarities create clear problems in genome to genome alignments. Regardless, with three independent human/mouse alignment programs overlapping ~80% of conserved noncoding sequences, a large fraction of human/mouse sequence conservation appears to have been correctly identified. Further refinements of genome alignment algorithms and strategies should reduce the relatively small amount of potential non-orthologous alignment.

One obvious category of DNA displaying conservation between human/mouse is exons. This is not unexpected due to the known functional importance of the proteins that they encode and the similar gene content between these two mammalian species. In both species, it has been estimated that there are ~200,000 exons, the vast majority of which share an orthologous relationship (Lander et al. 2001; Venter et al. 2001; Waterston et al. 2002). Making the assumption that most exons are conserved between human/mouse, this still leaves greater than 800,000 conserved elements that do not coincide with coding sequence. Thus, more than four out of five conserved human/mouse elements appear to reside in noncoding DNA.

With all this apparent conservation outside of coding sequence, what (if any) function(s) might this large number of elements perform? Crude estimates have suggested that ~5% of the mammalian genome is under selection due to functional constraints (Waterston et al. 2002). If we assume that ~1.5% of the genome is functionally constrained due to exons, then ~3.5% of the genome (twice as much as in exons) is conserved due to noncoding DNA functions. Based on the apparent existence of approximately 200,000 human/mouse conserved elements which overlap exons, we therefore might conclude that ~400,000 conserved elements are due to functions in noncoding DNA. While these estimates are based on many extrapolations, they provide a framework for understanding the nature of conserved human/mouse DNA. A major challenge ahead lies in determining which fraction of noncoding conservation is functional and their exact biological functions.

While we have very few clues as to the immediate functional significance of mammalian noncoding conservation, gene regulatory elements represent one category of functional noncoding DNA. Though our ability to identify gene regulatory elements is limited, recent success utilizing comparative genomic strategies have proven their ability to uncover important gene regulatory elements based solely on conservation (Dubchak et al. 2000; Duret et al. 1997; Gottgens et al. 2000; Hardison 2000; Hardison et al. 1997; Pennacchio et al. 2001b). While most transcription factors bind extremely short DNA sequences, it has been found that many gene regulatory elements reside within much larger blocks (50-1000bp) of conservation. It appears that gene regulatory elements are a composite of numerous adjacent transcription factor binding sites that dictate gene expression. Thus, screening for evolutionarily conserved noncoding DNA is a powerful strategy to reveal gene regulatory elements. Complementary efforts are also needed to catalog noncoding DNA with functions beyond gene regulation such as chromosomal pairing, replication and higher order chromatin structure.

Comparative Sequence Analysis

The recent explosion of genomic sequence availability from higher eukaryotes has sparked new strategies to uncover functional regions of the mammalian genome (Aparicio et al. 2002; Dehal et al. 2002; Lander et al. 2001; Venter et al. 2001; Waterston et al. 2002). Traditional approaches required the painstaking isolation of DNA fragments and largely random searches to identify functional coding and noncoding elements. Today, DNA sequence data are routinely obtained using computational homology searches (Altschul et al. 1990) and subsequent requests for clones maintained in large cDNA and genomic clone repositories (Lennon et al. 1996; Osoegawa et al. 2001; Osoegawa et al. 2000). In addition to immediately providing sequence information and reagents for biologists, this technical shift has had a profound secondary effect. Specifically, the wealth of sequence data has provided the means for comparing the genomic sequence from numerous vertebrate species. This strategy is based on the search for evolutionary conserved sequences and the hypothesis that highly conserved sequences are functionally important. The new paradigm of utilizing genomic sequence data to select regions of the human genome for functional studies has provided a vast dataset ripe for targeted analysis. While still in its infancy for vertebrate genomes, the power of this approach has become readily apparent (Aparicio et al. 2002; Duret et al. 1997; Elgar et al. 1996; Hardison 2000; Hardison et al. 1997; Hedges et al. 2002; Loots et al. 2000; Pennacchio et al. 2001b; Waterston et al. 2002).

Strategies

Several methods are available for comparative genome analysis of the mammalian genome (Delcher et al. 2002; Gottgens et al. 2001; Jareborg et al. 2000; Margot et al. 1989; Mayor et al. 2000; Schwartz et al. 2000). In all cases, their primary goal is to turn raw orthologous sequence information from multiple species into visually interpretable plots to drive biological discovery. Self-input data can be used with web-based tools such as VISTA or PipMaker ((Mayor et al. 2000) http://www-gsd.lbl.gov/vista/; (Schwartz et al. 2000) http://bio.cse.psu.edu/pipmaker/). These two programs provide similar data output with slight differences in their alignment algorithm (AVID or BLASTZ) and visualization format (VISTA or PIP) (Bray et al. 2003; Schwartz et al. 2003).

As an example of the VISTA and PIP outputs, the human/mouse mitochondria outer membrane protein 40 (*TOM40*) gene was compared by both programs (Figure 3). Though very little is known about the function of mammalian *TOM40*, note the high level of conservation of exons between human and mouse. In addition, a highly conserved noncoding sequence upstream of the *TOM40* proximal promoter is visible. Such a conserved peak represents a candidate noncoding sequence for a role in regulating the *TOM40* gene and warrants further experimental investigation.

In addition to custom comparative genomic analysis driven by user input sequences, pre-processed whole genome comparative data are also available. VISTA Genome Browser is a web-based browser for interactively visualizing comparative sequence data in a VISTA plot format (http://www.pipeline.lbl.gov). Alternatively, the UCSC Genome Browser contains a vast amount of genome annotation with features that include comparative sequence data (http:/genome.ucsc.edu) (Kent et al. 2002). These two browsers contain thorough gene annotation data that provide a searchable entry point and reference for comparative plots. Though most of the comparative genome data are currently for human versus mouse, the rat genome has recently been integrated into both databases.

Discoveries

Even with the only recent generation of genomic sequence for various vertebrates, numerous biological discoveries have been possible. Examples include the identification of new genes and the inventory of differences in gene family content. For instance, through comparative sequence analysis, the apolipoprotein A5 (*APOA5*) gene was recently uncovered in an extremely well-studied genomic interval based solely on its conservation between humans and mice (Pennacchio et al. 2001a). Initiated by this purely sequence based discovery, follow-up functional studies have consistently showed the importance of *APOA5* in influencing triglyceride levels in humans and mice (Endo et al. 2002; Nabika et al. 2002; Pennacchio et al. 2001a; Pennacchio et al. 2002; Ribalta et al. 2002; Talmud et al. 2002). Other insightful studies have targeted olfactory and pheromone receptor gene families and revealed vast differences in gene family number between humans and mice (Dehal et al. 2001). For both receptor classes, large numbers of human gene family members were found to have been inactivated since the last common ancestor of humans and mice. These gene content differences are strongly correlated with the reduced smell and pheromone response in humans relative to mice.

Comparative genomics has also yielded insights into gene regulation. For example, examination of an interleukin (IL) gene cluster on human chromosome 5 revealed a highly conserved noncoding sequence (CNS1) which was subsequently shown to be responsible for the co-activation of IL-4, IL-5 and IL-13 (Loots et al. 2000). In a second unrelated study, the stem cell leukemia (SCL) gene was examined in human, mouse, chicken, fugu and zebrafish (Gottgens et al. 2002). Through the use of "phylogenetic footprinting" (Gumucio et al. 1996) across these five species, two highly conserved promoter sequences were identified and shown to be necessary for full SCL expression in erythroid cells. These briefly described studies provide several examples of a comparative genomic starting point which can reveal both coding and noncoding functional elements.

Comparative Utility of Non-Mammalian Species

Examination of highly conserved human/fish sequence has also proved to be a powerful approach for uncovering novel mammalian genes (Aparicio et al. 2002;

Gilligan et al. 2002). Early on in the Human Genome Project the pufferfish, *Fugu rubripes*, was proposed as a reference genome to aid in annotating human sequence (Aparicio et al. 1997; Brenner et al. 1993; Elgar 1996). This was based on the evolutionary position of pufferfish relative to humans as well as its notably compact genome size. As an example, 85 kb of fugu sequence containing 17 known genes was compared to homologous regions in the human genome and based on conservation three novel human genes were identified (Gilligan et al. 2002). This one specific example confirms the utility of fugu as a complementary species for comparative genomics and supports its ability to pinpoint additional novel genes. Indeed, recent genome-wide comparisons between the human and fugu genomes have revealed ~1,000 novel human genes based solely on conservation (Aparicio et al. 2002).

Comparative genomics has also supported that human/fugu conservation can reveal gene regulatory elements. In a seminal study, Aparicio et al (1995) compared the mouse and fugu *Hoxb-4* gene and revealed several intervals of noncoding DNA displaying high levels of homology (Aparicio et al. 1995). Functional studies on one conserved element indicated the mouse sequence was responsible for neural expression in a transgenic mouse assay, and in parallel work that the orthologous fugu sequence could also direct hind-brain expression. Several additional examples where human/fugu comparisons led to putative gene regulatory elements are available (Abrahams et al. 2002; Aparicio et al. 1995; Bagheri-Fam et al. 2001; Gilligan et al. 2002; How et al. 1996; Kimura et al. 1997; Rowitch et al. 1998; Venkatesh et al. 1997). These studies highlight that a comparative genomic driven approach can be successful for identifying gene regulatory elements even in distant mammal/fish comparisons.

A second powerful way to use human/fugu comparisons is to identify distant mammalian regulatory elements. Since the fugu genome is about one-eighth the size of the mammalian genome, both exons and gene regulatory elements are on average ~90% closer in fugu relative to mammals. Furthermore, these sequences are likely to be functional since they have been conserved during 450 million years of vertebrate evolution. Thus, the identification of noncoding conserved sequences that are in close proximity to a fugu gene but are found at a distance from the mammalian ortholog, would be expected to reveal distant gene regulatory elements. As an example, comparisons were performed between 3,700 kb around the human SOX9 gene and 195 kb of orthologous fugu sequence (Bagheri-Fam et al. 2001). This analysis revealed five conserved noncoding elements up to 290 kb upstream of human SOX9, while all of these elements were found within 18 kb of the fugu promoter. Functional studies of these elements have not yet been performed, though this study suggests that fugu comparisons to the mammalian genome can be used to reduce the search space for long-range gene regulatory elements and to direct attention to sequences otherwise unlikely to be tested functionally (Bagheri-Fam et al. 2001; Gilligan et al. 2002; Hedges et al. 2002; Venkatesh et al. 2000).

Conclusion

The recent completion of the mouse genome has provided the first opportunity to compare the whole genomes of two mammalian species. In addition to cataloging genome similarities and differences between human/mouse, these data have driven the development of comparative sequence databases and web-accessible browsers for the retrieval of evolutionarily-conserved sequence data. Such resources are providing biologists with an entry point for comparative sequence-based analysis of a gene or interval of interest. Through this portal, several biological insights have already been gained, with many more anticipated in the near future.

While computational approaches for comparing and analyzing sequence from multiple species are rapidly advancing, the ability to assign biological function to sequence has lagged. Accordingly, the most important challenge in deciphering the vocabulary of the mammalian genome lies in the area of high throughput biological studies linked to sequence. Comparative genomic strategies are providing large datasets of highly conserved mammalian sequences but high throughput biological experimentation is necessary to ultimately prove their function.

With the recent draft sequence completion of the rat genome and additional mammalian species slated in the near future, powerful multiple species sequence comparisons will be possible. These data will further drive comparative genomic discoveries of additional coding and noncoding functional elements within the genomes of mammals. Furthermore, deep sequence alignments will allow for the reconstruction of the last common ancestor of *Mammalia* and will help to explain changes that are unique to a given species. Through a better understanding of genomic sequence speciation events, we will likely be able to explain the genetic basis for numerous biochemical and structural differences between mammalian species.

Acknowledgements

This work was supported in part by the NIH-NHLBI Programs for Genomic Application Grant HL66681 through the U.S. Department of Energy under contract no. DE-AC03-76SF00098.

Figure Legends

Figure 1: Pair-wise VISTA comparisons of the human, mouse and rabbit apolipoprotein A1 (*APOA1*) gene reveals different levels of DNA sequence conservation. Human/mouse, human/rabbit and mouse/rabbit sequence comparisons are plotted in the top, middle and bottom panel, respectively. As an example, in the top panel ~6kb of human sequence is represented on the x-axis and the percent similarity to mouse is plotted on the y-axis. The graphical plot is based on sliding window analysis of the underlying genomic alignment, in this illustration a 100bp window is used which slides at 40bp nucleotide increments. Gene orientation and exon location are depicted above the VISTA plot. Note the lower level of conservation for both coding and noncoding sequences for mouse versus rabbit comparisons relative to either human/mouse or human/rabbit.

Figure 2: UCSC Genome Browser output for mouse/human sequence comparison of the *APOA1* gene. In this case, mouse sequence is represented on the horizontal axis while sequence similarity to human is indicated in the "human consensus plot". The human L-scores take into account the context of the level of conservation. Conservation in relatively non-conserved regions receive higher L-scores than similar conservation in relatively highly conserved regions. An experimentally defined *APOA1* liver enhancer is indicated within the plot and is found within a high L-score interval.

Figure 3: Human versus mouse *TOM40* genomic sequence comparison. VISTA **(A)** and PipMaker **(B)** analysis of the identical human/mouse sequence. In both panels, human is the reference sequence with percent similarity to mouse plotted on the vertical axis. Both programs provide gene orientation (arrows) and exon location (rectangles) above each panel. Each PIP horizontal bar indicates regions

of similarity based on the percent identity of each gap-free segment in the alignment. Once a gap (insertion or deletion) is found within the alignment, a new bar is created to display the adjacent correspondent gap-free segment.

References

Abrahams BS, Mak GM, Berry ML, Palmquist DL, Saionz JR, Tay A, et al. (2002) Novel vertebrate genes and putative regulatory elements identified at kidney disease and NR2E1/fierce loci. Genomics 80, 45-53

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215, 403-10

Aparicio S, Brenner S (1997) How good a model is the Fugu genome? Nature 387, 140

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, et al. (2002) Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297, 1301-10

Aparicio S, Morrison A, Gould A, Gilthorpe J, Chaudhuri C, Rigby P, et al. (1995) Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes. Proc Natl Acad Sci U S A 92, 1684-8

Bagheri-Fam S, Ferraz C, Demaille J, Scherer G, Pfeifer D (2001) Comparative genomics of the SOX9 region in human and Fugu rubripes: conservation of short regulatory sequence elements within large intergenic regions. Genomics 78, 73-82

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, et al. (2002) Recent segmental duplications in the human genome. Science 297, 1003-7

Boguski MS (2002) Comparative genomics: the mouse that roared. Nature 420, 515-6

Bray N, Dubchak I, Pachter L (2003) AVID: A Global Alignment Program. Genome Res 13, 97-102

Brenner S, Elgar G, Sandford R, Macrae A, Venkatesh B, Aparicio S (1993) Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. Nature 366, 265-8

Couronne O, Poliakov A, Bray N, Ishkhanov T, Ryaboy D, Rubin EM, et al. (2003) Strategies and Tools for Whole-Genome Alignments. Genome Res 13, 73-80

Dehal P, Predki P, Olsen AS, Kobayashi A, Folta P, Lucas S, et al. (2001) Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. Science 293, 104-11

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, et al. (2002) The Draft Genome of Ciona intestinalis: Insights into Chordate and Vertebrate Origins. Science 298, 2157-2167

Delcher AL, Phillippy A, Carlton J, Salzberg SL (2002) Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res 30, 2478-83

Dubchak I, Brudno M, Loots GG, Pachter L, Mayor C, Rubin EM, et al. (2000) Active conservation of noncoding sequences revealed by three-way species comparisons. Genome Res 10, 1304-6.

Duret L, Bucher P (1997) Searching for regulatory elements in human noncoding sequences. Curr Opin Struct Biol 7, 399-406

Elgar G (1996) Quality not quantity: the pufferfish genome. Hum Mol Genet 5, 1437-42

Elgar G, Sandford R, Aparicio S, Macrae A, Venkatesh B, Brenner S (1996) Small is beautiful: comparative genomics with the pufferfish (Fugu rubripes). Trends Genet 12, 145-50

Endo K, Yanagi H, Araki J, Hirano C, Yamakawa-Kobayashi K, Tomura S (2002) Association found between the promoter region polymorphism in the apolipoprotein A-V gene and the serum triglyceride level in Japanese schoolchildren. Hum Genet 111, 570-2

Gilligan P, Brenner S, Venkatesh B (2002) Fugu and human sequence comparison identifies novel human genes and conserved non-coding sequences. Gene 294, 35

Gottgens B, Barton LM, Chapman MA, Sinclair AM, Knudsen B, Grafham D, et al. (2002) Transcriptional regulation of the stem cell leukemia gene (SCL)--comparative analysis of five vertebrate SCL loci. Genome Res 12, 749-59

Gottgens B, Barton LM, Gilbert JG, Bench AJ, Sanchez MJ, Bahn S, et al. (2000) Analysis of vertebrate SCL loci identifies conserved enhancers. Nat Biotechnol 18, 181-6

Gottgens B, Gilbert JG, Barton LM, Grafham D, Rogers J, Bentley DR, et al. (2001) Long-range comparison of human and mouse SCL loci: localized regions of sensitivity to restriction endonucleases correspond precisely with peaks of conserved noncoding sequences. Genome Res 11, 87-97

Gumucio DL, Shelton DA, Zhu W, Millinoff D, Gray T, Bock JH, et al. (1996) Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the beta-like globin genes. Mol Phylogenet Evol 5, 18-32

Hardison RC (2000) Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet 16, 369-72

Hardison RC, Oeltjen J, Miller W (1997) Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. Genome Res 7, 959-66

Hedges SB, Kumar S (2002) Genomics. Vertebrate genomes compared. Science 297, 1283-5

Hood L, Rowen L, Koop BF (1995) Human and mouse T-cell receptor loci: genomics, evolution, diversity, and serendipity. Ann N Y Acad Sci 758, 390-412

How GF, Venkatesh B, Brenner S (1996) Conserved linkage between the puffer fish (Fugu rubripes) and human genes for platelet-derived growth factor receptor and macrophage colony-stimulating factor receptor. Genome Res 6, 1185-91

Hudson TJ, Church DM, Greenaway S, Nguyen H, Cook A, Steen RG, et al. (2001) A radiation hybrid map of mouse genes. Nat Genet 29, 201-5

Jareborg N, Durbin R (2000) Alfresco--a workbench for comparative genomic sequence analysis. Genome Res 10, 1148-57

Jimenez G, Gale KB, Enver T (1992) The mouse beta-globin locus control region: hypersensitive sites 3 and 4. Nucleic Acids Res 20, 5797-803

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. (2002) The human genome browser at UCSC. Genome Res 12, 996-1006

Kimura C, Takeda N, Suzuki M, Oshimura M, Aizawa S, Matsuo I (1997) Cis-acting elements conserved between mouse and pufferfish Otx2 genes govern the expression in mesencephalic neural crest cells. Development 124, 3929-41

Koop BF, Hood L (1994) Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. Nat Genet 7, 48-53

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409, 860-921

Lennon G, Auffray C, Polymeropoulos M, Soares MB (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics 33, 151-2

Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D (1996) Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol Phylogenet Evol 5, 182-7

Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, et al. (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 288, 136-40.

Margot JB, Demers GW, Hardison RC (1989) Complete nucleotide sequence of the rabbit beta-like globin gene cluster. Analysis of intergenic sequences and comparison with the human beta-like globin gene cluster. J Mol Biol 205, 15-40

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. (2000) VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16, 1046-7

Nabika T, Nasreen S, Kobayashi S, Masuda J (2002) The genetic effect of the apolipoprotein AV gene on the serum triglyceride level in Japanese. Atherosclerosis 165, 201-4

Nadeau JH, Taylor BA (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. Proc Natl Acad Sci U S A 81, 814-8

Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, et al. (2001) A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res 11, 483-96

Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, et al. (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome Res 10, 116-28.

Pennacchio LA, Olivier M, Hubacek JA, Cohen JC, Cox DR, Fruchart JC, et al. (2001a) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. Science 294, 169-73

Pennacchio LA, Olivier M, Hubacek JA, Krauss RM, Rubin EM, Cohen JC (2002) Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. Hum Mol Genet 11, 3031-3038

Pennacchio LA, Rubin EM (2001b) Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet 2, 100-9

Ribalta J, Figuera L, Fernandez-Ballart J, Vilella E, Castro Cabezas M, Masana L, et al. (2002) Newly identified apolipoprotein AV gene predisposes to high plasma triglycerides in familial combined hyperlipidemia. Clin Chem 48, 1597-600

Rowitch DH, Echelard Y, Danielian PS, Gellner K, Brenner S, McMahon AP (1998) Identification of an evolutionarily conserved 110 base-pair cis-acting regulatory sequence that governs Wnt-1 expression in the murine neural plate. Development 125, 2735-46

Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison R, et al. (2003) Human-Mouse Alignments with BLASTZ. Genome Res 13, 103-107

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, et al. (2000) PipMaker--a web server for aligning two genomic DNA sequences. Genome Res 10, 577-86

Talmud PJ, Hawe E, Martin S, Olivier M, Miller GJ, Rubin EM, et al. (2002) Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. Hum Mol Genet 11, 3039-3046

Venkatesh B, Gilligan P, Brenner S (2000) Fugu: a compact vertebrate reference genome. FEBS Lett 476, 3-7

Venkatesh B, Si-Hoe SL, Murphy D, Brenner S (1997) Transgenic rats reveal functional conservation of regulatory controls between the Fugu isotocin and rat oxytocin genes. Proc Natl Acad Sci U S A 94, 12462-6

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. (2001) The sequence of the human genome. Science 291, 1304-51

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520-62