comment
reviews
reports
deposited research
refereed research
interactions
information

Opinion

What use is the human genome for understanding the mouse?

Paul Denny, Rachael Bate and Ann-Marie Mallon

Address: MRC UK Mouse Genome Centre and Mammalian Genetics Unit, Harwell, Oxon OX11 0RD, UK.

Correspondence: Paul Denny. E-mail: p.denny@har.mrc.ac.uk

Published: 22 October 2001

Genome ***Biology*** 2001, **2(11)**:comment2009.1–2009.5

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2001/2/11/comment/2009

Abstract

Having a working draft of the human genome sequence is proving invaluable to mouse genetic and genomic studies, providing a useful stepping-stone towards the finished sequence of the mouse genome.

All over the world, mouse geneticists have applauded the publication of the initial analysis of the human genome sequence(s) [1,2]. Why? One simplistic answer is that mouse and human are two flavors of mammal, and a genome sequence for one is a surrogate for the other. So perhaps the pertinent question then becomes: how can mouse geneticists make use of the human sequence? In this article, we briefly describe some ways in which the human working draft sequence can be used as a tool in mouse genomics, not only for assembling the mouse genome but also for identifying conserved sequence elements and providing new insights into genome evolution.

Before the genome

Even before the inception of 'The Human Genome Project', mouse and human genetics already formed a two-way street. An early example of this was the observation that inherited traits exhibiting sex-linkage in humans were also sex-linked in mice - for example, hypophosphatemia [3]. As genetic maps improved in both species, it became clear that there were blocks of conserved synteny, along chromosomes (synteny literally means 'on the same thread') [4]. Indeed, with the development of dense, genome-wide maps, it has become possible confidently to infer the location of a mouse homolog of a human gene, on the basis of the location of the genes that flank it in the human genome, and vice versa (Figure 1) [5,6].

Mice suffer from diseases similar to those of humans. Furthermore, this biological similarity can extend to defects in the

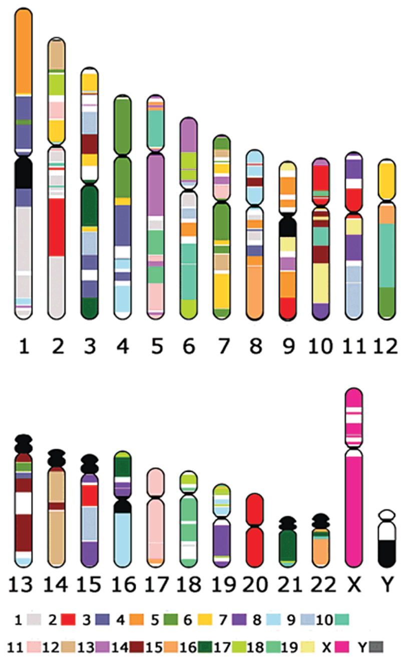

Figure 1
Schematic representation of mouse radiation-hybrid map segments overlaid on the human genome sequence. Each color corresponds to a particular mouse chromosome (bottom), and centromeres, subcentromeric heterochromatin and repetitive short arms are shown in black. Reproduced with permission from [6].

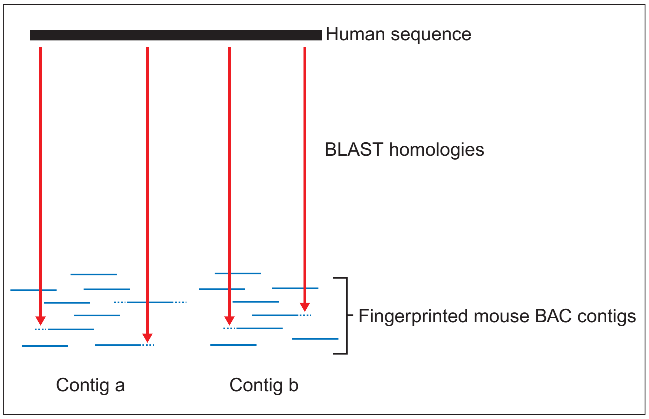

Figure 2
Building mouse clone contigs using a human scaffold. The solid black bar at the top of the diagram represents human genomic sequence, and the smaller clusters of horizontal lines are two fingerprinted mouse BAC contigs, a and b. Dashes at the ends of some of the BACs represent insert-end sequences. The vertical arrows indicate significant sequence homologies (detected using BLAST) between human genomic sequence and mouse BAC end sequence. These homologies are then used to anchor, orient and order the mouse BAC contigs relative to the human sequence. In some cases, the alignment of two or more contigs with the human sequence allows identification of BAC clone overlaps missed in the original contig building, so that contigs can be merged together.

same molecules: for example, mutations in the leptin gene cause morbid obesity [7] and in myosin VII cause deafness [8], both in humans and in mice. In complex genetic diseases, such as diabetes, or where mutations may be subtle rather than obviously deleterious, the mouse is of particular importance as it allows experimental testing of the validity of candidate mutations, by targeted mutagenesis. From the outset, the Human Genome Project recognized the importance of model organisms, from bacteria to mouse, and devoted funding to developing resources for their genetic and physical mapping [9]. As a genetically malleable social mammal, well suited to living and breeding in (relatively) modest space, the mouse has become the premier genetic model for humans. A secondary aspect of the publicly funded Human Genome Project that has been immensely valuable to scientists working on model organisms is the policy of rapid data release. This has meant that data could be used prior to formal publication.

Building mouse genome sequence using a human scaffold

One direct way in which the human genome sequence can be used in mouse genomics is as a scaffold to support the anchoring and merging of clone contigs (contiguous assemblies) of large-insert bacterial artificial chromosomes (BACs; Figure 2). Draft or finished human genomic sequence is compared with mouse sequence taken from the ends of BAC inserts [10]. BAC-insert ends showing highly significant similarity to the human genomic sequence are assumed to represent homologous sequences from conserved syntenic segments, where both gene function and gene order are conserved. Clone names for the BAC clones from which these sequences are derived are then used to search the public database of fingerprinted BAC clones (where the restriction digest band pattern on electrophoresis makes a 'fingerprint') constructed by the British Columbia Genome Sequencing Center (BC-GSC) [11]. This can be done in a number of ways: using the text version of the database on the BC-GSC website [11]; or by downloading the data onto a local computer and searching it using the fingerprinted contigs (FPC) software [12]; or by using an assembly and annotation website [13] maintained by the Center for Bioinformatics, University of Pennsylvania.

The version of the contig data generated in the last of these ways is made particularly powerful by the way in which it links together data from several sources: the sequence-tagged site (STS) content of the contigs, plus assemblies of expressed sequence tags (ESTs) in the database of transcribed sequences (DOTS) [14] and data from radiation hybrid (RH) maps. In some cases, this helps to confirm anchoring of mouse BAC contigs and also to orient the contigs on the mouse chromosome (M. Bucan, personal communication). Furthermore, a refinement of this approach has been used recently to produce a physical map of the whole mouse genome [15].

Identifying and visualizing conserved sequences

One of the most powerful uses of the human draft sequence again depends on the high level of sequence similarity between mouse and human. It is assumed that sequence elements with the highest similarity are those with critical functions, such as the transcribed and regulatory elements of genes. Evolutionary forces will have actively selected against mutations in these elements, whereas the sequences of non-functional genomic regions will acquire differences to an extent approximately proportional to the time passed since the divergence of the two organisms from a common ancestor. It can therefore be highly informative to take mouse and human sequences from a region of synteny, and to align them and graphically display the degree of sequence similarity (Figure 3) [16-18]. Known or predicted gene structures can be overlaid on the alignments displayed by the PipMaker and VISTA programs, allowing identification of novel evolutionarily conserved regions (ECRs) [19]. Such regions may represent coding exons not predicted by conventional methods, regulatory elements, genes expressed at low levels and so not represented in EST databases, or perhaps genes that are transcribed to make non-coding RNAs. A further development of the VISTA package incorporates

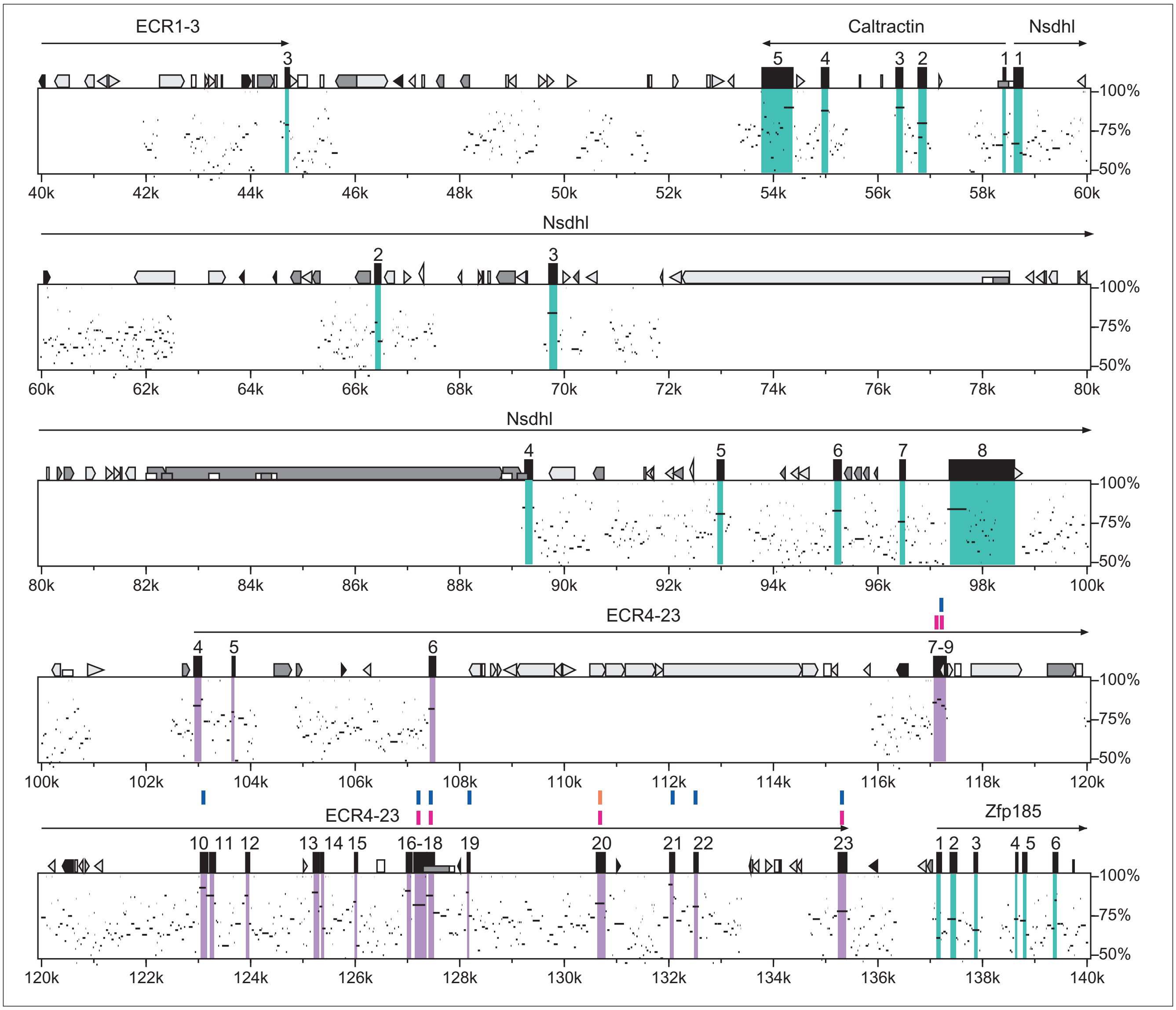

Figure 3
Percentage identity plot (PIP) of part of the 'bare patches' region of the mouse X chromosome between the genes *Nsdhl* and *Zfp185* [19]. The mouse genomic sequence is shown on the *x* axis, and the percentage identity (50-100%) on the *y* axis. Regions that are conserved between the two sequences and are greater than 50 bp in length are shown as lines within the plot. Confirmed and predicted exons are depicted as numbered black boxes along the top of the plot. Confirmed exons are also shaded in green within the plot, whereas putative exons are colored in purple. Repetitive elements are illustrated on the top line (grey pointed boxes are L1s, black pointed boxes are LINE2s, light grey triangles are SINES other than MIRs, black triangles are MIRs, dark grey triangles are LTR elements and dark grey boxes are other types of interspersed repeats; see [16,17] for details), alongside CpG islands (short dark grey boxes are CpG islands where CpG/GpC < 0.75, and short white boxes are CpG islands where the CpG/GpC ratio lies between 0.6 and 0.75). Computer predictions are shown for the putative genes in the ECRA4-A23 region above the plot, with gene predictions shown as blue boxes, open reading frames (ORFs) as pink boxes and sequence similarity with human ESTs as orange boxes.

prediction of transcription-factor binding sites in conserved regions, to aid identification of potential regulatory elements [20]. In some cases, experimental evidence supports the prediction that ECRs are transcribed ([19] and R.B., unpublished observations) or that they represent regulatory elements [21]. It seems that not all genomic regions acquire mutations at similar rates, however, so some sequences will be conserved as a result of insufficient time to diverge, adding 'noise' to the alignments. One way of improving the discrimination of actively and passively conserved sequences may be to make multiple comparisons with different species [22].

Genome evolution

Turning the idea of comparative analysis on its head, we can use the human genome to find areas of the mouse genome where conservation of gene content and order breaks down. What happens in these evolutionary 'breakpoint' regions? Studies of this kind are still quite rare, but there are suggestions of an emerging consensus. A correlation has been noted between genetic instability and sites that are rich in repetitive elements and may, therefore, be more prone to rearrangement through inappropriate homologous recombination [23]. Transposition events, leading to both insertion and deletion, are also apparent in regions of chromosomal rearrangements [24-26]. The evolutionary breakpoint disrupting the conservation of human 19p13.3 with mouse chromosomes 10 and 17, for example, is rich in simple tandem repeats [27], and repetitive elements were identified in the section of mouse chromosome 10 bridging the junction between conserved syntenic regions of human chromosomes 21 and 22 [28]. Indeed, many of the breaks in conservation of human chromosome 19 relative to mouse chromosomes seem to be associated with localized repetitive elements, such as tandemly repeated gene families [29]. As more extensive mouse genome sequence becomes available, it will be interesting to assess whether the prediction that regions rich in repetitive elements are associated with genome rearrangements holds up.

Finishing the mouse genome?

Finally, we would like to argue the case for producing a finished mouse genomic sequence. When mouse genome sequencing first became a serious endeavor, it was unclear what the quality of the 'product' might be. Some suggested that all that was needed was a low-to-medium coverage whole-genome shotgun (about 3-6-fold sequencing depth), which could be assembled and aligned with the finished human genome. Indeed, unassembled, low-coverage mouse shotgun sequence can be used efficiently to find exons in the human working draft sequence [30] and is a valuable resource for gene and marker discovery [31]. Lack of long-range contiguity hampers accurate prediction of mouse gene structure, however, and accurate prediction is invaluable for efficient mutation scanning. Studies of two critical classes of mutations would benefit from high-quality, finished sequence: point mutations such as those induced by the supermutagen ethylnitrosourea (ENU) [32], and mutations responsible for quantitative traits, as it is believed that many of the latter may be found in regulatory elements [33]. Despite the unequivocal utility of a draft human genome sequence [1,2], it is clear that draft sequence has limitations [34]. Even with a finished human genome sequence, interpretation of genome function will be enhanced by access to a second, finished mammalian genome. As the premier genetic model mammal, it makes sense to finish the mouse.

Acknowledgements

We thank Steve Brown for improvements to the manuscript.

References

1. International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al.*: **The sequence of the human genome.** *Science* 2001, **291:**1304-1351.
3. Eicher EM, Southard JL, Scriver CR, Glorieux FH: **Hypophosphatemia: mouse model for human familial hypophosphatemic (vitamin D-resistant) rickets.** *Proc Natl Acad Sci USA* 1976, **73:**4667-4671.
4. Shows TB, Brown JA, Chapman VM: **Comparative gene mapping of HPRT, G6PD, and PGK in man, mouse, and muntjac deer.** *Cytogenet Cell Genet* 1976, **16:**436-439.
5. **MGD** [http://www.informatics.jax.org/].
6. Hudson TJ, Church DM, Greenaway S, Nguyen H, Cook A, Steen RG, Van Etten WJ, Strivens MA, Trickett P, Heuston C, *et al.*: **A radiation hybrid map of mouse genes.** *Nat Genet* 2001, **29:**201-205.
7. O'Rahilly S: **Life without leptin.** *Nature* 1998, **392:**330-331.
8. Brown SDM, Steel KP: **Deafness (DFN) genes.** In *Encyclopedia of molecular medicine*. New York: John Wiley and sons, in press.
9. **National Human Genome Research Institute: Understanding Our Genetic Inheritance: The U.S. Human Genome Project. The First Five Years Fiscal Years 1991-1995** [http://www.nhgri.nih.gov/HGP/HGP_goals/5yrplan.html]
10. **TIGR: Mouse BAC Ends** [http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html]
11. **British Columbia Genome Sequence Center: A BAC fingerprint map of the mouse genome** [http://www.bcgsc.bc.ca/projects/mouse_mapping/]
12. Soderlund C, Humphray S, Dunham A, French L: **Contigs built with fingerprints, markers, and FPC V4.7.** *Genome Res* 2000, **10:**1772-1787.
13. **Mouse chromosome 5 annotation project** [http://www.cbil.upenn.edu/mouse/chromosome5/fpc-search.php3]
14. **Allgenes.org** [http://www.allgenes.org]
15. **Mouse Ensembl** [http://mouse.ensembl.org]
16. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W: **PipMaker-a web server for aligning two genomic DNA sequences.** *Genome Res* 2000, **10:**577-586.
17. **Pipmaker** [http://bio.cse.psu.edu/pipmaker]
18. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I: **VISTA: visualizing global DNA sequence alignments of arbitrary length.** *Bioinformatics* 2000, **16:**1046-1047.
19. Mallon AM, Platzer M, Bate R, Gloeckner G, Botcherby MR, Nordsiek G, Strivens MA, Kioschis P, Dangel A, Cunningham D, *et al.*: **Comparative genome sequence analysis of the Bpa/Str region in mouse and man.** *Genome Res* 2000, **10:**758-775.
20. **rVISTA** [http://www-gsd.lbl.gov/vista/rVISTA.html]
21. Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA: **Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons.** *Science* 2000, **288:**136-140.
22. Dubchak I, Brudno M, Loots GG, Pachter L, Mayor C, Rubin EM, Frazer KA: **Active conservation of noncoding sequences revealed by three-way species comparisons.** *Genome Res* 2000, **10:**1304-1306.
23. Carver EA, Stubbs L: **Zooming in on the human-mouse comparative map: genome conservation re-examined on a high-resolution scale.** *Genome Res* 1997, **7:**1123-1137.
24. DeBry RW, Seldin MF: **Human/mouse homology relationships.** *Genomics* 1996, **33:**337-351.
25. Ehrlich J, Sankoff D, Nadeau JH: **Synteny conservation and chromosome rearrangements during mammalian evolution.** *Genetics* 1997, **147:**289-296.
26. Kamnasaran D, O'Brien PC, Ferguson-Smith MA, Cox DW: **Comparative mapping of human chromosome 14q11.2-q13 genes with mouse homologous gene regions.** *Mamm Genome* 2000, **11:**993-999.
27. Puttagunta R, Gordon LA, Meyer GE, Kapfhamer D, Lamerdin JE, Kantheti P, Portman KM, Chung WK, Jenne DE, Olsen AS, *et al.*: **Comparative maps of human 19p13.3 and mouse chromosome 10 allow identification of sequences at evolutionary breakpoints.** *Genome Res* 2000, **10:**1369-1380.

28. Pletcher MT, Roe BA, Chen F, Do T, Do A, Malaj E, Reeves RH: **Chromosome evolution: the junction of mammalian chromosomes in the formation of mouse chromosome 10.** *Genome Res* 2000, **10:**1463-1467.
29. Dehal P, Predki P, Olsen AS, Kobayashi A, Folta P, Lucas S, Land M, Terry A, Ecale-Zhou CL, Rash S, *et al.*: **Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution.** *Science* 2001, **293:**104-111.
30. Bouck JB, Metzker ML, Gibbs RA: **Shotgun sample sequence comparisons between mouse and human genomes.** *Nat Genet* 2000, **25:**31-33.
31. **Ensembl** [http://www.ensembl.org]
32. Brown SD, Nolan PM: **Mouse mutagenesis-systematic studies of mammalian gene function.** *Hum Mol Genet* 1998, **7:**1627-1633.
33. Flint J, Mott R: **Finding the molecular basis of quantitative traits: successes and pitfalls.** *Nat Rev Genet* 2001, **2:**437-445.
34. Katsanis N, Worley KC, Lupski JR: **An evaluation of the draft human genome sequence.** *Nat Genet* 2001, **29:**88-91.